



42nd Annual Meeting
of the
Statistical Society of Canada

42^e Congrès annuel
de la
Société statistique du Canada

May 25 - 28, 2014

25 au 28 mai 2014

University of Toronto

Table of Contents • Table des matières

Table of Contents • Table des matières	1
Welcome to Toronto • Bienvenue à Toronto	2
Sponsors • Commanditaires	3
Exhibitors • Exposants	3
Organizers • Organismes	4
General Information • Informations générales	5
The Conference • Le congrès	7
Social Events • Activités sociales	9
Committees and Meetings • Comités et réunions	11
Social and Information Events • Événements sociaux et informatifs	13
List of Sessions • Liste des sessions	15
Program • Programme	19
Abstracts • Résumés	71
Author List • Liste des auteurs	298

Welcome to Toronto • Bienvenue à Toronto

The University of Toronto welcomes you to the 42nd Annual Meeting of the Statistical Society of Canada. The University of Toronto was founded as King's College in 1827 and now has three campuses Scarborough and Erindale and the historic St. George campus, with ten fully affiliated teaching hospitals in metropolitan Toronto. The University has over 65,000 undergraduate students, close to 16,000 graduate students, over 12,000 faculty and had approximately \$1.2 billion in research funds last year. The St. George campus is home to Hart House, which hosts Sunday Concerts as well as housing the Hart House Theatre. The campus surrounds the Royal Ontario Museum and is close to the Art Gallery of Ontario (AGO).

Toronto is on the North shore of Lake Ontario and is home to 2.6 million people and about 10 million trees. The city was originally founded as "The Town of York" by British colonial officials in 1793 on the border of Upper Canada, and became the "City of Toronto" in 1834. Tourist attractions include the CN tower, Casa Loma, the Ontario Science Centre, and the brand new Ripley's Aquarium. Just to the north of the St George campus is The Annex, an upscale shopping district. To the south-west is Kensington Market and Chinatown. A little further south is Queen Street West, Toronto's entertainment and fashion district. To the east is Bay Street, where you will find many of Toronto's financial firms. The city is host to the Toronto Symphony Orchestra, the Canadian Opera Company, and the National Ballet of Canada, and is home to the Blue Jays, the Raptors, the Argonauts, and the Maple Leafs.

By happy coincidence, Toronto's annual "Doors Open" event will be happening on May 24 & 25th. For this weekend only a wide range of buildings open their doors to the public, allowing you to see inside cultural centres, museums, blood donor centres, financial institutions and government buildings, all for free. A schedule is posted at <http://goo.gl/Nzugai>. The theme of this year's event is "Secrets and Spirits", but old favourites like Queen's Park (home to the Ontario Legislature) and City Hall will remain on the programme. Queen's Park will have free guided tours, which need to be booked in advance. The queues to visit the Mayor's office may be longer than usual.

For more information to help you plan your trip, please visit <http://www.seetorontonow.com>.

La University of Toronto vous souhaite la bienvenue au 42^e congrès annuel de la Société statistique du Canada. Baptisée King's College lors de sa fondation en 1827, la University of Toronto compte aujourd'hui trois campus (Scarborough, Erindale et le campus historique de St. George), ainsi que dix hôpitaux universitaires affiliés dans la région. L'université accueille plus de 65 000 étudiants de premier cycle et près de 16 000 étudiants de cycle supérieur, emploie plus de 12 000 enseignants et disposait pour 2013 d'environ 1,2 milliards de dollars de fonds de recherche. Hart House, situé sur le campus de St. George, accueille une série de concerts du dimanche et un théâtre. Le campus entoure le Musée royal de l'Ontario et n'est pas loin du Musée des beaux-arts de l'Ontario (ou AGO).

Toronto, située sur la rive nord du lac Ontario, compte 2,6 millions d'habitants et près de 10 millions d'arbres. Fondée en 1793 par des fonctionnaires coloniaux britanniques à la frontière du Haut-Canada, la « Ville de York » est devenue la « Ville de Toronto » en 1834. Parmi les attractions touristiques, signalons la Tour CN, Casa Loma, le Centres des sciences de l'Ontario et le tout nouveau Aquarium Ripley's. Juste au nord du campus de St George se trouve l'Annex, un quartier de boutiques de luxe ; au sud-ouest, le marché Kensington et le quartier chinois. Un peu plus au sud, vous tomberez sur la rue Queen ouest, le quartier de divertissement et de la mode de Toronto. À l'est se trouve la rue Bay, le quartier financier. La ville est le siège de l'Orchestre symphonique de Toronto, de la Compagnie d'opéra canadienne et du Ballet national du Canada, ainsi que d'équipes sportives professionnelles : les Blue Jays, les Raptors, les Argonauts et les Maple Leafs.

Par une coïncidence heureuse, les journées « Portes ouvertes » annuelles de Toronto se tiennent cette année les 24 et 25 mai. Toute une variété d'édifices ouvrent ces jours-là leurs portes au public, vous permettant de découvrir gratuitement des centres culturels, des musées, des centres de donneurs de sang, des institutions financières et des bâtiments administratifs. Vous en trouverez le calendrier à l'adresse suivante : <http://goo.gl/Nzugai>. Cette année, le thème de l'événement est « Secrets et esprits », mais les classiques tels Queen's Park (site de l'assemblée législative de l'Ontario) et l'Hôtel de ville restent aussi au programme. À Queen's Park, la visite guidée gratuite doit être réservée à l'avance. Les files d'attente pour visiter le bureau du Maire risquent d'être plus longues que d'habitude. Pour planifier votre séjour, veuillez consulter <http://fr.seetorontonow.ca>.

Sponsors • Commanditaires

The Statistical Society of Canada would like to thank each of the sponsors, whose generous contributions have made this conference possible:

La Société statistique du Canada désire remercier chacun de ses commanditaires dont la généreuse contribution a rendu possible la tenue de ce congrès :

- The University of Toronto
- Department of Statistical Sciences, University of Toronto
- The Fields Institute
- Pacific Institute for the Mathematical Sciences
- Centre de Recherches Mathématiques
- Canadian Statistical Sciences Institute /
Institut canadien des sciences statistiques

Exhibitors • Exposants

The exhibitors, providing displays for examination and purchase, can be found in the Great Hall of Hart House. The exhibitors will be available Monday, May 26 and Tuesday, May 27 from 9:00 am to 4:30pm.

Les kiosques des exposants sont situés dans le Great Hall de Hart House. Les exposants seront présents lundi 26 mai et mardi 27 mai de 9h00 à 16h30.

- Wiley <http://ca.wiley.com/WileyCDA>
- Nelson Education Ltd. <http://goo.gl/6vL6nG>
- Pearson Canada <http://www.pearsoncanada.ca>
- CRC Press <http://www.crcpress.com>
- Minitab <http://www.minitab.com>

Organizers • Organisateurs

Local Arrangements Committee • Comité des arrangements locaux

- Patrick Brown
University of Toronto & Cancer Care Ontario
- Jamie Stafford
University of Toronto
- Jon Lee
University of Toronto & Cancer Care Ontario)
- Eleanor Pullenayegum
University of Toronto & Hospital for Sick Children
- Stacie Bellemare
University of Toronto Continuing Professional Development Office

The local arrangements committee would like to thank all those who helped pull this event together. Among other responsibilities, our student volunteers have handled AV support and manned the registration desk. At the SSC Office, Larysa Valachko and Alan Kelm have provided crucial support for many aspects, including registration and exhibitors. Thanks to Rhonda Rosychuk and other previous local arrangements chairs we did not have to start from scratch, and have appreciated their willingness to answer our (numerous!) questions. The SSC Meetings Coordinator, Tim Swartz, and other SSC executive members have offered insights and opinions that helped inform key decisions. We also greatly appreciate the accurate and timely translations provided by Catherine Cox.

Le comité des arrangements locaux est très reconnaissant à la solide équipe de bénévoles locaux dévoués, ainsi qu'aux diverses facultés et institutions. Nous tenons tout particulièrement à remercier les étudiants diplômés bénévoles qui ont contribué, entre autres, au soutien audiovisuel et au bureau des inscriptions. Au bureau de la SSC, Larysa Valachko et Alan Kelm ont fourni un appui essentiel pour divers aspects, dont l'inscription et les exposants. Grâce à Rhonda Rosychuk et aux autres anciens présidents qui ont généreusement partagé leurs expériences et volontiers répondu à nos nombreuses questions. Le coordonnateur des congrès, Tim Swartz, et d'autres membres de l'exécutif de la SSC ont offert leurs avis et opinions et contribué à éclairer la prise de décisions clés. Nous avons également beaucoup apprécié les traductions précises et rapides fournies par Catherine Cox.

Program Committee • Comité du programme

- Changbao Wu (Chair • Président) *University of Waterloo*
- Joan Hu *Simon Fraser University*, Biostatistics - Biostatistique
- Michael Kouritzin *University of Alberta*, Probability - Probabilité
- Tom Loughin *Simon Fraser University*, Business and Industrial Statistics- Industrielle et Gestion
- Susana Rubin-Bleuer *Statistics Canada*, Survey Methods - Enquête
- Kevin Thorpe, *University of Toronto*, Statistical Education - Éducation en statistique
- Gord Willmot *University of Waterloo*, Actuarial Science - Science Actuarielle

General Information • Informations générales

Directions • Emplacement

Campus maps are available via the QR code below, at <http://goo.gl/1hsgoa>, in the conference bag and in this program. Those staying at Victoria University or any of the conference hotels will have a short walk to the session venues. Those preferring to take public transit will likely find the Toronto Transit Commission (TTC) to be an easy way to get around Toronto. Museum and Queen's Park subway stations are closest to the conference venues.

Cartes du campus sont disponibles avec le code QR ci-dessous, à <http://goo.gl/kV3IoT>, dans le sac de conférence et dans ce programme. Ceux qui s'héberge à l'Université de Victoria ou un hôtel de la conférence sont une courte distance à pied des lieux de la session. Ceux qui préfèrent se déplacer avec le transport en commun trouveront la Toronto Transit Commission (TTC) un moyen facile de se déplacer à Toronto. Les stations Museum et Queen's Park sont les plus proches des lieux du congrès.



Registration • Inscription

Registration will be available at:

- Sunday, May 25: 8:00 am to 4:30 pm
Stone Lobby Medical Sciences Building
<http://goo.gl/maps/tMzq5>
- Sunday, May 25: 6:30pm to 7:30pm
Royal Ontario Museum
- Monday, May 26: 7:30 am to 5:00 pm
Hart House Great Hall
- Tuesday, May 27: 8:15 am to 5:00 pm
Hart House Great Hall

Vous pourriez vous inscrire

- Dimanche 25 mai : 08h00 à 16h30
Stone Lobby Medical Sciences Building
<http://goo.gl/maps/tMzq5>
- Dimanche 25 mai : 18h30 à 19h30
Musée royal de l'Ontario
- Lundi 26 mai : 7h30 à 17h00
Hart House Great Hall
- Mardi 27 mai : 8h15 à 17h00
Hart House Great Hall

Parking on Campus • Stationnement sur le campus

Parking or driving in Toronto is a bad idea. If you do arrive in Toronto by car you are best off leaving your car at your hotel. The closest lot with a strictly positive probability of having space available is at Princess Margaret Hospital off Muray Street (see <https://goo.gl/maps/1NsKc>).

Stationner ou conduire à Toronto n'est pas recommandé. Si vous arrivez à Toronto en voiture, vous êtes suggéré de garder votre voiture à votre hôtel. Le terrain de stationnement le plus proche ayant une probabilité strictement positive pour avoir un espace disponible est au Princess Margaret Hospital sur la rue Muray (voir <https://goo.gl/maps/1NsKc>).

Campus Security • Sécurité sur le campus

Toronto is a very safe city, not only in comparison with major American cities but also in relation to other cities in Canada (see <http://goo.gl/3VITZF>). That said, you may wish to avoid Jarvis street and the areas to its immediate east at night.

Toronto est une ville très sécuritaire, non seulement par rapport aux grandes villes américaines mais aussi par rapport à d'autres villes au Canada (voir <http://goo.gl/QHq9e1>). Cela dit, vous avez raison d'éviter le quartier à l'est de la rue Jarvis les soirs.

- Fire/Police/Ambulance: 911
- Campus Police:
416-978-2222 (emergencies, 24 hrs)
416-978-2323 (general enquiries)
- Toronto Police Service: 416-808-2222

- Pompiers/Police/Ambulance : 911
- Force policière du campus :
416-978-2222 (urgences)
416-978-2323 (renseignements généraux)
- Force policière de Toronto : 416-808-2222

Internet Access • Accès internet

UofT: username SSC2014, password UofT

eduroam: Visit your home organization's link at <http://goo.gl/aRYJ18>.

UofT : username SSC2014, mot de passe UofT

eduroam : Visitez le lien de votre organisme chez <http://goo.gl/76jH6k>

Food on Campus and in the City • Nourriture sur le campus et en ville

A selection of food locations are listed on the conference web site (<http://goo.gl/IZRf0k>).

The Arbor room in Hart House, one level below the Great Hall, has good selection of hot and cold food served cafeteria style. Large cafeterias which include the likes of Tim Horton's can be found in the Medical Sciences Building and Sidney Smith Hall.

Among the very agreeable pubs on Bloor street are the Duke of York, Bedford Academy, and The Madison. Those with slim wallets and/or searching for an `authentic' experience might consider O'Grady's on College street.

Proper restaurants for the gastronomically inclined can be found on Baldwin and Harbord streets. The Gallery Grill in Hart House is the ultimate lunch destination in the area, reserve via 416-978-2445 or

<http://harthouse.ca/hours-reservations/>.

There is a grocery store at Bloor and Spadina, and another in the Manulife Building at Bay and Bloor.

Une sélection de restaurants est disponible sur le site web (<http://goo.gl/L5vivid>).

L'Arbor Room à Hart House qui est situé au première étage du Great Hall a une bonne sélection du nourriture chaude et froide en style cantine. Le Sidney Smith Hall et le Medical Sciences Building ont de larges cantines avec des cafés Tim Hortons.

Le Duke of York, Bedford Academy, et Le Madison font partis des pubs reconnus situés sur la rue Bloor. Ceux qui cherchent une expérience « authentique » (ou plutôt bas de gamme) et abordable peuvent considerer le pub O'-Grady's sur la rue College.

Les restaurants gastronomiques sont situées sur le rue Baldwin et Harbord. Le Gallery Grill qui est situés au Hart House est le meilleur choix pour le déjeuner. Pour réserver une table appeler 416-978-2445 ou visitez

<http://harthouse.ca/hours-reservations/>.

On peut trouver une épicerie sur la rue Bloor et Spadina, et une autre dans le Manulife Building situé à l'intersection de Bay et Bloor.

Athletics Facilities • Installations sportives

The University of Toronto has several athletic facilities. The University Athletic Centre is at Harbord and Spadina and has gyms, pools, an indoor running track, a strength and conditioning centre, cardio machines, tennis and squash courts. Day passes are available for \$18; a week pass is \$47 (<http://goo.gl/GkXS76>). Hart House has a gym, running track and pool; a month pass is \$73 (<http://harthouse.ca/fitness>).





L'Université de Toronto est munie de plusieurs installations sportives. Le Centre Sportif de l'Université se trouve à l'intersection de Harbord et Spadina, et comprend des gymnases, des piscines, une piste intérieure, un centre de force et de conditionnement, machines de cardio, de tennis et de squash. Un abonnement de jour coûte \$ 18 ; un billet d'une semaine est \$ 47 (<http://goo.gl/GkXS76>). A Hart House, il-y-a un gymnase, une piste de course et une piscine ; un abonnement durant un mois coûte \$ 73 (<http://harthouse.ca/fitness>).





The Conference • Le congrès

Language • Langue

An important feature of our meetings is the presentation of the abstracts and of the plenary session visual aids in both official languages. This translation was very ably carried out under the supervision of the Bilingualism Committee (chaired by Pierre-Jérôme Bergeron) and the translators Catherine Cox, Maya Giroux, Pascale Giroux-Denis and Caroline Petit-Turcotte.

Une caractéristique importante de nos congrès est la présentation des résumés et des supports visuels des sessions plénières dans les deux langues officielles. Cette traduction a été très habilement menée sous la supervision du Comité du bilinguisme (présidé par Pierre-Jérôme Bergeron) et des traductrices Catherine Cox, Maya Giroux, Pascale Giroux-Denis et Caroline Petit-Turcotte.

At the time that they submitted their abstract, speakers were asked to provide the language in which they intend to give their oral presentation as well as the language of their visual aids. Icons are used to provide this information for each paper. For the oral presentation, we have used the icons  and , whereas  and  indicate the language of the visual aids. The letter inside identifies the language: E for English and F for French. Please note that the visual aids for the plenary talks will be provided in both languages.

Lorsque les conférenciers ont soumis leur résumé, ils ont spécifié la langue dans laquelle ils comptaient faire leur présentation orale, ainsi que la langue du support visuel. À titre informatif, nous avons inclus cette information à l'aide d'icônes pour chaque présentation. Pour la présentation orale nous avons utilisé les icônes , , alors que  et  indiquent le support visuel. La lettre à l'intérieur identifie la langue : F pour le français et E pour l'anglais (English). Veuillez noter que le support visuel des conférences plénières sera présenté dans les deux langues.

Rooms • Salles

Coffee breaks, exhibits and posters are found in the Great Hall of Hart House (HH on maps). Plenary sessions on Monday and Tuesday mornings are in the Medical Sciences Auditorium (MS). The remainder of scientific sessions will occur in Hart House and University College (UC). Instructions for speakers can be found on the meeting website <http://goo.gl/j3qIb9>.

Les pauses-café, les exposants et les affiches sont au Great Hall de Hart House (HH sur les plans). Les séances plénières les matins de lundi et mardi sont au Medical Sciences Auditorium (MS). Les autres sessions scientifiques auront lieu à Hart House et University Collège (UC). Les instructions pour les présentateurs peuvent être trouvés au site du congrès <http://goo.gl/WSDfxS>.

Poster Sessions • Séances d'affichage

Contributed posters and Case Study posters will be in the Great Hall and adjoining areas, with Case study posters being displayed Monday and contributed research posters will appear Tuesday. The authors will be with their posters until the end of coffee break at 3:30 pm. Instructions for poster presenters can be found on the meeting website <http://goo.gl/j3qIb9>.

Les séances d'affiches libres et d'études de cas seront organisées dans le Grand Hall et des salles voisines. Les affiches des études de cas seront présentées lundi et les affiches libres, y compris les affiches d'étudiants des groupes de probabilité et de science actuarielle, seront présentées mardi. Les auteurs seront présent aux affiche jusqu'au fin de la pause-café à 15h30. Les présentateurs trouveront plus d'instructions sur le site Web du congrès. <http://goo.gl/WSDfxS>

Workshops • Ateliers

Sunday, May 25, 9:00 - 4:30 pm • Dimanche 25 mai, 9h00 - 16h30

Medical Sciences Building

Workshops organized by the sections will be held Sunday 9:00 am to 4:30 pm in the Medical Sciences Building at the University of Toronto. Come to the Stone Lobby <http://goo.gl/maps/tMzq5>. Coffee breaks will be held on the ground floor 10:20am-10:40am and 2:50pm-3:10pm. Lunch break is 12noon-1:30pm.

Les ateliers organisés par les sections auront lieu dimanche 9h00-16h30 dans le Medical Sciences Building de l'Université de Toronto. Rendez-vous au Stone Lobby <http://goo.gl/maps/tMzq5>. Les pauses-café auront lieu au rez de chaussée 10h20-10h40 et 14h50-15h10. La pause-déjeuner est de 12h-13h30.

Other Meetings • Autres reunions

SSC Annual General Meeting: Monday 5-6pm
Hart House East Common Room

Section AGM's: Tuesday 5-6pm
various rooms in Hart House and UC

Committee meetings: Most noontime business meetings will occur in HH and UC, with lunch for business meeting participants available in the Great Hall.

NSERC 2014 Results: Sunday 5-6
South Sitting Room, Hart House.

NSERC Application Workshop: Tuesday 5-6 in UC 256.
This workshop will be presented by NSERC Research Grants staff and will cover the Form 180 process, the Discovery Grant evaluation process principles, criteria and ratings), the Conference Model and tips for preparing a Discovery Grant application. Following the Workshop, there will be an opportunity for participants to ask questions.

Assemblée générale annuelle : lundi 17h-18h, Hart House East Common Room

AGM des groupes : mardi 17h-18h, plusieurs salles à Hart House et UC

Information concours 2014 CRSNG : dimanche 17h-18h Hart House South Sitting Room

Atelier subventions du CRSNG : mardi 17h-18h, UC 256.

Cet atelier, présenté par le personnel des subventions à la découverte du CRSNG, couvrira le processus du formulaire 180, le processus d'évaluation des subventions à la découverte (principes, critères et cotes), le modèle de conférence et des conseils pour la préparation d'une demande de subvention à la découverte. À la fin de l'atelier, les participants pourront poser des questions.

Social Events • Activités sociales

Welcome Reception • Réception de bienvenue

Sunday, May 25, 6:30 - 9:00 pm • Dimanche 25 mai, 18h30 - 21h00 Royal Ontario Museum

For those who like to kick their statistical meetings off in style, please join us at the opening reception at the Royal Ontario Museum on Sunday evening. This striking space features the newly added Michael Lee Chin crystal as well as carefully curated cultural and historical treasures. The Asian galleries on the ground floor will be open for our exclusive use. All conference participants are cordially invited to attend.

Pour ceux qui aiment commencer leurs réunions statistiques avec du style, il nous fera plaisir de vous recevoir à la réception d'ouverture au Musée royal de l'Ontario dimanche soir. Cet espace saisissant comporte le cristal Michael Lee Chin nouvellement ajoutée, ainsi que des trésors culturels et historiques soigneusement organisée. Les galeries asiatiques au rez de chaussée seront ouvertes pour notre usage exclusif. Tous les participants à la conférence sont cordialement invités à y assister.

Barbeque • Barbecue

Monday, May 26, 6:00 - 8:00 pm • Lundi 26 mai, 18h00 - 20h00 Hart House Great Hall

The student BBQ is open to both students and self-identified recent graduates. It is a great opportunity to catch up with old friends and make new ones! You can even win a door prize! The BBQ will take place in the Hart House Great Hall and/or courtyard. Hot dogs, burgers, and veggie burgers will be served and a cash bar will be available.

Le barbecue est ouvert aux étudiants et aux diplômés récents. Il s'agit d'une excellente occasion de retrouver des anciens amis et s'en faire des nouveaux! Vous pouvez même gagner un prix de présence! Le barbecue aura lieu à la Hart House Grand Hall et / ou dans la cour. Hot-dogs, des hamburgers et des hamburgers végétariens seront servis et un bar payant sera disponible.

Banquet

Tuesday, May 27, 6 pm • Mardi 27 mai, 18h Liberty Grand Ballroom, Exhibition Place

All conference participants are invited to Tuesday evening's banquet, which will be held at the Liberty Grand at Exhibition place.

Busses will leave from Hart House circle commencing at 6pm, with return trips operating regularly and as late as necessary. Public transit is also possible; take the subway Westbound on the Bloor-Danforth line to Dufferin station (make sure you pick up a transfer before you get on the subway), then find the Dufferin bus and travel south to the Dufferin Gates of Exhibition Place.

Tous les participants du congrès sont invités au banquet de mardi soir, qui aura lieu à la Liberty Grand, située à Exhibition Place. Les bus partiront de Hart House circle à partir de 18h00, avec des voyages retours qui fonctionnent régulièrement et assez tard. Le transport en commun est également possible; prenez le métro (TTC) direction ouest sur la ligne Bloor-Danforth jusqu'à la station Dufferin (assurez-vous que vous prenez un transfert avant de monter dans le métro), puis prenez le bus Dufferin vers le sud et démontez à Dufferin Gates of Exhibition Place.

Other Social Events • Autres événements sociaux

New investigators lunch Monday noon-1.30pm
Hart House Circle
Meet at the registration desk in the Great Hall.

COWIS dinner Monday at 7pm
Bedford Academy 36 Prince Arthur Ave
reservations under the name `SSC'

U of T alumni Monday 6:30pm-8:00pm
Hart House South Dining Room.

Repas des nouveaux chercheurs lundi midi-13h30
Hart House Circle
Rendez-vous au bureau d'inscription dans le Grand Hall.

Souper du Comité de promotion de la femme en statistique
lundi 19h, Bedford Academy 36 ave Prince Arthur
réservations sous le nom de `SSC'

Réception des anciens étudiants de l'Université de Toronto
lundi 18h30-20h
Hart House South Dining Room

Committees and Meetings • Comités et réunions

Saturday May 24

samedi 24 mai

18:00 - 23:30 Executive Committee 1 / Comité exécutif 1
Intercontinental Toronto Yorkville Hotel, Maifair room

Sunday May 25

dimanche 25 mai

9:30 - 12:00 Department Heads Meeting / Rencontre des directeurs de départements
South Sitting Room, Hart House

12:00 - 17:00 Board of Directors 1 / Conseil d'administration 1
South Dining Room, Hart House

Monday May 26

lundi 26 mai

12:00 - 13:30 Business and Industrial Statistics Section Executive Committee / Comité exécutif du groupe de statistique industrielle et de gestion
UC 55

12:00 - 13:30 Committee on SSC Bylaws / Comité de statuts de la SSC
UC 328

12:00 - 13:30 Actuarial Science Section Executive Committee 1 / Comité exécutif du groupe d'actuariat 1
UC 248

12:00 - 13:30 Biostatistics Section Executive Committee 1 / Comité exécutif du groupe de biostatistique 1
UC 255

12:00 - 13:30 Probability Section Executive Committee 1 / Comité exécutif du groupe de probabilité 1
UC 257

12:00 - 13:30 Research Committee / Comité de recherche
South Dining Room, Hart House

12:00 - 13:30 Statistical Education Section Executive Committee 1 / Comité exécutif du groupe d'éducation en statistique 1
UC 244

12:00 - 13:30 Strategic Planning Committee / Comité de la planification stratégique
UC 52

12:00 - 18:00 Award for Case Studies in Data Analysis Committee 1 / Comité du prix pour les études de cas et l'analyse de données 1
South Sitting Room, Hart House

17:00 - 18:00 SSC Annual General Meeting / Assemblée générale annuelle de la SSC
East Common Room, Hart House

Tuesday May 27

mardi 27 mai

12:00 - 13:30 CJS Editorial Board / Comité de rédaction de la RCS
UC 244

12:00 - 13:30 Biostatistics Section Executive Committee 2 / Comité exécutif du groupe de biostatistique 2
UC 53

12:00 - 13:30 Census at School Committee / Comité de recensement à l'école
UC 52

12:00 - 13:30 Office Committee / Comité du Bureau
UC 257

12:00 - 13:30 Public Relations Committee / Comité des relations publiques
UC 255

12:00 - 13:30	Statistical Education Committee / Comité d'éducation en statistique <i>UC 144</i>
12:00 - 13:30	Committee on New Investigators / Comité des nouveaux chercheurs <i>UC 248 (Rain location / en cas de pluie)</i>
12:00 - 13:30	Accreditation General Meeting / Assemblée générale des membres accrédités <i>UC 256</i>
12:00 - 18:00	Award for Case Studies in Data Analysis Committee 2 / Comité du prix pour les études de cas et l'analyse de données 2 <i>South Sitting Room, Hart House</i>
17:00 - 18:00	Business and Industrial Statistics Section General Meeting / Groupe de statistique industrielle et de gestion assemblée générale <i>Debates Room, Hart House</i>
17:00 - 18:00	Student Research Presentation Award Committee / Comité du prix pour la présentation de recherche étudiante <i>UC 152</i>
17:00 - 18:00	Actuarial Science Section General Meeting / Groupe d'actuariat assemblée générale <i>South Dining Room, Hart House</i>
17:00 - 18:00	Biostatistics Section General Meeting / Groupe de biostatistique assemblée générale <i>East Common Room, Hart House</i>
17:00 - 18:00	Probability Section General Meeting / Groupe de probabilité assemblée générale <i>Music Room, Hart House</i>
17:00 - 18:00	Statistical Education Section General Meeting / Groupe d'éducation en statistique assemblée générale <i>UC 144</i>
17:00 - 18:00	Survey Methods Section General Meeting / Groupe de méthodes d'enquête assemblée générale <i>UC 52</i>

Wednesday May 28**mercredi 28 mai**

12:00 - 13:30	Accreditation Committee / Comité d'accréditation <i>UC 256</i>
12:00 - 13:30	Actuarial Science Section Executive Committee 2 / Comité exécutif du groupe d'actuariat 2 <i>UC 248</i>
12:00 - 13:30	Probability Section Executive Committee 2 / Comité exécutif du groupe de probabilité 2 <i>UC 257</i>
12:00 - 13:30	Program Committee / Comité du programme <i>UC 53</i>
12:00 - 13:30	Statistical Education Section Executive Committee 2 / Comité exécutif du groupe d'éducation en statistique 2 <i>UC 244</i>
12:00 - 13:30	Student and Recent Graduate Committee / Comité des étudiants et diplômés récents <i>UC 257</i>
12:00 - 13:30	Treasurers / Trésoriers <i>UC 52</i>
16:45 - 18:00	Board of Directors 2 / Conseil d'administration 2 <i>UC - Croft Chapter House</i>
18:00 - 19:00	Executive Committee 2 / Comité exécutif 2 <i>UC - Croft Chapter House</i>

Social and Information Events • Événements sociaux et informatifs

Schedule for Coffee breaks Horaire pour Pauses-café

Sunday May 25

dimanche 25 mai

10:20 - 10:40	Coffee break / Pause-café <i>Stone Lobby, Medical Sciences Building</i>
14:50 - 15:10	Coffee break / Pause-café <i>Stone Lobby, Medical Sciences Building</i>

Monday May 26

lundi 26 mai

9:50 - 10:20	Coffee break / Pause-café <i>Great Hall</i>
15:00 - 15:30	Coffee break / Pause-café <i>Great Hall</i>

Tuesday May 27

mardi 27 mai

9:50 - 10:20	Coffee break / Pause-café <i>Great Hall</i>
15:00 - 15:30	Coffee break / Pause-café <i>Great Hall</i>

Wednesday May 28

mercredi 28 mai

9:50 - 10:20	Coffee break / Pause-café <i>Great Hall</i>
14:30 - 15:00	Coffee break / Pause-café <i>Great Hall</i>

Schedule for Information Events Horaire pour Événements informatifs

Sunday May 25

dimanche 25 mai

17:00 - 18:00	Information on 2014 NSERC Discovery Grant Competition / Information sur le concours 2014 des subventions à la découverte du CRSNG <i>South Sitting Room, Hart House</i>
---------------	--

Tuesday May 27

mardi 27 mai

17:00 - 18:00	NSERC Discovery Grant Application Workshop / Atelier sur les demandes de subventions à la découverte du CRSNG <i>UC 256</i>
---------------	--

Schedule for Registration Horaire pour Inscription

Sunday May 25 dimanche 25 mai

8:00 - 16:30 Registration / Inscription
Stone Lobby, Medical Sciences Building

18:30 - 19:30 Registration / Inscription
Royal Ontario Museum / Musée royal de l'Ontario

Monday May 26 lundi 26 mai

7:30 - 17:00 Registration / Inscription
Hart House Great Hall

Tuesday May 27 mardi 27 mai

8:15 - 17:00 Registration / Inscription
Hart House Great Hall

Schedule for Social events Horaire pour Événements Sociaux

Sunday May 25 dimanche 25 mai

18:30 - 21:00 Opening Reception / Réception de bienvenue
Royal Ontario Museum

Monday May 26 lundi 26 mai

13:00 - 14:30 New Investigators lunch / Repas des nouveaux chercheurs
Hart House Circle

18:00 - 20:00 U of T Alumni reception / Réception des anciens étudiants de l'Université de Toronto
South Dining Room, Hart House

18:00 - 20:00 Student BBQ / Barbecue étudiant
Great Hall

19:00 - 21:00 COWIS dinner / Souper du Comité de promotion de la femme en statistique
Bedford Academy under the name "SSC", 36 Prince Arthur Ave, Toronto

Tuesday May 27 mardi 27 mai

18:00 - 22:00 Banquet
Board busses in Hart House Circle

List of Sessions • Liste des sessions

- 1A-A1 Inaugural Session
Séance inaugurale
- 1A-A2 SSC Presidential Invited Address
Allocution de l'invité du président de la SSC
- 1B-I1 Big Data
Données volumineuses
- 1B-I2 Machine Learning
Apprentissage machine
- 1B-I3 Recent Developments in Finite Mixture Models
Récents progrès en modèles de mélanges finis
- 1B-I4 Statistical Genetic and Genomic Modeling and Analysis of Complex Traits
Modélisation génétique et génomique statistique et analyse des caractères complexes
- 1B-I5 Statistical Issues in Health Economics and Health Technology Assessment
Problèmes statistiques en économie de la santé et évaluation des technologies de la santé
- 1B-I6 Statistical Methods in Actuarial Science
Méthodes statistiques en science actuarielle
- 1B-C1 Biostatistics - New Estimation Methods
Biostatistique - nouvelles méthodes d'estimation
- 1B-C2 Probability Theory and Applications
Théorie des probabilités et applications
- 1B-C3 Statistical Inference and Applications 1
Inférence statistique et applications 1
- 1B-C4 Survey Methodology 2
Méthodologie d'enquête 2
- 1C-D1 Case Study 1
Étude de cas 1
- 1C-D2 Case Study 2
Étude de cas 2
- 1D-I1 Opportunities and Challenges in the First Years of an Academic Career (A Panel Discussion)
Possibilités et défis en début de carrière universitaire (table ronde)
- 1D-I2 Optimal Design of Experiments
Plans d'expérience optimaux
- 1D-I3 Recent Advances in Spatial Statistics
Récents progrès en statistique spatiale
- 1D-I4 Statistical Analysis with Data Complications
Analyse statistique et complications liées aux données
- 1D-I5 Stochastic Spatial Processes
Processus spatiaux stochastiques
- 1D-I6 Teaching Introductory Statistics with R
Enseigner la statistique de base avec R
- 1D-C1 Actuarial Science and Finance 1
Actuariat et finance 1
- 1D-C2 Biostatistics - Applications 1
Biostatistique - applications 1
- 1D-C3 Biostatistics - Bayesian Analysis
Biostatistique - analyse bayésienne
- 1D-C4 Statistical Inference and Applications 2
Inférence statistique et applications 2
- 1D-C5 Survey Methodology 1
Méthodologie d'enquête 1

- 1E-I1 Delivering a Successful Introductory Stats Service Course
Cours de service de base en statistique : meilleures pratiques
- 1E-I2 Financial Mathematics
Mathématique financière
- 1E-I3 New Developments in Nonlinear Time Series Modelling and Robust Methods in Time Series Analysis with Applications
Récents progrès en modélisation de séries chronologiques non linéaires; méthodes robustes d'analyse des séries chronologiques et applications
- 1E-I4 Publishing and Reproducible Research (A Panel Discussion)
Publication et reproductibilité de la recherche (table ronde)
- 1E-I5 Recent Developments in Small Area Estimation
Récents progrès en estimation pour les petits domaines
- 1E-I6 Survival Methods in Genetics
Méthodes de survie en génétique
- 1E-C1 Biostatistics - Applications 2
Biostatistique - applications 2
- 1E-C2 Biostatistics - Clinical Trials and Diagnostic Protocols
Biostatistique - essais cliniques et protocoles diagnostiques
- 1E-C3 Clustering, Classification and Dimension Reduction
Partitionnement, classification et réduction de dimension
- 1E-C4 Probability Theory and Estimation
Théorie des probabilités et estimation
- 1E-C5 Statistical Inference and Applications 5
Inférence statistique et applications 5
- 2A-A Gold Medal Address
Allocution du récipiendaire de la Médaille d'or
- 2B-I1 Actuarial Finance
Finance actuarielle
- 2B-I2 New Developments in High Dimensional Data Analysis
Récents progrès en analyse de données de grande dimension
- 2B-I3 New Directions in Causal Inference
Nouvelles directions en inférence causale
- 2B-I4 Semi and Unsupervised Learning
Apprentissage semi- et non supervisé
- 2B-I5 Shape-Constrained Maximum Likelihood: Methods and Applications
Maximum de vraisemblance avec contraintes de forme : méthodes et applications
- 2B-I6 Stochastic Processes and Their Applications
Processus stochastiques et leurs applications
- 2B-C1 Biostatistics - Measurement Errors and Prediction Errors
Biostatistique - erreurs de mesure et erreurs de prédiction
- 2B-C2 Empirical and Composite Likelihood and Higher Order Asymptotics
Vraisemblances empirique et composite et comportement asymptotique d'ordre supérieur
- 2B-C3 Nonregular Statistical Test Problems
Problèmes de tests statistiques non réguliers
- 2B-C4 Statistical Inference and Applications 3
Inférence statistique et applications 3
- 2C-P1 Actuarial Science Poster Session
Séance d'affiches en science actuarielle
- 2C-P2 Poster Session
Séance d'affichage
- 2C-P3 Probability Student Poster Session
Séance d'affiches étudiants en probabilité
- 2D-I1 David Binder Memorial Session
Séance commémorative David Binder

- 2D-I2 Gaussian Process Regression: Innovations and Applications
Régression par processus gaussien : innovations et applications
- 2D-I3 Graduate Student Presentations in Actuarial Science
Présentations d'étudiants de cycle supérieur en science actuarielle
- 2D-I4 New Methods for Analyzing Brain Imaging data
Nouvelles méthodes d'analyse des données d'imagerie du cerveau
- 2D-I5 Statistical Challenges in High-Throughput Genomic Data
Problèmes statistiques des données génomiques à débit élevé
- 2D-I6 Weakly Interacting Processes in Finance and Economics
Processus à faible interaction en finance et économie
- 2D-C1 Biostatistics - Applications 3
Biostatistique - applications 3
- 2D-C2 Biostatistics - Association Studies and Correlation Structure
Biostatistique - études d'association et structure de corrélation
- 2D-C3 Business and Industrial Statistics 1
Statistique industrielle et de gestion 1
- 2D-C4 Quantile, Spatial and Logistic Regression
Régression quantile, spatiale et logistique
- 2D-C5 Statistical Inference and Applications 4
Inférence statistique et applications 4
- 2E-I1 Bridging the Gap between Methods Development and Implementation in Biostatistics (A Panel Discussion)
Comment combler le fossé entre le développement et la mise en œuvre de méthodes en biostatistique (table ronde)
- 2E-I2 Challenges and New Developments in Complex Multiple Comparison Procedures
Défis et progrès récents pour les procédures de comparaisons multiples complexes
- 2E-I3 Copula Theory and Methods
Théorie et méthodes liées aux copules
- 2E-I4 Recent Advances in the Analysis of Complex Surveys
Récents progrès en analyse d'enquêtes complexes
- 2E-I5 Statistical Challenges in Modern Astronomy
Problèmes statistiques en astronomie moderne
- 2E-I6 Statistics of Sea Surface Winds and Non-Gaussianity in Climate Variability
Statistique des vents de surface de la mer et non-gaussianité de la variabilité climatique
- 2E-C1 Actuarial Science and Finance 2
Actuariat et finance 2
- 2E-C2 Biostatistics - New Inferential Strategies
Biostatistique - nouvelles stratégies inférentielles
- 2E-C3 Business and Industrial Statistics 2
Statistique industrielle et de gestion 2
- 2E-C4 Probability Theory and Finance
Théorie des probabilités et finance
- 2E-C5 Robust and Shrinkage Estimation
Estimation robuste et par rétrécissement
- 3A-A1 CRM-SSC Prize in Statistics Address
Allocution du récipiendaire du Prix CRM-SSC en statistique
- 3A-A2 SSC Impact Award Address
Allocution du récipiendaire du Prix pour impact de la SSC
- 3B-I1 Bioinformatics and Copula Models
Bioinformatique et modèles de copules
- 3B-I2 Modern Statistical Methods for Complex Health Data
Méthodes statistiques modernes pour données sanitaires complexes
- 3B-I3 New Challenges and Developments in Survival Analysis and its Applications
Nouveaux problèmes et progrès en analyse de survie et ses applications

- 3B-I4 New Developments in Statistical Neuroscience
Récents progrès en neuroscience statistique
- 3B-I5 Statistical Education in Light of Philosophy of Science (A Panel Discussion)
Philosophie de la science et éducation statistique
- 3B-I6 Stochastic Modelling in Industry
Modélisation stochastique en industrie
- 3B-C1 Actuarial Science and Finance 3
Actuariat et finance 3
- 3B-C2 Biostatistics - Survival, Longitudinal and Event History Analysis
Biostatistique - analyses de survie, longitudinales et d'historique d'événements
- 3B-C3 Some Specific Issues in Modelling and Estimation
Quelques problématiques spécifiques en modélisation et estimation
- 3C-A1 CJS Award Address
Allocution du récipiendaire du Prix de la RCS
- 3C-A2 Isobel Loutit Invited Address
Allocution invitée Isobel Loutit
- 3C-A3 Pierre Robillard Award Address
Allocution du récipiendaire du Prix Pierre-Robillard
- 3C-A4 Survey Methods Section Presidential Invited Address
Allocution de l'invité du président du Groupe des méthodes d'enquête
- 3D-I1 Bridging the Gap between Theory and Practice in Statistics
Comblent l'écart entre la théorie et la pratique en statistique
- 3D-I2 Business Analytics: Where does the Statistician Fit?
Méthodes analytiques de gestion : le rôle du statisticien
- 3D-I3 Evaluating the Effectiveness of our Introductory Courses
Comment évaluer l'efficacité de nos cours d'introduction?
- 3D-I4 Levy Processes and Applications
Processus de Lévy et leurs applications
- 3D-I5 Predictive Density Estimation
Estimation de densité prédictive
- 3D-I6 Statistical Analysis for Infectious Diseases
Analyse statistique des maladies infectieuses
- 3D-C1 Actuarial Science and Finance 4
Actuariat et finance 4
- 3D-C2 Biostatistics - Variable Selection
Biostatistique - sélection de variables

Program • Programme

Saturday May 24

samedi 24 mai

18:00 - 23:30

Intercontinental Toronto Yorkville Hotel, Maifair room

Executive Committee 1

Comité exécutif 1

Sunday May 25**dimanche 25 mai****8:00 - 16:30****Stone Lobby, Medical Sciences Building****Registration
Inscription****9:00 - 16:30****Fitzgerald Building 139****Applied Analysis of Survey Data
Analyse appliquée des données d'enquête**

Organizer and Chair / Responsable et présidente: Susana Rubin-Bleuer (Statistics Canada)
 Survey Methods Section
 Groupe de méthodes d'enquête

9:00 - 16:30 Steven Heeringa (University of Michigan)

9:00 - 16:30**Fitzgerald Building 129****Probabilistic Simulation via Distributed Algorithms
Simulation probabiliste via algorithmes distribués**

Organizer and Chair / Responsable et président: Michael Kouritzin (University of Alberta)
 Probability Section and Actuarial Science Section
 Groupe de probabilité et Groupe de science actuarielle

9:00 - 16:30 Hao Yu (Western University)

9:00 - 16:30**MedSci 3163****Spatial-Temporal Models in Epidemiology
Modèles spatio-temporels en épidémiologie**

Organizer and Chair / Responsable et présidente: Joan Hu (Simon Fraser University)
 Biostatistics Section
 Groupe de biostatistique

9:00 - 16:30 Gavin Shaddick (University of Bath)

9:00 - 16:30 Jim Zidek (University of British Columbia)

9:00 - 16:30**MedSci 2172****Statistical Learning Methods in Data Mining
Méthodes d'apprentissage statistique en forage de données**

Organizer and Chair / Responsable et président: Thomas Loughin (Simon Fraser University)
 Business and Industrial Statistics Section
 Groupe de statistique industrielle et de gestion

9:00 - 16:30 Hugh Chipman (Acadia University)

9:00 - 16:30 Rob Tibshirani (Stanford University)

9:00 - 16:30**MedSci 3153****Using WeBWork for On-line Homework in Statistics Courses
WeBWork pour les devoirs de statistique en ligne**

Organizer and Chair / Responsable et président: Kevin Thorpe (University of Toronto)
 Statistics Education Section
 Groupe d'éducation en statistique

9:00 - 16:30 Bruce Dunham (University of British Columbia)

9:00 - 16:30 Andy Leung (University of British Columbia)

9:30 - 12:00

South Sitting Room, Hart House

Department Heads Meeting

Rencontre des directeurs de départements

10:20 - 10:40

Stone Lobby, Medical Sciences Building

Coffee break

Pause-café

12:00 - 17:00

South Dining Room, Hart House

Board of Directors 1

Conseil d'administration 1

14:50 - 15:10

Stone Lobby, Medical Sciences Building

Coffee break

Pause-café

17:00 - 18:00

South Sitting Room, Hart House

Information on 2014 NSERC Discovery Grant Competition

Information sur le concours 2014 des subventions à la découverte du CRSNG

18:30 - 21:00

Royal Ontario Museum

Opening Reception

Réception de bienvenue

18:30 - 19:30

Royal Ontario Museum

Registration

Inscription

Monday May 26**lundi 26 mai****7:30 - 17:00****Hart House Great Hall****Registration
Inscription****8:30 - 8:45 Session 1A-A1****MedSci Auditorium****Inaugural Session
Séance inaugurale**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Mike Evans (University of Toronto)

8:30 - 8:35 Jamie Stafford (Chair of Local Arrangement)



8:35 - 8:40 Patrick Brown (Chair of Local Arrangement)

8:40 - 8:45 Mike Evans (SSC President)

8:45 - 9:50 Session 1A-A2 (abstract/résumé - page 72)**MedSci Auditorium****SSC Presidential Invited Address****Allocution de l'invité du président de la SSC**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Mike Evans (University of Toronto)

8:45 - 9:50 Michael Jordan (University of California Berkeley), Venkat Chandrasekaran (Caltech), John Duchi, Martin Wainwright & Yuchen Zhang (University of California, Berkeley)
On the Computational and Statistical Interface and "Big Data" / Statistique, informatique et l'inférence à grande échelle  



9:50 - 10:20**Great Hall****Coffee break
Pause-café****10:20 - 11:50 Session 1B-I1** (abstracts/résumés - page 73)**Hart House - Debates Rm****Big Data****Données volumineuses**



Invited Paper Session / Présentations sur invitation



Organizer and Chair / Responsable et présidente: Susana Rubin-Bleuer (Statistics Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

10:20 - 10:50 Hugh Chipman (Acadia University), Robert McCulloch (University of Chicago), Matt Pratola (Ohio State University), Dave Higdon (Los Alamos National Laboratory), Steve Scott (Google), Edward George (University of Pennsylvania)
Ensemble Learning for Big Data / Apprentissage par ensembles pour données volumineuses  

10:50 - 11:20 Steven Heeringa (University of Michigan), Roberto Edmundo Mellipilan (University of Michigan, Program in Survey Methods)
Survey Research and Big Data: Estrangement or a Long-term Relationship? / Recherche-enquête et données volumineuses : éloignement ou relation à long terme ?  

11:20 - 11:50 Jean Pignal (Statistics Canada), Ferhana Ansari (), Lily Ma (Statistics Canada)
Big Data and Official Statistics / Les données volumineuses et les statistiques officielles  

10:20 - 11:50 **Session 1B-I2** (abstracts/résumés - page 75) **Hart House - East Common**

Machine Learning**Apprentissage machine**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Ruslan Salakhutdinov (University of Toronto)







- 10:20 - 10:50 Ali Ghodsi, Michael Biggs, Stephen Vavasis & Tameem Hesham (University of Waterloo)
Nonnegative Matrix Factorization via Rank-One Downdate and its Application in Sum-Product Networks Structure Learning / Factorisation de matrices non négatives par l'algorithme de recul de rang un et ses applications dans l'apprentissage de la structure de réseaux sommes et produits  
- 10:50 - 11:20 Peter Orbanz (Columbia University)
Random Walk Models of Graphs and Networks / Modèles de marche aléatoire de graphes et de réseaux  
- 11:20 - 11:50 Ruslan Salakhutdinov & Nitish Srivastava (University of Toronto)
Multimodal Learning with Deep Boltzmann Machines / Apprentissage multimodal profond à l'aide de machines de Boltzmann  

10:20 - 11:50 **Session 1B-I3** (abstracts/résumés - page 77) **Hart House - Music Rm**

Recent Developments in Finite Mixture Models**Récents progrès en modèles de mélanges finis**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Pengfei Li (University of Waterloo)



- 10:20 - 10:50 Jiahua Chen (University of British Columbia), Pengfei Li (University of Waterloo), Yukun Liu (East China Normal University)
Sample Size Calculation for Homogeneity Test / Calcul de la taille de l'échantillon pour un test d'homogénéité  
- 10:50 - 11:20 Hajo Holzmann (Philipp University of Marburg, Germany), Florian Schwaiger (Marburg University, Germany)
Hidden Markov Models with State-Dependent Mixtures: Minimal Representation, Model Testing and Applications to Clustering / Les modèles à chaîne de Markov cachée avec des mélanges dépendants de l'état : représentation minimale, test du modèle et application à l'analyse de regroupement.  
- 11:20 - 11:50 Hiroyuki Kasahara (University of British Columbia), Katsumi Shimotsu (University of Tokyo)
Testing the Number of Components in Heteroscedastic Normal Mixtures / Test du nombre de composantes dans des mélanges normaux hétéroscédastiques  

10:20 - 11:50 **Session 1B-I4** (abstracts/résumés - page 79) **UC - 140**

Statistical Genetic and Genomic Modeling and Analysis of Complex Traits**Modélisation génétique et génomique statistique et analyse des caractères complexes**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Brad McNeney (Simon Fraser University)

- 10:20 - 10:50 Joseph Beyene (McMaster University)
Integrative Analysis of High-Dimensional Biological Data / Analyse intégrative des données biologiques de grande dimension  

- 10:50 - 11:20 Angelo Canty (McMaster University), Nabin Srestha (McMaster University), Marie-Pierre Sylvestre (Université de Montréal), John Brunzell (University of Washington), Andrew Boright (University of Toronto), Andrew Paterson (Hospital for Sick Children, Toronto), Shelley Bull (Mount Sinai Hospital, Toronto)
GWAS for Lipoprotein Profiles Using an Empirically Fitted Null Distribution / Études d'association pangénomique pour les profils de lipoprotéines à l'aide d'une distribution nulle ajustée empiriquement  
- 11:20 - 11:50 Jo Knight (University of Toronto), Sarah Gagliano (Centre for Addiction and Mental Health), Michael Barnes (Queen Mary University of London), Michael Weale (King's College London)
Data Integration to Boost Association Signals / Intégration de données pour amplifier des signaux d'association  

10:20 - 11:50 Session 1B-I5 (abstracts/résumés - page 81) **UC - 163**







Statistical Issues in Health Economics and Health Technology Assessment

Problèmes statistiques en économie de la santé et évaluation des technologies de la santé

Invited Paper Session / Présentations sur invitation

Chair/Président: Nickolas Mitsakakis (University of Toronto)

Organizers/Responsables: Nickolas Mitsakakis (University of Toronto) and/et Eleanor Pullenayegum (University of Toronto & Hospital for Sick Children)

- 10:20 - 10:50 Heejung Bang (University of California at Davis)
Analysis of "Censored" Medical Costs for the Last 15 Years: Something Old, Something New and Something Resurrected / Analyse des coûts médicaux «censurés» pour les 15 dernières années : du vieux, du neuf et du ressuscité  
- 10:50 - 11:20 Kristian Thorlund (McMaster University)
Integration of Network Meta-Analysis and Health Economic Evaluations -- Reducing Uncertainty and Avoiding Biases / Intégration de la méta-analyse en réseau et des évaluations économiques sur la santé afin de réduire l'incertitude et d'éviter les biais  
- 11:20 - 11:50 Andy Willan (University of Toronto)
Should Stochastic Uncertainty Play a Role in the Pricing of New Health Care Interventions? / L'incertitude stochastique devrait-elle jouer un rôle dans l'établissement du prix des nouvelles interventions en matière de soins de santé ?  

10:20 - 11:50 Session 1B-I6 (abstracts/résumés - page 83) **UC - 179**

Statistical Methods in Actuarial Science







Méthodes statistiques en science actuarielle

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Yi Lu (Simon Fraser University)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe de science actuarielle

- 10:20 - 10:50 Arthur Charpentier (Université du Québec à Montréal), Emmanuel Flachaire (Université d'Aix-Marseille, GREQAM-EHESS,)
Quantiles Estimation from Heavy-Tailed Distribution / Estimation de quantiles à partir de distributions à ailes relevées  
- 10:50 - 11:20 Edward Furman & Jianxi Su (York University), Ricardas Zitikis (University of Western Ontario)
Background Risk Models: Theory and Applications / Modèles de risque rémanent : théorie et applications  
- 11:20 - 11:50 Natalia Nolde (University of British Columbia)
The Effect of Aggregation on Extremes from Asymptotically Independent Light-Tailed Risks / Effet de l'agrégation sur les valeurs extrêmes de distributions de risques à ailes plates asymptotiquement indépendantes  













10:20 - 11:50 **Session 1B-C1** (abstracts/résumés - page 85) **UC - 161**

Biostatistics - New Estimation Methods

Biostatistique - nouvelles méthodes d'estimation

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Lei Sun (University of Toronto)

- 10:20 - 10:35 Christian Boudreau (University of Waterloo), Hugo Varet (Université Paris Descartes), Kuan Liu (London Health Sciences Centre), Sandrine Katsahian (Hôpital Henri Mondor)
Frailty Models for Competing Risks / Modèles à effets aléatoires pour risques compétitifs  
- 10:35 - 10:50 Kun Liang (University of Waterloo)
False Discovery Rate Estimation for Large-scale Homogeneous Discrete P-values / Estimation du taux de fausses découvertes pour des valeurs p discrètes homogènes à grande échelle  
- 10:50 - 11:05 Lasantha Premarathna (Simon Fraser University), Carl James Schwarz (Simon Fraser University.)
Partial Stratification in Two-Sample Capture-Recapture Experiments / Stratification partielle dans les expériences de capture-recapture à deux échantillons  
- 11:05 - 11:20 Krishna Saha (Central CT State University)
Statistical Analysis of Means and Dispersions from Several Treatment Groups for Extra-Dispersed Count Data from Toxicology Studies / Analyse statistique des moyennes et dispersions de plusieurs groupes de traitement pour des données de dénombrement surdispersées tirées d'études en toxicologie  
- 11:20 - 11:35 Doneal Thomas & Andrea Benedetti (McGill University)
Estimating Generalized Linear Mixed Models with Binary Outcomes for the Analysis of Individual Patient Data Meta-Analyses: Which method is best? / Estimation des modèles mixtes linéaires généralisés avec résultats binaires pour l'analyse de méta-analyses sur données individuelles : quelle est la meilleure méthode ?  
- 11:35 - 11:50 Saman Muthukumarana (University of Manitoba), Michael Evans (University of Toronto)
Bayesian Hypothesis Assessment in Two-arm Trials Using Relative Belief Ratios / Évaluation bayésienne d'hypothèses utilisant les rapports de confiance relative dans des essais à deux groupes  







10:20 - 11:35 **Session 1B-C2** (abstracts/résumés - page 89) **UC - A101**





Probability Theory and Applications

Théorie des probabilités et applications

Contributed Paper Session / Séance de communications libres

Chair/Président: Serge Provost (Western University)

- 10:20 - 10:35 Farid El Ktaibi & Gail Ivanoff (University of Ottawa)
Blockwise Bootstrap for Empirical Processes and Change-Point Detection / Méthodes de rééchantillonnage pour les processus empiriques et détection de changement de structure dans la loi marginale  
- 10:35 - 10:50 Samira Sadeghi (University of Alberta)
Outer-product of Linear Processes with Heavy-tail and Long-range Dependence / Produit extérieur de processus linéaires à ailes lourdes et dépendance à long terme  
- 10:50 - 11:05 Ibrahim Abdelrazeq, Gail Ivanoff & Rafal Kulik (University of Ottawa)
Goodness of Fit Test: Recovered Noise for CAR(1) Processes / Test d'adéquation : récupération du bruit dans des processus CAR(1)  

- 11:05 - 11:20 Sean Jewell, Neil Spencer & Alexandre Bouchard-Côté (University of British Columbia)
Modelling Graffiti Prevalence in Metro Vancouver under a Hierarchical Bayesian Non-parametric Framework / Modélisation de la prévalence de graffitis dans la région métropolitaine de Vancouver dans un contexte bayésien hiérarchique non paramétrique  
- 11:20 - 11:35 Simon Mak (Georgia Institute of Technology), Derek Bingham & Yi Lu (Simon Fraser University)
A Spatial Compound Poisson Process for Hurricane Damage / Processus spatial de Poisson composé pour évaluer les dommages résultant d'un ouragan  

10:20 - 11:50 Session 1B-C3 (abstracts/résumés - page 92) **UC - Croft House**

Statistical Inference and Applications 1
Inférence statistique et applications 1



Contributed Paper Session / Séance de communications libres
Chair/Président: Tolulope Sajobi (University of Calgary)

- 10:20 - 10:35 Malak ALmutairi & Jianan Peng (Acadia University)
Simultaneous Inference for Low Dose Risk Estimation with Abbott-adjusted Log-Logistic Response Model Benchmark Analysis / Inférence simultanée pour l'estimation du risque à faible dose avec le modèle log-logistique ajusté selon la formule d'Abbott  
- 10:35 - 10:50 Marti Anderson (Massey University, Auckland, New Zealand)
A Comparison of Permutation, Bootstrap and Monte Carlo Techniques Applied to the Multivariate Behrens-Fisher Problem: Behaviour and Consequences for Statistical Inference / Comparaison des techniques de permutation, bootstrap et Monte Carlo appliquées à un problème multivarié de Behrens-Fisher : comportement et conséquences pour l'inférence statistique  
- 10:50 - 11:05 Mohamed Belalia & Taoufik Bouezmarni (Université de Sherbrooke), Alexandre Leblanc (University of Manitoba)
Bernstein Polynomials for Estimating a Conditional Distribution Function / Estimation de la fonction de répartition conditionnelle par les polynômes de Bernstein.  
- 11:05 - 11:20 Jinjun Tong (Memorial University of NL), J Concepción Loredó-Osti (Memorial University of Newfoundland)
Estimation of the State-Space Model with Stable Noise / Estimation d'un modèle à espace d'états avec un bruit stable  
- 11:20 - 11:35 Seong-Hwan Jun (University of British Columbia)
A Beginner's Perspective on Causal Inference: Potential Outcomes Framework, Graphical Models and Structural Causal Model / Perspective d'un novice sur l'inférence causale : cadre pour les résultats potentiels, modèles graphiques et modèle causal structurel  
- 11:35 - 11:50 Lucas Thorpe & Mei Ling Huang (Brock University)
On Nonparametric Methods of Extreme Value Distributions / Méthodes non paramétriques de distributions de valeurs extrêmes  

10:20 - 11:50 Session 1B-C4 (abstracts/résumés - page 95) **Hart House - South Dining**

Survey Methodology 2
Méthodologie d'enquête 2

Contributed Paper Session / Séance de communications libres
Chair/Président: David Haziza (University of Montreal)

- 10:20 - 10:35 Tounkara Fode & Rivest Louis-Paul (Université Laval)
Copula Model-based Approach and Empirical Bayes Techniques for Small Area Proportions / Approche basée sur le modèle de copules et les techniques empiriques bayésiennes sur les proportions pour les petits domaines  

- 10:35 - 10:50 Naima Gouzi (Statistique Canada)
Integrating Establishment Characteristics When Calibrating a Business Survey / Intégrer les caractéristiques des établissements lors du calage d'une enquête-entreprise  
- 10:50 - 11:05 Amanda Halladay (Statistics Canada)
Methodological Challenges with the 2015 Canadian Community Health Survey - Nutrition / Défis méthodologiques du volet de nutrition de l'Enquête sur la santé dans les collectivités canadiennes de 2015  
- 11:05 - 11:20 Eric Pelletier & Pierre-Olivier Julien (Statistique Canada)
The Volatility and the Rarity of Scientific Research and Development in Canada -- How to "Search" for It? / La volatilité et la rareté de la recherche et développement scientifique au Canada -- Comment la « chercher » ?  
- 11:20 - 11:35 Michael Rotondi (York University)
Towards the Estimation of Effect Measures in Studies Using Respondent-Driven Sampling / Vers une estimation des mesures d'effet dans les études avec échantillonnage fondé sur les répondants (EFR)  
- 11:35 - 11:50 Kyle Vincent (Bank of Canada), Saman Muthukumarana & Jenna Tichon (University of Manitoba)
Bayesian Item Response Analysis of Method-of-Payment Habits in Banking Surveys / Analyse bayésienne des réponses par items dans les sondages bancaires sur les habitudes de moyens de paiement  

12:00 - 13:30**UC 248****Actuarial Science Section Executive Committee 1
Comité exécutif du groupe d'actuariat 1**

12:00 - 18:00**South Sitting Room, Hart House****Award for Case Studies in Data Analysis Committee 1
Comité du prix pour les études de cas et l'analyse de données 1**

12:00 - 13:30**UC 255****Biostatistics Section Executive Committee 1
Comité exécutif du groupe de biostatistique 1**

12:00 - 13:30**UC 55****Business and Industrial Statistics Section Executive Committee
Comité exécutif du groupe de statistique industrielle et de gestion**

12:00 - 13:30**UC 328****Committee on SSC Bylaws
Comité de statuts de la SSC**

12:00 - 13:30**UC 257****Probability Section Executive Committee 1
Comité exécutif du groupe de probabilité 1**

12:00 - 13:30**South Dining Room, Hart House****Research Committee
Comité de recherche**

12:00 - 13:30 **UC 244**

Statistical Education Section Executive Committee 1
Comité exécutif du groupe d'éducation en statistique 1

12:00 - 13:30 **UC 52**

Strategic Planning Committee
Comité de la planification stratégique

12:00 - 17:30 **Session 1C-D1** **Great Hall**

Case Study 1
Étude de cas 1

Case Study / Étude de cas

Organizer and Chair / Responsable et président: Georges Monette (York University)

- 12:00 - 17:30 Dylan Moore, Dayeol Choi, Christopher Balette (University of Toronto)
- 12:00 - 17:30 Maxime Turgeon, Kevin Mcgregor, Sahir Bhatnagar (McGill University)
- 12:00 - 17:30 Chunlin Wang, Peter Raboud, Lin Chen, Allison Devlin, Marco Sung, Nathan Wilson, Massey Cashore (University of Waterloo)
- 12:00 - 17:30 Tingting Yu, Seong-Hwan Jun (University of British Columbia)
- 12:00 - 17:30 Shariah Shams, Tim To, Ruoyong Xu, Changchang Xu, Lei Miao, Gabriel Lau (University of Toronto)
- 12:00 - 17:30 Shivon Sue-Chee, Lingling Fan, Wei Lin (University of Toronto)
- 12:00 - 17:30 Armend Lokku, Nadia Muhe (University of Toronto)
- 12:00 - 17:30 Elena Szefer, Abdollah Safari (Simon Fraser University)
- 12:00 - 17:30 Justine Zehr, Steve Fereira, Bertrand Saulnier (Université de Montréal)
- 12:00 - 17:30 Osvaldo Espin-Garcia, Simon Greaves, Sepehr Jafar-Pour, Yitian Tao, Kate Tsiplova, Coralie Wong, Shary (Xinyu) Zhang (University of Toronto)
- 12:00 - 17:30 Teresa Chan, Farhood Mohammadi, Pavel Petrov, Samira Sadeghi, Xiaozhou Wang (University of Alberta)
- 12:00 - 17:30 Karl Knopf, Yuanhao Lai, Jiaoying Li, Kexin Luo, Shengchao Yan, Wei Yue (University of Western Ontario)
- 12:00 - 17:30 Qing Xu, Pingchuan Cao, Mengnan Zhu (University of Toronto)

12:00 - 17:30 **Session 1C-D2** **Great Hall**

Case Study 2
Étude de cas 2

Case Study / Étude de cas

Organizer and Chair / Responsable et président: Georges Monette (York University)

- 12:00 - 17:30 Di Wang, Namdar Homayounfar, Peter Lee, Serena Lu, Shiva Ashta (University of Toronto)
- 12:00 - 17:30 Lihu Liu, Moyan Mei, Maria Reyes, Michael Butler, Yun Cai (Dalhousie University)
- 12:00 - 17:30 Shamika Thirimanna, Mi Shen, Pawel Morzywolek, Daniel Rodriguez Duque (University of Waterloo)
- 12:00 - 17:30 Yu Luo, Zhe Li, Ya-Ting Chang, Shuanglu Liu, Bo Hu (University of Waterloo)
- 12:00 - 17:30 Shariah Shams, Tim To, Ruoyong Xu Changchang Xu, Lei Miao, Gabriel Lau (University of Toronto)
- 12:00 - 17:30 Fei Huang, Yuying Xie, Di Shu, Ken Xiao, Xin Liu (University of Western Ontario)
- 12:00 - 17:30 Simon Greaves, Xinyu Zhang, Yao Song, Yicen Guo, Xi Liu (University of Toronto)

12:00 - 17:30	Lauren Erdman, Jennifer Gillis, Naim Panjwani, Osvaldo Espin-Garcia, Armend Lokku (University of Toronto)
12:00 - 17:30	Ahsan Bhatti, Chengwei Qin, Hedan Han, Keya Biswas, Regina Kampo, Sandip Barui, Peter Macdonald (McMaster University)
12:00 - 17:30	Hervé Mensah, Imad Yanni, Ghizlene Zerghini, Jean-Phillipe Mallette (HEC Montréal)
12:00 - 17:30	Adam Rahman (University of Waterloo)
12:00 - 17:30	Xiaozhou Wang, Sile Tao, Qian Shi (University of Alberta)
12:00 - 17:30	Andy Koh, Gabriela Gonzalez, Jordan Collins, John Jarvis, Sriskondarajah Sujanthan (York University)
12:00 - 17:30	Nadia Muhe, Coralie Wong, Yitian F. Tao (University of Toronto)
12:00 - 17:30	Yumi Kondo, Md Mahsin, Kyle Hambrook, Yanling (Tara) Cai (University of British Columbia)
12:00 - 17:30	Xingliang Huang, Honglei Xie, Ning Zhang (University of Waterloo)

13:00 - 14:30**Hart House Circle****New Investigators lunch****Repas des nouveaux chercheurs****13:30 - 15:00 Session 1D-I1 (description - page 98) UC - 163****Opportunities and Challenges in the First Years of an Academic Career (A Panel Discussion)****Possibilités et défis en début de carrière universitaire (table ronde)**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Anne-Sophie Charest (Laval University)

Session sponsored by the New Investigator Committee

Séance parrainée par le Comité des nouveaux chercheurs

13:30 - 15:00 Masoud Asgharian (McGill University)

13:30 - 15:00 Hanna Jankowski (York University)

13:30 - 15:00 Pengfei Li (University of Waterloo)

13:30 - 15:00 Session 1D-I2 (abstracts/résumés - page 99) UC - 140**Optimal Design of Experiments****Plans d'expérience optimaux**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Julie Zhou (University of Victoria)

13:30 - 14:00	Saumen Mandal & Monsur Chowdhury (University of Manitoba) Optimal Experimental Design for Maximum Likelihood Estimation / Plan d'expérience optimal pour l'estimation du maximum de vraisemblance  
14:00 - 14:30	Peter Qian (University of Wisconsin-Madison) Samurai Sudoku-Based Space-Filling Designs / Plans de grilles Samurai inspirés du Sudoku  
14:30 - 15:00	Weng Kee Wong (University of California Los Angeles), Jiaheng Qiu (UCLA), Ray-Bing Chen (National Cheng Kung University), Weichung Wang (National Taiwan University) Using Animal Instincts to Find Efficient Experimental Designs / Utilisation de l'instinct animal pour trouver des plans d'expériences efficaces  

13:30 - 15:00 Session 1D-I3 (abstracts/résumés - page 101) **Hart House - Debates Rm**
Recent Advances in Spatial Statistics
Récents progrès en statistique spatiale

Invited Paper Session / Présentations sur invitation

Chair/Président: Ye Li (University of Toronto)

Organizer/Responsable: Mahmoud Torabi (University of Manitoba)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 13:30 - 14:00 Ying MacNab (University of British Columbia)
 Bayesian Hierarchical Spatial Mixture for Modeling Sparse Data on a Discontinuous and Mod-
 estly Large Irregular Lattice / Modélisation bayésienne hiérarchique de mélanges pour don-
 nées clairsemées sur un treillis irrégulier discontinu d'envergure modeste E E
- 14:00 - 14:30 Rhonda Rosychuk (University of Alberta), Hsing-Ming Chang (University of British Columbia
 Okanagan)
 A Spatial Scan Statistic for Compound Poisson Data, Using Negative Binomial Distribution
 and Accounting for Population Stratification / Statistique de balayage spatial pour des don-
 nées de Poisson composées selon une distribution binomiale négative en tenant compte de
 la stratification de la population E E
- 14:30 - 15:00 Mahmoud Torabi & Mahmoud Torabi (University of Manitoba)
 Spatial Generalized Linear Mixed Models with Multivariate CAR Models for Areal Data / Mod-
 èles linéaires mixtes généralisés spatiaux avec modèles autorégressifs conditionnels multi-
 variés pour des données géospatiales E E

13:30 - 15:00 Session 1D-I4 (abstracts/résumés - page 103) **Hart House - East Common**
Statistical Analysis with Data Complications
Analyse statistique et complications liées aux données

Invited Paper Session / Présentations sur invitation

Chair/Présidente: Wei Liu (York University)

Organizer/Responsable: Lang Wu (University of British Columbia)

- 13:30 - 14:00 Joan Hu (Simon Fraser University)
 Challenges and Strategies for Statistical Inference with Administrative Databases / Défis et
 stratégies en matière d'inférence statistique sur des bases de données administratives E E
- 14:00 - 14:30 Sanjoy Sinha (Carleton University), Abdus Sattar (Case Western Reserve University)
 A Robust Approach to Analyzing Longitudinal Data with Informative Dropouts and Outliers
 / Approche robuste pour l'analyse de données longitudinales comportant des cessations de
 suivi et des données aberrantes informatives E E
- 14:30 - 15:00 Leilei Zeng, Lan Wen & Richard Cook (University of Waterloo)
 Model Misspecification in Progression-Free Survival Analysis Due to Intermittent Assessment
 / Erreurs de spécification de modèles dues à une évaluation intermittente en analyse de survie
 sans progression E E







13:30 - 15:00 Session 1D-I5 (abstracts/résumés - page 105) **UC - 179**
Stochastic Spatial Processes
Processus spatiaux stochastiques

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Tom Salisbury (York University)

Session sponsored by the Probability Section





Séance parrainée par le Groupe de probabilité

- 13:30 - 14:00 Louigi Addario-Berry (McGill University)
Growing Random Maps / Comment s'enfuir de l'Hexagone  
- 14:00 - 14:30 Deniz Sezer (University of Calgary)
Exit Densities and the Martin Boundary of Super-Brownian Motion / Densités de passage et frontière de Martin du supermouvement brownien  
- 14:30 - 15:00 Janosch Ortmann & Jeremy Quastel (University of Toronto), Daniel Remenik (Universidad de Chile)
Exact Formulas for Random Growth off a Flat Interface / Formules exactes pour la croissance aléatoire à partir d'une interface plane  

13:30 - 15:00 Session 1D-I6 (abstracts/résumés - page 107) **UC - 161**

Teaching Introductory Statistics with R
Enseigner la statistique de base avec R









Invited Paper Session / Présentations sur invitation
Chair/Président: Duncan Murdoch (Western University)
Organizer/Responsable: John Braun (Western University)
Session sponsored by the Statistics Education Section
Séance parrainée par le Groupe d'éducation en statistique





- 13:30 - 14:00 Joel Dubin (University of Waterloo)
Teaching R to Introductory Biostatistics Students in a Professional Distance-Education Masters Program / Enseignement de R dans un cours d'introduction à la biostatistique aux étudiants d'un programme de maîtrise professionnelle à distance  
- 14:00 - 14:30 Kevin Thorpe (University of Toronto)
Using R in Introductory Biostatistics Courses / Utilisation du logiciel R dans des cours d'introduction à la biostatistique  
- 14:30 - 15:00 Alison Gibbs, Discussant (University of Toronto)

13:30 - 15:00 Session 1D-C1 (abstracts/résumés - page 109) **UC - A101**

Actuarial Science and Finance 1
Actuariat et finance 1

Contributed Paper Session / Séance de communications libres
Chair/Président: Paul Marriott (University of Waterloo)

- 13:30 - 13:45 Mathieu Boudreault (Université du Québec à Montréal), Geneviève Gauthier (HEC Montréal), Tommy Thomassin (Standard Life)
Estimation of Asset Correlations with Transient Shocks in a Multi-name Credit Risk Model / Estimation de corrélations d'actifs avec des chocs transitoires dans un modèle sans-nom de risque de crédit  
- 13:45 - 14:00 Marie-Pier Côté & Christian Genest (McGill University)
Copula-based Model for Risk Aggregation / Modèle d'agrégation de risques basé sur des copules  
- 14:00 - 14:15 Hassan Omid Firouzi (University of Montreal), Andrew Luong (Laval University)
Optimal Portfolio Problem for Entropic Value at Risk: When the Underlying Distribution is Non Elliptical / Problème de portefeuille optimal pour la valeur à risque entropique : quand la distribution sous-jacente est non elliptique  
- 14:15 - 14:30 Etienne Marceau, Hélène Cossette, Etienne Larrivée-Hardy & Julien Trufinh (Université Laval)
Infinite-time Ruin Measures for Compound Renewal Risk Models with Dependence / Mesures de ruine pour les modèles de risque basés sur les processus de renouvellement composés avec dépendance  

- 14:30 - 14:45 Shanoja Naik & Taehan Bae (University of Regina)
Calibration of Wishart Stochastic Correlation Model for Sovereign Default Risk / Calage du modèle de corrélation stochastique de Wishart pour le risque de faillite souveraine  
- 14:45 - 15:00 Chengguo Weng (University of Waterloo)
CPPI under Regime Switching and Transaction Cost / CPPI en cas de changement de régime et coûts de transaction  

13:30 - 15:00 Session 1D-C2 (abstracts/résumés - page 112) **Hart House - Music Rm**

Biostatistics - Applications 1

Biostatistique - applications 1

Contributed Paper Session / Séance de communications libres

Chair/Président: Joseph Beyene (McMaster University)

- 13:30 - 13:45 Farnoosh Abbas Aghababazadeh, Mayer Alvo & David Bickel (University of Ottawa)
Optimal Reference Set for Estimating Local False Discovery Rate / Ensemble de référence optimal pour l'estimation du taux local de fausses découvertes  
- 13:45 - 14:00 Kian Jahromi & Dianliang Deng (University of Regina)
Stable Trend Detection for Gene Expression Under Multiple Biological Conditions / Détection d'une tendance stable pour l'expression génique dans des conditions biologiques multiples.  
- 14:00 - 14:15 Vanda Milheiro Lourenço (FCT and CMA - NOVA University of Lisbon), Paulo C. Rodrigues (Federal University of Bahia), Ana M. Pires (IST - Universidade Técnica de Lisboa)
A Robust Coefficient of Determination for Heritability Estimation in Genetic Association Studies / Coefficient robuste de détermination pour l'estimation de l'héritabilité dans les études d'association génétique  
- 14:15 - 14:30 Narges Nazeri Rad & Jerald F Lawless (University of Waterloo)
Estimation of Prevalence Probabilities and Assessment of Multistate Models / Estimation des probabilités de prévalence et évaluation des modèles multi-états  
- 14:30 - 14:45 Marylène Troupé & Jean Vaillant (Université des Antilles et de la Guyane), Léonide Célini (Université de Paris Est - Créteil)
Loglinear Models in the Presence of Overdispersion with Application to Aphis Gossypii Count Data / Modèles loglinéaires en présence de surdispersion. Application aux données de comptage d'Aphis Gossypii.  
- 14:45 - 15:00 Yishu Wang & Michal Abrahamowicz (McGill University)
Assessing the Impact of Exposure Measurement Errors on the Results of Time-to-Event Analyses of Adverse Effects of Medications: a Simulation Study / Évaluation de l'impact des erreurs de mesures de l'exposition sur les résultats des analyses du temps avant événement des effets négatifs de médicaments : une étude par simulation  



13:30 - 15:00 Session 1D-C3 (abstracts/résumés - page 116) **UC - Croft House**











Biostatistics - Bayesian Analysis

Biostatistique - analyse bayésienne

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Kristian Thorlund (McMaster University)

- 13:30 - 13:45 Ahalya Sivathayalan & Patrick Farrell (Carleton University), Daniel Krewski (University of Ottawa)
Introduction and Assessment of a New PoD for Human Exposure Guidelines / Introduction et évaluation d'un nouveau point de départ pour les recommandations sur l'exposition humaine  

- 13:45 - 14:00 Sayantee Jana, Jemila Hamid & Narayanaswamy Balakrishnan (McMaster University)
A Bayesian Growth Curve Model for High Dimensional Data / Modèle bayésien de courbe de croissance pour des données de grande dimension  
- 14:00 - 14:15 Sarah Gagliano & Jo Knight (Centre for Addiction and Mental Health), Michael Weale (King's College London), Michael Barnes (Queen Mary University of London)
Investigation of Predictive Accuracy Measures for Genetic Models / Exploration des mesures de précision des modèles génétiques  
- 14:15 - 14:30 Khader Khadraoui (Université Laval)
Nonparametric Bayesian Regression under Combinations of Local Shape Constraints / Régression bayésienne non paramétrique sous combinaisons de contraintes locales de forme  
- 14:30 - 14:45 Elaheh Torkashvand, Mohammad Jafari Jozani & Mahmoud Torabi (University of Manitoba)
Pseudo-Empirical Bayes Estimation of Small Area Means Based on the James-Stein Estimation in Linear Regression Models with Functional Measurement Error / Estimation bayésienne pseudo-empirique des moyennes pour petites régions fondée sur l'estimation de James-Stein dans les modèles de régression linéaire avec erreur de mesure fonctionnelle  
- 14:45 - 15:00 Maria Esther Perez Trejo & Robert Platt (McGill University)
Bayesian Modeling of Clustered Measured Binary Outcomes / Modélisation bayésienne des résultats binaires mesurés regroupés  










13:30 - 15:00 **Session 1D-C4** (abstracts/résumés - page 120) **UC - 144**

Statistical Inference and Applications 2

Inférence statistique et applications 2

Contributed Paper Session / Séance de communications libres

Chair/Président: Zhou Zhou (University of Toronto)

- 13:30 - 13:45 Shirin Golchi & David Campbell (Simon Fraser University)
Sequentially-constrained Monte Carlo / Monte Carlo avec contraintes séquentielles  
- 13:45 - 14:00 Avik Halder & Glen Takahara (Queen's University)
On Simulation of the Beta-Dirichlet Process / Sur la simulation du processus beta-Dirichlet  
- 14:00 - 14:15 Wan-Chen Lee (University of Manitoba)
On Curved Exponential Imbedding and its Applications / À propos de l'imbrication de lois exponentielles courbes et de ses applications  
- 14:15 - 14:30 Francois Perron (University of Montreal), Ayi Ajavon (Université de Montréal)
On the Extensions of Subcopulas / Sur les possibles prolongements d'une sous-copule  
- 14:30 - 14:45 Serge Provost (The University of Western Ontario)
Improved Saddlepoint Density Approximations through Polynomial Adjustments / Approximations en point de selle de la densité par ajustements par polynôme  
- 14:45 - 15:00 Aaron Springford & David Thomson (Queen's University at Kingston)
Analysis of Time series with Unknown Times Using a Hierarchical Chronology Model / L'analyse des séries chronologiques avec temps latents employant un modèle hiérarchique  











13:30 - 14:45 **Session 1D-C5** (abstracts/résumés - page 123) **Hart House - South Dining**

Survey Methodology 1

Méthodologie d'enquête 1

Contributed Paper Session / Séance de communications libres

Chair/Président: Christian Boudreau (University of Waterloo)

- 13:30 - 13:45 Claude Girard (Statistics Canada)
Can the Approximation to the Design-based Variance of Model Assisted Estimators Provided by Linearization be Trusted under Model Misspecification? / Peut-on toujours se fier à l'approximation obtenue par linéarisation de la vraie variance par rapport au plan de sondage d'estimateurs s'appuyant sur un modèle?  
- 13:45 - 14:00 Caren Hasler & Yves Tillé (University of Neuchâtel)
Fast Balanced Sampling for Highly Stratified Population / Échantillonnage équilibré dans des populations hautement stratifiées  
- 14:00 - 14:15 Wilson Lu (Acadia University)
Algorithms for Unequal Probability Sampling Without Replacement / Algorithmes pour l'échantillonnage sans remise à probabilités inégales  
- 14:15 - 14:30 Xichen She & Changbao Wu (University of Waterloo)
Analysis of Complex Survey Data with Missing Ordinal Responses / Analyse de données d'enquêtes complexes avec réponses ordinales manquantes  
- 14:30 - 14:45 Audrey-Anne Vallée & David Haziza (Université de Montréal)
Approximation of the Variance in the Presence of Imputed Data / Approximation de la variance en présence de données imputées  

15:00 - 15:30

Great Hall

Coffee break
Pause-café15:30 - 17:00 **Session 1E-I1** (abstracts/résumés - page 126)

Hart House - Debates Rm

Delivering a Successful Introductory Stats Service Course**Cours de service de base en statistique : meilleures pratiques**







Invited Paper Session / Présentations sur invitation

Chair/Président: Bruce Dunham (University of British Columbia)

Organizer/Responsable: Gordon Hines (University of Guelph)

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

- 15:30 - 16:00 Jeanette O'Hara Hines (University of Waterloo)
Statistics Service Courses and Evolving Undergraduate Science Education / Cours de service en statistique et évolution de l'enseignement scientifique du premier cycle  
- 16:00 - 16:30 Gordon Hines (University of Guelph), R. Jeanette O'Hara Hines (University of Waterloo)
Useful Tactics For Teaching A Successful Stats Service Course / Stratégies utiles pour l'enseignement réussi d'un cours de service en statistique  
- 16:30 - 17:00 Tom Nudds, Gordon Hines & Beren Robinson (University of Guelph)
A Successful Structure for a Practical Introductory Statistics Service Course / Structure réussie pour un cours de service d'introduction à la statistique  



15:30 - 17:00 **Session 1E-I2** (abstracts/résumés - page 128)

Hart House - East Common

Financial Mathematics**Mathématique financière**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Adam Kolkiewicz (University of Waterloo)

- 15:30 - 16:00 Matt Davison & Christian Maxwell (Western University)
The impact of Regulatory Uncertainty on the Operation of Green Energy Projects / Impact de l'incertitude réglementaire sur le fonctionnement de projets d'énergie verte  







- 16:00 - 16:30 Adam Metzler (Wilfrid Laurier University), R.M. Reesor (Western University)
Pricing and Understanding Contingent Capital Bonds using Structural Models / Tarification et compréhension des obligations de capital contingent à l'aide de modèles structuraux  
- 16:30 - 17:00 Anatoliy Swishchuk (University of Calgary), Giovanni Salvi (University of Rome 'La Sapienza')
Covariance and Correlation Swaps for Markov-modulated Volatilities / Swaps de covariance et de corrélation pour des volatilités  

15:30 - 17:00 Session 1E-I3 (abstracts/résumés - page 130) **Hart House - Music Rm**
New Developments in Nonlinear Time Series Modelling and Robust Methods in Time Series Analysis with Applications

Récents progrès en modélisation de séries chronologiques non linéaires; méthodes robustes d'analyse des séries chronologiques et applications

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Ying Zhang (Acadia University)

- 15:30 - 16:00 Herold Dehling (Ruhr-Universität Bochum, Germany), Roland Fried (TU Dortmund), Aeneas Roach (Ruhr-Universität Bochum), Murad Taqqu (Boston University), Martin Wendler (Ruhr-Universität Bochum)
Robust Change-Point Tests for Time Series / Tests de point de rupture robustes pour des séries chronologiques  
- 16:00 - 16:30 Ian McLeod & Nagham Mohammad (Western University)
Censored Time Series Analysis / L'analyse des séries chronologiques censurées  
- 16:30 - 17:00 Hao Yu, Zi Zhen Liu & Reg Kulperger (Western University)
The PLAC-Weighted Adaptive LASSO for VAR(p) Models / LASSO adaptatif pondéré selon la fonction d'autocorrélation à décalage partiel pour les modèles VAR(p)  

15:30 - 17:00 Session 1E-I4 **Hart House - South Dining**

Publishing and Reproducible Research (A Panel Discussion)

Publication et reproductibilité de la recherche (table ronde)

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Karen Kopciuk (University of Calgary)

- 15:30 - 17:00 Theresa Hyslop (Duke University)
- 15:30 - 17:00 Veronique Kiermer (Nature Publishing Group)
- 15:30 - 17:00 Robert Platt (McGill University)





15:30 - 17:00 Session 1E-I5 (abstracts/résumés - page 132) **UC - A101**



Recent Developments in Small Area Estimation

Récents progrès en estimation pour les petits domaines

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Mahmoud Torabi (University of Manitoba)

- 15:30 - 16:00 Jiming Jiang (University of California, Davis), Senke Chen (Freddie Mac Foundation), Thuan Nguyen (Oregon Health and Science University)
Observed Best Prediction for Small Area Counts / Meilleure prévision observée pour des décomptes de petits domaines  
- 16:00 - 16:30 Gauri Datta (University of Georgia), Maples Jerry (US Census Bureau), Mahmoud Torabi (University of Manitoba), Jiashen You (US Census Bureau)
Best Predictive Small Area Estimation in Measurement Error Model / Meilleure estimation prédictive pour petits domaines dans un modèle d'erreur de mesure  







- 15:30 - 17:00 Thuan Nguyen (Oregon Health and Science University), Jiming Jiang (University of California, Davis), J. Sunil Rao (University of Miami)
Observed Best Prediction via Nested-Error Regression with Potentially Misspecified Mean and Variance / Meilleure prévision observée par une régression à erreur emboîtée avec moyenne et variance mal spécifiées  

15:30 - 17:00 Session 1E-I6 (abstracts/résumés - page 134) **UC - 161**

Survival Methods in Genetics
Méthodes de survie en génétique

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: M'Hamed Lajmi Lakhal Chaieb (Université Laval)









- 15:30 - 16:00 Laurent Briollais (Samuel Lunenfel Research Institute), Laurent Briollais (Lunenfeld-Tanenbaum Research Institute), Yun-Hee Choi (Western University)
Statistical Challenges in Modeling Time-to-Event Data in Family Studies: Lessons Learned from the Analysis of Lynch Syndrome and Breast Cancer Families / Défis statistiques de la modélisation de temps avant l'événement dans les études familiales : Leçons apprises dans l'analyse de familles atteintes du syndrome de Lynch et du cancer du sein  
- 16:00 - 16:30 Martin Leclerc, Jacques Simard & Lajmi Lakhal-Chaieb (Université Laval)
Ascertainment-Adjusted SNP Set Analysis for Clustered Age-at-Onset Outcomes / Analyse d'ensembles de SNPs ajustée pour le biais de détermination dans le cas de réponses d'âge d'apparition groupées  
- 16:30 - 17:00 Yildiz Yilmaz (Memorial University)
Analysis of Time-to-Event Phenotypes in Genetic Association Studies of Cancer / Analyse des phénotypes de temps avant l'événement dans les études d'association génétique sur le cancer  





15:30 - 17:00 Session 1E-C1 (abstracts/résumés - page 136) **UC - 179**

Biostatistics - Applications 2
Biostatistique - applications 2

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Katherine Davies (University of Manitoba)

- 15:30 - 15:45 Olaf Berke (University of Guelph), Bimal Chhetri (Simon Fraser University), Zvonimir Poljak (University of Guelph)
Diagnostic Misclassification Bias in Spatial Point Data Analysis -- A Simulation Study / Diagnostique des biais d'erreurs de classification dans une analyse de points de données spatiales - Une étude de simulation  
- 15:45 - 16:00 Mohammad Kaviul Khan, Katerina Maximova & Paul Veugelers (University of Alberta)
Are Overweight Children More Likely to Underestimate Their Weight Status: A Novel Approach to Quantify Weight Status Misperception / Les enfants atteints d'embonpoint sont-ils plus susceptibles de sous-estimer leur poids : une nouvelle approche pour quantifier les perceptions erronées de l'état pondéral  
- 16:00 - 16:15 Zihang Lu (The Hospital For Sick Children), Teterina Anastasia (Toronto General Hospital)
Assessing the Performance of Different Indexes in Finding Cut-off Points for Continuous Biomarkers / Évaluation de la performance de divers indices afin de déterminer les seuils de biomarqueurs continus  
- 16:15 - 16:30 Abdollah Safari (Simon Fraser University), Rosa Aghdam (Shahid Beheshti University), Mehdi Sadeghi & Hamid Pezeshk (University of Tehran)
Generating a Random Gene Expression Dataset with a given Gene Regulatory Network / Comment générer un ensemble de données d'expression génique aléatoire avec un réseau de régulation génétique donné  

- 16:30 - 16:45 Thuva Vanniyasingam, Charles E. Cunningham, Alison Niccols, Gary Foster, Lehana Thabane & (CONNECTIONS) On behalf of CONNECTIONS (McMaster University)
A Discrete Choice Experiment of Evidence-based Practice Dissemination in Addiction Agencies for Women: An Empirical Comparison of Methods for Analyzing Clustered Data / Expérience à choix discret de la dissémination de pratiques fondées sur la preuve dans les organismes de toxicomanie pour les femmes : comparaison empirique de méthodes d'analyse de données regroupées  
- 16:45 - 17:00 Ying Wu & Richard Cook (University of Waterloo)
Variable Selection in Multi-State Analysis by Penalized Likelihood / Sélection de variables dans une analyse multi-états par vraisemblance pénalisée  

15:30 - 17:00 Session 1E-C2 (abstracts/résumés - page 140) **UC - 163**

Biostatistics - Clinical Trials and Diagnostic Protocols

Biostatistique - essais cliniques et protocoles diagnostiques

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Rhonda Rosychuk (University of Alberta)

- 15:30 - 15:45 Cecilia Cotton, Oana Danila, Stefan Steiner, Daniel Severn & Jock MacKay (University of Waterloo)
Using Available Information in the Assessment of Diagnostic Protocols / Utiliser l'information disponible dans l'évaluation des protocoles de diagnostic  
- 15:45 - 16:00 Shun Fu Lee (Population Health Research Institute (PHRI)), Allan Donner & Neil Klar (University of Western Ontario)
Power of the Adjusted Q Statistic for the Meta-Analysis of Cluster Randomization Trials / Puissance de la statistique Q ajustée pour la méta-analyse d'essais randomisés par grappes  
- 16:00 - 16:15 Selvakkadunko Selvaratnam, Yanqing Yi, Alwell Oyet & Veeresh Gadag (Memorial University of Newfoundland)
Estimation of Generalized Linear Mixed Model for Response Adaptive Designs of Multi-Center Clinical Trials / Estimation d'un modèle mixte linéaire généralisé pour des plans s'adaptant aux données dans le cas d'essais cliniques multicentres  
- 16:15 - 16:30 Lan Wen, Leilei Zeng & Richard Cook (University of Waterloo)
The Design and Analysis of Cluster-Randomized Longitudinal Trials Based on Transition Models / Conception et analyse d'essais longitudinaux randomisés en grappes fondés sur des modèles de transition  
- 16:30 - 16:45 Yanqing Yi (Memorial University of Newfoundland), Yanqing Yi
Dynamic Programming and Optimal Designs of Adaptive Clinical Trials / Programmation dynamique et plans optimaux pour les essais cliniques adaptatifs  
- 16:45 - 17:00 Yujie Zhong & Richard Cook (University of Waterloo)
Robust Design of Cluster-Randomized Trials With Censored Event Times / Plans robustes d'essais randomisés par grappes avec temps d'événements censurés  



15:30 - 17:00 Session 1E-C3 (abstracts/résumés - page 144) **UC - 144**











Clustering, Classification and Dimension Reduction

Partitionnement, classification et réduction de dimension

Contributed Paper Session / Séance de communications libres

Chair/Président: Longhai Li (University of Saskatchewan)

- 15:30 - 15:45 Naomi Altman (Penn State University)
Extending PCA to Non-elliptical Data / Généralisation de l'ACP à des données non elliptiques  

- 15:45 - 16:00 Tortora Cristina, Paul D. McNicholas & Ryan P. Browne (University of Guelph)
Mixture of Generalized Hyperbolic Factor Analyzers / Mélange d'analyseurs factoriels hyperboliques généralisés  
- 16:00 - 16:15 Geoffroy Mouret, Jean-Jules Brault & Vahid Partovi Nia (École Polytechnique de Montréal)
No Mean Clustering Algorithm / Algorithme de regroupement sans moyennes  
- 16:15 - 16:30 Yang Tang, Ryan Browne & Paul McNicholas (University of Guelph)
Mixture of Latent Trait Models for High-Dimensional Categorical Data / Mélange de modèles de traits latents pour données catégoriques de grandes dimensions  
- 16:30 - 16:45 Yuhong Wei & Paul McNicholas (University of Guelph)
Mixture Model Averaging for Clustering and Classification / Moyenne de modèles de mélange pour le groupage et la classification  
- 16:45 - 17:00 Lu Xin & Mu Zhu (University of Waterloo), Hugh Chipman (Acadia University)
Continuous-time Stochastic Blockmodels for Transactional Networks / Modèles stochastiques par blocs en temps continu pour les réseaux transactionnels  

15:30 - 17:00 Session 1E-C4 (abstracts/résumés - page 147) **UC - Croft House**

Probability Theory and Estimation

Théorie des probabilités et estimation

Contributed Paper Session / Séance de communications libres

Chair/Président: Deli Li (Lakehead University)

- 15:30 - 15:45 Ernest Dankwa (Memorial University)
Bias Study of the Naive Estimator in a Longitudinal Binary Mixed-effects Model with Measurement Error and Misclassification in Covariates / Étude des biais de l'estimateur naïf dans un modèle longitudinal binaire à effets mixtes avec les erreurs de mesure et de classification dans les covariables  
- 15:45 - 16:00 Shakhawat Hossain (University of Winnipeg), Ejaz Ahmed (Brock University), Grace Yi (University of Waterloo), Baojiang Chen (University of Nebraska Medical Center)
Shrinkage and Pretest Estimators in Partially Linear Models for Longitudinal Data / Estimateurs de rétrécissement et d'essai préliminaire dans des modèles partiellement linéaires de données longitudinales  
- 16:00 - 16:15 Masoud M Nasari & Miklos Csorgo (Carleton University)
On the Accuracy of the Bootstrap / À propos de la précision du bootstrap  
- 16:15 - 16:30 Matthew Schmirler (University of Saskatchewan)
Strand Passage Probabilities in an Interacting Self-Avoiding Polygon Model / Probabilités de passage de brins dans un modèle de polygones auto-évitant interactif  
- 16:30 - 16:45 Trevor Thomson, Shakhawat Hossain & Melody Ghahramani (University of Winnipeg)
Application of Shrinkage Estimation in Linear Regression Models with Autoregressive Errors / Application de l'estimation du rétrécissement dans les modèles de régression linéaire avec erreurs autorégressives  
- 16:45 - 17:00 Monica H.T. Wong & Paul D. McNicholas (University of Guelph)
Biclustering of Gene Expression Data with Applications in Nutrigenomics / Bipartitionnement parcimonieux des données d'expression de gènes avec des applications en nutriginomique  

15:30 - 17:00 Session 1E-C5 (abstracts/résumés - page 150) **UC - 244**

Statistical Inference and Applications 5

Inférence statistique et applications 5

Contributed Paper Session / Séance de communications libres

Chair/Président: Saman Muthukumarana (University of Manitoba)

- 15:30 - 15:45 Amy Morin & Alisha Albert-Green (University of Western Ontario), Douglas Woolford (Wilfrid Laurier University), David Martell (University of Toronto)
Analyzing Forest Fire Lifetime Data Using Proportional Hazards Frailty Models / Analyse des cycles de vie des feux de forêt à l'aide des modèles de fragilité à risques proportionnels  
- 15:45 - 16:00 Wei Lin, Don Fraser & Nancy Reid (University of Toronto)
Adjustments to r : Insides from Taylor Expansions / Ajustements à r : le fonctionnement du développement en série de Taylor  
- 16:00 - 16:15 Matthieu Wilhelm (Université de Neuchâtel), Laura M. Sangalli (Politecnico di Milano)
Generalized Models for Spatial Regression / Modèles généralisés pour régression spatiale  
- 16:15 - 16:30 Sile Tao & Ivan Mizera (University of Alberta)
Prediction of Ice Hockey Performance: A Field Test of Empirical Bayes Procedure / Prédiction de la performance en hockey sur glace : test de terrain d'une procédure bayésienne empirique  
- 16:30 - 16:45 Maryam Sohrabi & Mahmoud Zarepour (University of Ottawa)
Asymptotic Theory for M-Estimates in Unstable AR(P) Processes with Infinite Variance Innovations / Théorie asymptotique pour les M-estimations dans les processus AR(P) instables avec innovations de variance infinie  
- 16:45 - 17:00 Audrey Béliveau, Richard Lockhart & Carl Schwarz (Simon Fraser University), Michael Schaub (Swiss Ornithological Institute), Roger Pradel (Centre d'écologie fonctionnelle et évolutive)
Explicit Integrated Population Modelling: Escaping the Conventional Assumption of Independence / Modélisation explicite intégrée de populations : échapper à l'hypothèse classique d'indépendance  

17:00 - 18:00

East Common Room, Hart House

SSC Annual General Meeting

Assemblée générale annuelle de la SSC

18:00 - 20:00

Great Hall

Student BBQ

Barbecue étudiant

18:00 - 20:00

South Dining Room, Hart House

U of T Alumni reception

Réception des anciens étudiants de l'Université de Toronto

19:00 - 21:00

Bedford Academy under the name "SSC", 36 Prince Arthur Ave, Toronto

COWIS dinner

Souper du Comité de promotion de la femme en statistique

Tuesday May 27**mardi 27 mai****8:15 - 17:00****Hart House Great Hall****Registration
Inscription****8:45 - 9:50****Session 2A-A**

(abstract/résumé - page 153)

MedSci Auditorium**Gold Medal Address****Allocution du récipiendaire de la Médaille d'or**

Special Session / Conférence spéciale

Chair/Président: Don McLeish (University of Waterloo)

Organizer/Responsable: Christian Léger (University of Montreal)

8:45 - 9:50

Jeffrey Rosenthal (University of Toronto)

The Magic of Monte Carlo / La magie de Monte Carlo

**9:50 - 10:20****Great Hall****Coffee break
Pause-café****10:20 - 11:50****Session 2B-I1**

(abstracts/résumés - page 154)

Hart House - South Dining**Actuarial Finance****Finance actuarielle**

Invited Paper Session / Présentations sur invitation


Organizer and Chair / Responsable et président: Jean-Francois Renaud (Université du Québec à Montréal)

Session sponsored by the Actuarial Science Section


Séance parrainée par le Groupe de science actuarielle

10:20 - 10:50

Runhuan Feng (University of Illinois at Urbana-Champaign), Hans Volkmer (University of Wisconsin-Milwaukee)

Spectral Methods for the Valuation of Variable Annuity Guaranteed Benefits / Méthodes spectrales pour l'évaluation des prestations garanties de rente variable **10:50 - 11:20**

Anne Mackay & Carole Bernard (University of Waterloo)

Reducing Surrender Incentives through Fee Structure in Variable Annuities / Structures de frais visant à réduire l'incitatif à la résiliation pour des produits d'assurance liés aux marchés **11:20 - 11:50**

Tom Salisbury & Moshe Milevsky (York University)

Optimal Tontines / Tontines optimales **10:20 - 11:50****Session 2B-I2**

(abstracts/résumés - page 156)


Hart House - East Common**New Developments in High Dimensional Data Analysis****Récents progrès en analyse de données de grande dimension**

Invited Paper Session / Présentations sur invitation


Organizer and Chair / Responsable et président: Abbas Khalili (McGill University)



10:20 - 10:50

Jinch Lv, Yingying Fan & Zemin Zheng (University of Southern California)

High-Dimensional Thresholded Regression and Shrinkage Effect / Régression seuillée de grande dimension et effet de rétrécissement **10:50 - 11:20**

Anand Vidyashankar (George Mason University)

Finite Sample Post Model Selection Inference / Inférence post-sélection de modèle sur des échantillons finis 

- 11:20 - 11:50 Hui Zou (University of Minnesota), Jianqing Fan (Princeton University), Lingzhou Xue (Penn State University)
A General Theory on The Strong Oracle Property of Folded Penalized Estimation / Théorie générale sur la propriété oracle forte de l'estimation pénalisée repliée  

10:20 - 11:50 Session 2B-I3 (abstracts/résumés - page 158) **Hart House - Debates Rm**

New Directions in Causal Inference

Nouvelles directions en inférence causale







Invited Paper Session / Présentations sur invitation

Chair/Président: Lawrence McCandless (Simon Fraser University)

Organizers/Responsables: Lawrence McCandless (Simon Fraser University) and/et Russell Steele (McGill University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 10:20 - 10:50 Peter Austin (University of Toronto)
The Performance of Different Propensity Score Methods for Estimating Marginal Hazard Ratios / La performance de différentes méthodes de score de propension pour estimer les rapports de risque marginaux  
- 10:50 - 11:20 Mireille Schnitzer (Université de Montréal)
Nonparametric Causal Inference through Targeted Maximum Likelihood Estimation: Options for Using Machine Learning in Causal Inference / Inférence causale non paramétrique par l'estimation du maximum de vraisemblance ciblé : options pour l'utilisation de l'apprentissage machine en inférence causale  
- 11:20 - 11:50 Dave Stephens (McGill University)
Discussion of New Directions in Causal Inference / Discussion à propos de nouvelles orientations en inférence causale  

10:20 - 11:50 Session 2B-I4 (abstracts/résumés - page 160) **Hart House - Music Rm**

Semi and Unsupervised Learning







Apprentissage semi- et non supervisé

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Paul McNicholas (University of Guelph)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 10:20 - 10:50 Salvatore Ingrassia (University of Catania, Italy), Antonio Punzo (University of Catania)
Diagnostic Tools for Data Classification via Mixture Models with Covariates / Outils de diagnostic pour la classification de données à l'aide de modèles de mélanges avec covariables  
- 10:50 - 11:20 Geoff McLachlan & Sharon Lee (University of Queensland, Australia)
On Finite Mixtures of Skew Distributions / À propos de mélanges finis de distributions asymétriques  
- 11:20 - 11:50 Irene Vrbik & Paul McNicholas (University of Guelph)
Fractionally-Supervised Classification via Skew-Normal and Skew-t Mixture Models / Classification partiellement supervisée au moyen de modèles de mélanges de lois asymétriques normales et de Student  







10:20 - 11:50 **Session 2B-I5** (abstracts/résumés - page 162) **UC - 140**

Shape-Constrained Maximum Likelihood: Methods and Applications

Maximum de vraisemblance avec contraintes de forme : méthodes et applications

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Hanna Jankowski (York University)

- 10:20 - 10:50 Charles Doss (University of Minnesota), Fadoua Balabdaoui (Université Paris Dauphine), Jon Wellner (University of Washington)
Semiparametric Modeling with Log-Concave Densities / Modélisation semiparamétrique avec des densités log-concaves  
- 10:50 - 11:20 Bodhi Sen (Columbia University), Mary Meyer (Colorado State University)
Testing Against a Parametric Regression Function / Test pour l'existence d'une fonction paramétrique de régression  
- 11:20 - 11:50 Ming Yuan (University of Wisconsin-Madison)
Information-Theoretic Properties of Log-Concave Projections and Their Applications / Propriétés des projections log-concaves du point de vue de la théorie de l'information et de leurs applications  

10:20 - 11:50 **Session 2B-I6** (abstracts/résumés - page 164) **UC - 161**

Stochastic Processes and Their Applications







Processus stochastiques et leurs applications

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Barbara Szyszkowicz (Carleton University)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

- 10:20 - 10:50 Miklos Csorgo & M. Masoud Nasari (Carleton University)
Randomized Empirical Processes with Applications to Infinite Super-Populations and Big Data Sets / Processus empiriques randomisés et applications à des superpopulations infinies et à des jeux de données volumineux  
- 10:50 - 11:20 Don Dawson (Carleton University)
Long Time Behaviour of a Class of Multidimensional Diffusions / Comportement à long terme d'une classe de processus de diffusion multidimensionnels  
- 11:20 - 11:50 Don McLeish (University of Waterloo)
The Importance of Importance Sampling: Measure Change in Probability, Statistics and Rare Event Simulation / Importance de l'échantillonnage par importance : changement de mesure en probabilité, statistique et simulation d'événements rares  



10:20 - 11:50 **Session 2B-C1** (abstracts/résumés - page 166) **UC - 179**

Biostatistics - Measurement Errors and Prediction Errors

Biostatistique - erreurs de mesure et erreurs de prédiction

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Elif Acar (University of Manitoba)

- 10:20 - 10:35 Longhai Li, Shi Qiu, Bei Zhang & Cindy X. Feng (University of Saskatchewan)
Approximating Cross-validated Predictive Evaluation in Bayesian Latent Variables Models with Integrated IS and WAIC / Approximation de l'évaluation prédictive par validation croisée dans des modèles bayésiens à variables latentes à l'aide des critères IS et WAIC intégrés  

- 10:35 - 10:50 Shomoita Alam & Yan Yuan (University of Alberta)
Average Precision – An Overall Measure for Evaluating Classification Prediction Performance / Précision moyenne : mesure globale pour l'évaluation de la performance prévisionnelle de classification  
- 10:50 - 11:05 Shahadut Hossain (UAE University)
Covariate Mismeasurement in Logistic Regression: Bayesian Adjustment of Main and Interaction Effects and the Sample Size Implications / Erreur de mesure dans les covariables en régression logistique : ajustement bayésien des effets principaux et des effets d'interaction, et implications pour la taille d'échantillon  
- 11:05 - 11:20 Shofiqul Islam, Jemila Hamid, Lehana Thabane & Joseph Beyene (McMaster University)
A Copula-based Method to Classify Individuals into Different Disease Categories Using Correlated Biomarkers / Méthode basée sur les copules pour classifier les patients dans différentes catégories de maladies à l'aide de biomarqueurs corrélés  
- 11:20 - 11:35 David Soave & Lei Sun (University of Toronto)
Improved Detection of Variants with Main or Interaction Effects using a Robust Location and Scale Testing Framework / Comment améliorer la détection de variantes avec effets principaux et d'interaction à l'aide d'un cadre de vérification de l'emplacement et de l'échelle robuste  
- 11:35 - 11:50 Ying Yan & Grace Yi (University of Waterloo)
Model Misspecification and Model Checking for Survival Data with Covariate Measurement Error / Erreur de spécification du modèle et vérification du modèle pour les données de survie avec erreur de mesure des covariables  

10:20 - 11:50 Session 2B-C2 (abstracts/résumés - page 170) **UC - A101**

Empirical and Composite Likelihood and Higher Order Asymptotics

Vraisemblances empirique et composite et comportement asymptotique d'ordre supérieur

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Cindy Fu (York University)

- 10:20 - 10:35 Shiva Ashta & Nancy Reid (University of Toronto)
Higher Order Asymptotics and Normal Coefficients of Variation / Propriétés asymptotiques d'ordre supérieur et coefficients de variation normaux  
- 10:35 - 10:50 Song Cai & Jiahua Chen (The University of British Columbia)
Empirical Likelihood Inference under Density Ratio Models Based on Type I Censored Multiple Samples / Inférence de la vraisemblance empirique sous les modèles de rapport de densité sur des échantillons multiples censurés de type I  
- 10:50 - 11:05 Mahdis Azadbakhsh, Xin Gao & Hanna Jankowski (York University)
Multiple Comparisons using Composite Likelihood in Clustered Data / Comparaisons multiples par vraisemblance composite pour des données en grappes  
- 11:05 - 11:20 Min Chen, Changbao Wu & Mary Thompson (University of Waterloo)
Empirical Likelihood Methods for Testing the Difference of Two Distributions in Pretest-Posttest Studies / Méthodes de vraisemblance empiriques afin de vérifier la différence entre deux distributions dans des études pré-test et post-test  
- 11:20 - 11:35 Tharshanna Nadarajah, Asokan Mulayath Variyath & J Concepción Loredó-Osti (Memorial University of Newfoundland)
Empirical Likelihood Based Longitudinal Studies / Études longitudinales basées sur la vraisemblance  
- 11:35 - 11:50 Ximing Xu & Hal Whitehead (Dalhousie University)
A Composite Likelihood Approach to Whale's Network Analysis / Approche de vraisemblance composée pour l'analyse des réseaux de baleines  

10:20 - 11:50 Session 2B-C3 (abstracts/résumés - page 173) **UC - 244**

Nonregular Statistical Test Problems

Problèmes de tests statistiques non réguliers

Contributed Paper Session / Séance de communications libres

Chair/Président: Farouk Nathoo (University of Victoria)

- 10:20 - 10:35 Joel Sango & Duchesne Pierre (Université de Montréal)
 Test for Non-Correlation Under Weak Dependence Between Two Stationary Processes / Test de non-corrélation sous faible dépendance entre deux processus stationnaires  
- 10:35 - 10:50 Luai Al Labadi & Emad Masuadi (UAE University), Mahmoud Zarepour (University of Ottawa)
 Two-Sample Bayesian Nonparametric Goodness of Fit Test / Test d'adéquation non paramétrique bayésien pour deux échantillons  
- 10:50 - 11:05 Debaraj Sen (Concordia University), Krishna K. Saha (Central CT State University)
 Robust Test for Testing Homogeneity of Dispersions for Several Groups of Counts Using Dissimilarity Measures / Test robuste de l'homogénéité des dispersions dans plusieurs groupes de dénombrements par mesures de dissimilitude  
- 11:05 - 11:20 Joseph François Tagne Tatsinkou, Pierre Duchesne & Pierre Lafaye de Micheaux (Université de Montréal)
 On Smooth Test of Goodness of Fit for Vector Time Series with Structured Parameterization / Un test lisse d'ajustement pour les séries chronologiques multivariées avec paramétrisation structurée  
- 11:20 - 11:35 Zheng Sun (Simon Fraser University)
 Testing Goodness-of-Fit for Latent Variables / Comment tester la qualité d'ajustement de variables latentes  
- 11:35 - 11:50 Joshua Pohlkamp-Hartt (Queen's University)
 Bootstrapped Multitaper F-test for Signal Detection / Test F « multitaper » bootstrap pour la détection de signaux  









10:20 - 11:50 Session 2B-C4 (abstracts/résumés - page 176) **UC - Croft House**



Statistical Inference and Applications 3



Inférence statistique et applications 3

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Xin Gao (York University)

- 10:20 - 10:35 Alexandre Leblanc, Stephane Durocher, Jason Morrison, Matthew Skala & Robert Fraser (University of Manitoba)
 On the Probability of Being Interior to a Random Subset of a Sample / À propos de la probabilité d'être à l'intérieur d'un sous-ensemble aléatoire d'un échantillon  
- 10:35 - 10:50 Tom Loughin & William Ruth (Simon Fraser University)
 The Effect of Heteroscedasticity on Regression Trees / Effet de l'hétéroscédasticité sur les arbres de régression  
- 10:50 - 11:05 Mina Mirshahi (École Polytechnique de Montréal), Masoud Asgharian (McGill University), Vahid Partovi Nia (École Polytechnique de Montréal)
 Cubic Splines for Shapes in Two-Dimensional Space / Des splines cubiques pour des formes dans un espace à deux dimensions  
- 11:05 - 11:20 Pavel Petrov & Giseon Heo (University of Alberta)
 Applications of Persistent Homology to DNA data / Applications de l'homologie persistante à des données sur l'ADN  

11:20 - 11:35 Wanhua Su (MacEwan University), Peng Zhang (Zhejiang University)
Quantile Methods for Estimating Bi-normal Parameters in ROC Analysis / Méthodes de quantiles pour estimer les paramètres bi-normaux dans une analyse ROC  

11:35 - 11:50 Karim Rahim & David Thomson (Queen's University)
Change-point Detection using Level-of-change in Spectra / Détection des points de rupture à l'aide du degré de changement des spectres  

12:00 - 13:30 **UC 256**
Accreditation General Meeting
Assemblée générale des membres accrédités

12:00 - 18:00 **South Sitting Room, Hart House**
Award for Case Studies in Data Analysis Committee 2
Comité du prix pour les études de cas et l'analyse de données 2

12:00 - 13:30 **UC 53**
Biostatistics Section Executive Committee 2
Comité exécutif du groupe de biostatistique 2

12:00 - 13:30 **UC 244**
CJS Editorial Board
Comité de rédaction de la RCS

12:00 - 13:30 **UC 52**
Census at School Committee
Comité de recensement à l'école

12:00 - 13:30 **UC 248 (Rain location)**
Committee on New Investigators
Comité des nouveaux chercheurs

12:00 - 13:30 **UC 257**
Office Committee
Comité du Bureau

12:00 - 13:30 **UC 255**
Public Relations Committee
Comité des relations publiques

12:00 - 13:30 **UC 144**
Statistical Education Committee
Comité d'éducation en statistique

12:00 - 17:30 **Session 2C-P1** (abstracts/résumés - page 179) **Great Hall**

Actuarial Science Poster Session

Séance d'affiches en science actuarielle

Contributed Poster Session / Séance d'affichage libre

- 12:00 - 17:30 Huan Gao, Rogemar Mamon & Xiaoming Liu (The University of Western Ontario)
Risk Measurement of a Guaranteed Annuity Option in a Stochastic Environment / Mesure du risque d'une option de rente garantie dans un environnement stochastique
- 12:00 - 17:30 Luyao Lin, Li Chen, Yi Lu & Gary Parker (Simon Fraser University)
Insurance and Investment Risks of a Joint Life Insurance Portfolio / Risques d'assurance et d'investissement d'un portefeuille d'assurance-vie conjointe
- 12:00 - 17:30 Yuzhou Zhang & Bruce L. Jones (Western University)
Modelling Disability Using Mixtures of Markov Chains / Modélisation de l'invalidité à l'aide de mélanges de chaînes de Markov

12:00 - 17:30 **Session 2C-P2** (abstracts/résumés - page 181) **Great Hall**

Poster Session

Séance d'affichage

Contributed Poster Session / Séance d'affichage libre

- 12:00 - 17:30 Justin Angevaere, Daniel Gillis & Gerarda Darlington (University of Guelph)
Efficient Bayesian Inference for Conditionally Autoregressive Models / Inférence bayésienne efficace pour modèles conditionnellement autorégressifs
- 12:00 - 17:30 Carolyn Augusta & Rob Deardon (University of Guelph)
A Comparison of MCMC and MLE in Spatial Epidemic Models / Une comparaison entre MCMC et MLE pour des modèles spatiaux de maladie infectieuse
- 12:00 - 17:30 JOSE MARTIN CADENA-BARAJAS (UNIVERSIDAD VERACRUZANA), Sergio Francisco Juarez-Cerrillo (Universidad Veracruzana), David A. Stephens (McGill University)
Bayesian Modeling of Wind Speed and Direction / Modélisation bayésienne de la vitesse et de la direction du vent
- 12:00 - 17:30 Yanling Cai & James Zidek (University of British Columbia)
"Quality Over Quantity": Damage Accumulation Models for Lumber Products / Qualité ou quantité : modèles d'accumulation des dommages pour les produits de bois d'œuvre
- 12:00 - 17:30 Hao Chen & William Welch (University of British Columbia)
Sequential Experimental Design for Percentile Estimation -- A Floor System Example / Plan d'expérience séquentiel pour l'estimation des percentiles -- un exemple des systèmes de planchers
- 12:00 - 17:30 Osvaldo Espin-Garcia (University of Toronto / Public Health Ontario), Michael Lebenbaum (Public Health Ontario), Laura Rosella & Ye Li (Public Health Ontario / University of Toronto)
Model Validation Via the 632+ Bootstrap Method in a Longitudinal Survey Sampling Context: An Application for the Obesity Population Risk Tool / Validation de modèles via la méthode du bootstrap 632+ dans un contexte d'échantillonnage de sondage longitudinal : application à l'Obesity Population Risk Tool
- 12:00 - 17:30 ANDREW GARRO (Lakehead University)
Sampling Techniques: A Detailed Analysis of Three Sector Funds / Techniques d'échantillonnage : analyse détaillée de trois fonds sectoriels
- 12:00 - 17:30 Mohammad Sajjad Ghaemi, Bruno Agard, Martin Trépanier & Vahid Partovi-Nia (École Polytechnique de Montréal)
Computational Complexity in Hierarchical Clustering Algorithms / Complexité de calcul dans les algorithmes hiérarchiques

- 12:00 - 17:30 Christopher Gravel & Patrick Farrell (Carleton University), Daniel Krewski (University of Ottawa)
Resampling for Outcome Misclassification: Internal Validation Sampling Concerns / Rééchantillonnage pour erreur de classement des résultats : problèmes d'échantillonnage de validation interne
- 12:00 - 17:30 Yumi Kondo, Yumi Kondo, Yinshan Zhao & John Petkau (University of British Columbia)
Who Is a Responder? Identification of Treatment Responders Based on Longitudinally Collected Multiple Endpoints in Multiple Sclerosis Patients / Qui est un répondant ? Identification des répondants au traitement en fonction de plusieurs critères collectés longitudinalement chez les patients atteints de sclérose en plaques
- 12:00 - 17:30 Violeta Kovacev-Nikolic & Giseon Heo (University of Alberta), Peter Bubenik (Cleveland State University), Dragan Nikolic (University of Alberta (at the time of initial research))
Topological Data Analysis of the Maltose Binding Protein / Analyse de données topologiques de la protéine de fixation du maltose
- 12:00 - 17:30 Jia Li & Taraneh Abarin (Memorial University of Newfoundland)
Bias Study of the Naive Estimator in a Longitudinal Linear Mixed-effects Model with Measurement Error and Misclassification in Covariates / Étude du biais de l'estimateur naïf dans un modèle à effets mixtes linéaire longitudinal avec erreurs de mesure et de classement des covariables
- 12:00 - 17:30 Zhong Li & Kristina Sendova (University of Western Ontario)
On a Dual Risk with Dependence Between Gains Sizes and Inter-Gain Times / Double risque avec dépendance entre la taille des gains et le temps entre gains
- 12:00 - 17:30 Erin Lundy (Western University), Elizabeth Juarez-Colunga (University of Colorado Denver), Charmaine Dean (Western University)
A Joint Model for Detecting Patterns of Antisocial Behaviour Among Serious Juvenile Offenders / Modèle conjoint pour la détection de tendances de comportement antisocial chez les délinquants juvéniles graves
- 12:00 - 17:30 Mahmud Mamun, H.S. Shin, W. Burr, M. Smith-Doiron, B. Jovic, D. Stieb & B. Jessiman (Health Canada)
Air Health Indicator (AHI) / Indicateur air-santé (IAS)
- 12:00 - 17:30 Lucia Mirea (Maternal-Infant Care Research Centre, Mount Sinai Hospital and the University of Toronto), Junmin Yang (Maternal-Infant Care Research Centre, Mount Sinai Hospital), Shoo K. Lee & Prakesh S. Shah (Pediatrics, Mount Sinai Hospital and the University of Toronto)
Evaluation of Propensity Score Methods for Multiple Treatment Groups / Évaluation des méthodes de score de propension pour les groupes de traitement multiples
- 12:00 - 17:30 Khurram Nadeem, Ying Zhang & Hugh Chipman (Acadia University)
Integrating Population Dynamics Models and Distance Sampling Data: A Hierarchical State-Space Approach / Intégration de modèles de dynamiques de la population et données d'échantillonnage de la distance : une approche espace d'états hiérarchique
- 12:00 - 17:30 Kunasekaran Nirmalkanna & Boxin Tang (Simon Fraser University)
Optimal Designs of Two-Level Factorials when $N \equiv 1$ and $2 \pmod{4}$ under a Baseline Parameterization / Plans factoriels optimaux à deux modalités où $N \equiv 1$ et $2 \pmod{4}$ avec un paramétrage comportant un niveau de base
- 12:00 - 17:30 Mohammad Nourmohammadi, Mohammad Jafari Jozani & Brad Johnson (University of Manitoba)
Distribution-Free Tolerance Intervals with Randomized Nomination Samples / Intervalles de tolérance non paramétriques avec des échantillons nominatifs randomisés
- 12:00 - 17:30 ALBERTO PADILLA (BANCO DE MEXICO)
Two-Stage Cluster Sampling: Bounds for the Variance, Coefficient of Variation and Design Effect of Proportions / Échantillonnage en grappes à deux étapes : bornes pour la variance, le coefficient de variation et l'effet du plan sur les proportions
- 12:00 - 17:30 Leila Sinaei (Shiraz University)
Estimation Of Mean In Double Sampling For Stratification Using Multivariate Auxiliary Information / Estimation de la moyenne avec un échantillonnage double stratifié à l'aide d'information auxiliaire multidimensionnelle

- 12:00 - 17:30 Anu Stanley, Rob Deardon & Zeny Feng (University of Guelph)
Using School Absenteeism in Disease Surveillance Models: Refining of Surveillance Threshold / Utilisation de l'absentéisme scolaire dans des modèles de surveillance épidémiologique : le raffinement du seuil de surveillance
- 12:00 - 17:30 Jiaying (Vicky) Weng (Simon Fraser University)
Maximin Strong Orthogonal Arrays / Tableaux orthogonaux robustes maximin
- 12:00 - 17:30 Jinkun Xiao & A.Ian McLeod (University of Western Ontario)
Bootstrapping Improvement to Portmanteau Test in ARFIMA Model / Amélioration au test portemanteau par le bootstrap dans un modèle ARFIMA

12:00 - 17:30 Session 2C-P3 (abstracts/résumés - page 193) **Great Hall**

Probability Student Poster Session
Séance d'affiches étudiants en probabilité

Contributed Poster Session / Séance d'affichage libre

- 12:00 - 17:30 Chi Dong (University of Alberta)
Nonlinear Filtering on Completely Regular Spaces / Filtrage non linéaire sur des espaces complètement réguliers
- 12:00 - 17:30 Haosui Duanmu (University of Toronto)
Nonstandard Analysis and Its Application to Convergence of Markov Processes / Analyse non standard appliquée à la convergence des processus de Markov

13:30 - 15:00 Session 2D-I1 (abstracts/résumés - page 195) **Hart House - South Dining**







David Binder Memorial Session
Séance commémorative David Binder

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Mary Thompson (University of Waterloo)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 13:30 - 14:00 Abdellatif Demnati (Statistics Canada), Rao J.N.K. (Carleton University)
Variance Estimation from Complex Survey Data Using Linearization Method: Impact of David Binder / Estimation de la variance par linéarisation à partir de données d'enquêtes complexes : influence de David Binder  
- 14:00 - 14:30 Milorad Kovacevic (United Nations Development Programme)
Measuring Multidimensional Poverty and Inequality / Mesure multidimensionnelle de la pauvreté et de l'inégalité  
- 14:30 - 15:00 Natalie Shlomo (University of Manchester), Rodolphe Priam (University of Southampton)
Calibration of Small Area Estimates in Business Surveys / Calibrage des estimations pour les petits domaines dans les enquêtes auprès des entreprises  







13:30 - 15:00 Session 2D-I2 (abstracts/résumés - page 197) **Hart House - East Common**

Gaussian Process Regression: Innovations and Applications
Régression par processus gaussien : innovations et applications

Invited Paper Session / Présentations sur invitation

Chair/Président: Mu Zhu (University of Waterloo)

Organizer/Responsable: Martin Lysy (University of Waterloo)

- 13:30 - 14:00 Derek Bingham (Simon Fraser University), Robert Gramacy (Booth School of Business, The University of Chicago)
Fast Gaussian Process Calibration of Computer Models with Large Ensembles / Calage rapide de modèles de processus gaussiens présentant de grands ensembles  
- 14:00 - 14:30 Jason Loeppky (University of British Columbia), Erik Rosolowsky (University of Alberta)
Comparing Simulated Emission from Molecular Clouds using Gaussian Process Regression / Comparaison d'émissions simulées de nuages moléculaires à l'aide de la régression de processus gaussiens  
- 14:30 - 15:00 Martin Lysy & Ethan Fang (University of Waterloo)
Parameter Inference for Diffusion Processes via Gaussian Process Regression / Inférence sur les paramètres de processus de diffusion à l'aide de la régression de processus gaussiens  

13:30 - 15:00 Session 2D-I3 (abstracts/résumés - page 199) **Hart House - Debates Rm**

Graduate Student Presentations in Actuarial Science



Présentations d'étudiants de cycle supérieur en science actuarielle

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Jiandong Ren (Western University)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe de science actuarielle

- 13:30 - 13:45 Anas Abdallah (Université Laval), Jean-Philippe Boucher (UQAM), Hélène Cossette (Université Laval)
Modeling Dependence Between Loss Triangles Using Hierarchical Archimedean Copulas / Modélisation de la dépendance entre triangles de développement à l'aide des copules archimédiennes hiérarchiques  
- 13:45 - 14:00 Fangda Liu, Jun Cai & Christiane Lemieux (University of Waterloo)
Optimal Reinsurance with Regulatory Initial Capitals and Default Risks / Réassurance optimale grâce à une réglementation en matière de capital initial et de risque de défaut de paiement  
- 14:00 - 14:15 Hassan Omid (Université de Montréal), Zied Ben-Salah (Concordia University), Helene Guerin (University of Rennes), Manuel Morales (University of Montreal)
On the Depletion Problem for an Insurance Risk Process: New Non-ruin Quantities in Collective Risk Theory / À propos du problème d'épuisement pour un processus de risque d'assurance : de nouvelles données quantitatives de non ruine en matière de théorie du risque collectif  
- 14:15 - 14:30 Oscar Quijano & José Garrido (Concordia University)
Linear Credibility for GLM Losses / Modèles linéaires généralisés avec crédibilité linéaire pour les pertes  
- 14:30 - 14:45 Dameng Tang, X. Sheldon Lin, Andrei L. Badescu & Lan Gong (University of Toronto)
Fitting Erlang-Based Mixture Models to Operational Loss Data / Ajustement de modèles de mélanges d'Erlang à des données de perte opérationnelle  
- 14:45 - 15:00 Chen Yang & Kristina Sendova (Western University)
On the Threshold Strategy for Dividend Payments Under the Dual Model Perturbed by Diffusion / À propos de la stratégie de seuil pour les versements de dividendes selon le modèle dual perturbé par la diffusion  
-

13:30 - 15:00 **Session 2D-I4** (abstracts/résumés - page 202) **Hart House - Music Rm**

New Methods for Analyzing Brain Imaging data
Nouvelles méthodes d'analyse des données d'imagerie du cerveau

Invited Paper Session / Présentations sur invitation

Chair/Présidente: Michelle Zhou (Simon Fraser University)

Organizer/Responsable: Ivor Cribben (University of Alberta)







- 13:30 - 14:00 Ivor Cribben (University of Alberta), Christian Habeck (Columbia University Medical Center)
 A New Method for Jointly Estimating Dynamic Functional Connectivity / Nouvelle méthode pour l'estimation conjointe de la connectivité fonctionnelle dynamique  
- 14:00 - 14:30 Linglong Kong (University of Alberta), Hongtu Zhu & Hongyu An (University of North Carolina at Chapel Hill), L. Andria Ford (Washington University at St. Louis)
 Simultaneously Sparse Partial Functional Linear Regression in Hyper-Acute Ischemic Stroke Study / Régression linéaire fonctionnelle partielle simultanément éparsée dans une étude sur l'accident ischémique cérébral hyper-aiguë  
- 14:30 - 15:00 Farouk Nathoo & Arif Babul (University of Victoria), Alexander Moiseev (Simon Fraser University), Naznin Virji-Babul (University of British Columbia), Mirsa Faisal Beg (Simon Fraser University)
 Variational Bayes High-Dimensional Discrete-Time Dynamical Systems for Combined MEG and EEG Data / Systèmes bayésiens dynamiques variationnels de grande dimension à temps discret pour des données combinées de MEG et d'EEG  

13:30 - 15:00 **Session 2D-I5** (abstracts/résumés - page 204) **UC - 140**

Statistical Challenges in High-Throughput Genomic Data
Problèmes statistiques des données génomiques à débit élevé

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Kun Liang (University of Waterloo)

- 13:30 - 14:00 Hongkai Ji (Johns Hopkins University), Bing He & Fang Du (Johns Hopkins Bloomberg School of Public Health)
 A Model-based Approach for Characterizing Correlation in Big Genomic Data / Approche pour la caractérisation des corrélations données génomiques volumineuses basée sur un modèle  
- 14:00 - 14:30 Lei Sun & Andriy Derkach (University of Toronto), Jerry Lawless (University of Waterloo)
 Methods for Association Analysis of Rare Variants: A Review and Some New Developments / Méthodes d'analyse d'association de variantes rares : bilan et nouveaux développements  
- 14:30 - 15:00 Wei Sun (University of North Carolina at Chapel Hill)
 A Study of RNA-seq Data in Cancer Patients / Étude des données de séquençage d'ARN des patients cancéreux  

13:30 - 15:00 **Session 2D-I6** (abstracts/résumés - page 206) **UC - 179**







Weakly Interacting Processes in Finance and Economics
Processus à faible interaction en finance et économie

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Bruno Remillard (HEC Montreal)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

- 13:30 - 14:00 Christoph Frei (University of Alberta)
Financial Markets with Interacting Investors and Their Relation to Backward SDEs / Marchés financiers en présence d'investisseurs en interaction et leur relation à l'égard d'équations différentielles stochastiques (EDS) rétrogrades  
- 14:00 - 14:30 Minyi Huang (Carleton University)
Mean Field Capital Accumulation Optimization and Decentralized Strategies / Stratégies décentralisées et optimisation d'accumulation de domaine de capital moyen  
- 14:30 - 15:00 Michael Kouritzin (University of Alberta)
Resampled Particle Filtering by McKean-Vlasov Thinking / Tri de particules ré-échantillonnées par l'approche McKean-Vlasov  

13:30 - 15:00 **Session 2D-C1** (abstracts/résumés - page 208) **UC - 52**

Biostatistics - Applications 3
Biostatistique - applications 3

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Yeying Zhu (University of Waterloo)

- 13:30 - 13:45 Caitlin Daly & Jemila Hamid (McMaster University)
Establishing Paediatric Reference Intervals: Adjusting for Complex Data / Établir des intervalles de référence pédiatriques : ajuster pour les données complexes  
- 13:45 - 14:00 Shahedul Khan (University of Saskatchewan)
A Multistate Modeling Approach to Study the Effects of Air Pollutants on Recurrent Hospital Admissions Due to Respiratory Diseases / Approche de modélisation multi-états pour l'étude des effets des polluants atmosphériques sur les admissions répétitives à l'hôpital pour cause de maladie respiratoire  
- 14:00 - 14:15 Farhood Mohammadi & Keumhee Carriere Chough (University of Alberta)
Comparison of Missing data Methods for Repeated Measures Binary Data / Comparaison des méthodes pour données manquantes appliquées aux données binaires de mesures répétées  
- 14:15 - 14:30 Thierry Chekouo Tekougang & Francesco Stingo (UT MD Anderson Cancer Center), James Doecke (CSIRO Computational Informatics, Australia), Kim-Anh Do (UT MD Anderson Cancer Center)
A Bayesian Approach to Biomarker Selection Through miRNA Regulatory Network with Application to Kidney Cancer / Une approche bayésienne de la sélection des biomarqueurs par un réseau de régulation microARN avec application au cancer du rein  
- 14:30 - 14:45 Shabnam Vatanpour, Farzana Yasmin & Xiaoming Wang (University of Alberta), Saumyadipta Pyne (CR Rao Advanced Institute of Mathematics, Statistics and Computer Science/Public Health Foundation of India), Irina Dinu (University of Alberta)
Linear Combination Test and Reduction in Microarray Studies / Test des combinaisons linéaires et de réduction dans les études de microréseau  
- 14:45 - 15:00 Guosheng Yin (The University of Hong Kong)
Two-stage Single-to-double Arm Design for Phase II Clinical Trials / Plan à branches simples et doubles à deux étapes pour les essais cliniques de phase II  

13:30 - 15:00 **Session 2D-C2** (abstracts/résumés - page 212) **UC - 244**

Biostatistics - Association Studies and Correlation Structure
Biostatistique - études d'association et structure de corrélation

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Yildiz Yilmaz (Memorial University)

- 13:30 - 13:45 Lajmi Lakhal-Chaieb (Université Laval), Karim Oualkacha (Université du Québec à Montréal), Celia Greenwood (McGill University)
A Family-based Test for Rare Variants Association in the Presence of Non-Normal Outcomes / Un test d'association pour les variantes rares avec des données familiales et des observations non-normales  
- 13:45 - 14:00 Muhammad Abu Shadeque Mullah & Andrea Benedetti (McGill University)
Smoothing in a Semi-Parametric Generalized Linear Mixed Model and its Effect on the Estimation of Covariance Structures for Correlated Data / Lissage dans un modèle linéaire généralisé mixte semi-paramétrique et son effet sur l'estimation des structures de covariance des données corrélées  
- 14:00 - 14:15 Chul Gyu Park (Carleton University)
A Representation for Binary Random Vector with Fixed Mean and Correlation Matrix / Représentation pour un vecteur aléatoire binaire à matrice de corrélation et à moyenne fixes  
- 14:15 - 14:30 Hela Romdhani, Aurelie Labbe & Heungsun Hwang (McGill University)
Genome-Wide Association Study Using Generalized Structured Component Analysis / Étude d'association pangénomique avec l'analyse en composantes structurées généralisées  
- 14:30 - 14:45 Ji-Hyung Shin & Shelley B. Bull (University of Toronto and Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital)
Can Sparse Data Methods Rescue Low Frequency Variants in Genome Wide Association Analysis? / Les méthodes de données éparées peuvent-elles sauver les variantes à faible fréquence dans l'analyse d'associations pangénomiques?  
- 14:45 - 15:00 Jianping Sun (McGill University), Karim Oualkacha (Université du Québec à Montréal), Celia Greenwood (McGill University, Lady Davis Institute for Medical Research)
A Method for Analyzing Multiple Continuous Phenotypes in Rare Variant Association Studies / Méthode d'analyse de phénotypes continus multiples dans les études d'associations de variantes rares  









13:30 - 15:00 Session 2D-C3





(abstracts/résumés - page 216)

UC - A101**Business and Industrial Statistics 1****Statistique industrielle et de gestion 1**

Contributed Paper Session / Séance de communications libres

Chair/Président: Hugh Chipman (Acadia University)

- 13:30 - 13:45 Kanika Anand (Acadia University)
Sequential Design for Noisy Computer Experiments with a Bayesian Ensemble Model / Plan séquentiel pour des expériences informatiques perturbées par des bruits avec un modèle d'ensemble bayésien  
- 13:45 - 14:00 Neil Spencer (University of British Columbia), Pritam Ranjan & Franklin Mendivil (Acadia University)
Obtaining Two-Level Factorial Designs with Pre-specified Randomization Restrictions / Comment obtenir des plans factoriels à deux niveaux avec des restrictions de randomisation préspecifiées  
- 14:00 - 14:15 Nathaniel Stevens & Stefan Steiner (University of Waterloo), Ryan Browne (University of Guelph), Jock MacKay (University of Waterloo)
Incorporating Baseline Information When Assessing Measurement Systems / Comment inclure les données de base dans l'évaluation de systèmes de mesure  
- 14:15 - 14:30 Si Chen (University of Waterloo), Zilin Wang (Wilfrid Laurier University)
A Bias Correction for Capital Buffer Estimation in Retail Credit Risk Management / Une correction du biais dans l'estimation de la marge de capital dans la gestion du risque du crédit au détail  

- 14:30 - 14:45 Matthew van Bommel, Hugh Chipman & Pritam Ranjan (Acadia University)
Stage-Wise Surrogate Modeling of Tidal Power Functions / Modélisation de substituts par étapes des fonctions d'énergie marémotrice  
- 14:45 - 15:00 Fateha Khanam Bappee & Xu (Sunny) Wang (St. Francis Xavier University)
Adaptively Chosen k in Weighted k-Nearest Neighbour Method for Drug Discovery Data / Sélection adaptative du k dans une méthode des k plus proches voisins pondérée pour des données de découverte de médicaments  

13:30 - 15:00 Session 2D-C4 (abstracts/résumés - page 219) **UC - Croft House**

Quantile, Spatial and Logistic Regression

Régression quantile, spatiale et logistique

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Cecilia Cotton (University of Waterloo)





- 13:30 - 13:45 Taoufik Bouezmarni (Université de Sherbrooke), Benedikt Funk (Technische Universität Dortmund), Félix Camirand Lemyre (Université de Sherbrooke)
Regression Estimation Based on Bernstein Density Copulas / Estimation de la régression en utilisant les densités de copules de Bernstein.  
- 13:45 - 14:00 Charles Fontaine & Taoufik Bouezmarni (Université de Sherbrooke), Anouar El Ghouh (Université catholique de Louvain)
Semiparametric Estimation of the Regression Function in Situation of Right-Censored Data / Estimation semi-paramétrique de la fonction de régression dans une situation de données censurées à droite  
- 14:00 - 14:15 Chithran Vadaverkkot Vasudevan, Zhao Zhi Fan & Asokan Variyath (Memorial University of Newfoundland)
Censored Quantile Regression with Auxiliary Covariates / Régression quantile censurée avec covariables auxiliaires  
- 14:15 - 14:30 Mark Wolters & Charmaine Dean (Western University)
Detection of Smoke in Satellite Images Using Autologistic Regression / Détection de la fumée dans les images satellite par régression auto-logistique  
- 14:30 - 14:45 Yin Xu, Mei Ling Huang & Wai Kong Yuen (Brock University)
On High Quantile Regression / Régression des quantiles supérieurs  
- 14:45 - 15:00 Xuemao Zhang (Trent University)
Teaching Introductory Statistics Using RExcel / Enseigner l'introduction à la statistique à l'aide de RExcel  









13:30 - 15:00 Session 2D-C5 (abstracts/résumés - page 222) **UC - 144**

Statistical Inference and Applications 4

Inférence statistique et applications 4

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Leilei Zeng (University of Waterloo)

- 13:30 - 13:45 Jules Joseph Sunda de TIBEIRO & Vartan CHOULAKIAN (Université de Moncton)
Graph Partitioning by Correspondence Analysis and Taxicab Correspondence Analysis / Partitionnement des graphes par l'analyse factorielle des correspondances et par l'analyse factorielle des correspondances du taxi  
- 13:45 - 14:00 Gabriel Escarela (Universidad Autónoma Metropolitana - Iztapalapa)
Analysis and Prediction of Daily Maxima of Ozone Levels in Guadalajara City / Analyse et prédiction des maximums quotidiens des niveaux d'ozone dans la ville de Guadalajara  

- 14:00 - 14:15 Feng (Frank) Feng (City of Edmonton - Office of Traffic Safety)
Time Series Analysis for Collision Forecasting / Analyse de séries chronologiques pour la prédiction de collisions  
- 14:15 - 14:30 David Han (University of Texas)
Time and Cost Constrained Optimal Designs of Constant-stress and Step-stress Accelerated Life Tests / Plans optimaux d'essais de vie accélérée sous contraintes de temps et de coût pour une tension constante ou échelonnée  
- 14:30 - 14:45 Rajat Malik, Rob Deardon & Grace P.S. Kwong (University of Guelph)
Parameterizing Individual-Level Models of Infectious Disease Spread Using Sampling-Based Likelihood Approximations / Établissement des paramètres de modèles au niveau individuel pour la propagation de maladies infectieuses par des approximations de la vraisemblance basée sur l'échantillonnage  
- 14:45 - 15:00 G. Arthur Mihram (Princeton, NJ), Danielle Mihram, Ph. D. (University of Southern California)
Re-examining Quinn's Conclusion: Don't Either Our Statistical Methodologies or Works in Applied Mathematics Qualify as Science? / Réexamen de la conclusion de Quinn : nos méthodes statistiques et nos travaux en mathématiques appliquées ne se qualifient-ils pas au titre de science?  

15:00 - 15:30**Great Hall****Coffee break****Pause-café****15:30 - 17:00 Session 2E-I1** (description - page 226)**Hart House - Debates Rm****Bridging the Gap between Methods Development and Implementation in Biostatistics (A Panel Discussion)**
Comment combler le fossé entre le développement et la mise en œuvre de méthodes en biostatistique (table ronde)

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Eleanor Pullenayegum (Hospital for Sick Children)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

15:30 - 17:00 Melanie Barwick (University of Toronto)

15:30 - 17:00 Brian Feldman (University of Toronto)

15:30 - 17:00 Martin Offringa (University of Toronto)

15:30 - 17:00 Robert Platt (McGill University)

15:30 - 17:00 Lehana Thabane (McMaster University)



15:30 - 17:00 Session 2E-I2 (abstracts/résumés - page 227)**Hart House - Music Rm****Challenges and New Developments in Complex Multiple Comparison Procedures****Défis et progrès récents pour les procédures de comparaisons multiples complexes**

Invited Paper Session / Présentations sur invitation



Chair/Président: Wenqing He (Western University)



Organizer/Responsable: Jianan Peng (Acadia University)

15:30 - 16:00 Frank Bretz & Willi Maurer (Novartis Pharma AG)

Multiple Testing in Group Sequential Trials using Graphical Approaches / Tests multiples dans des essais séquentiels par groupes à l'aide de méthodes graphiques  

16:00 - 16:30 Siu Hung Cheung (Chinese University of Hong Kong)

Non-inferiority Studies with Multiple New Treatments / Études de non-infériorité en présence de nouveaux traitements multiples  

- 16:30 - 17:00 Jianan Peng (Acadia University)
Multiple Confidence Intervals for Selected Parameters Adjusted for the False Coverage-Statement Rate in Dose-Response Microarray Experiments / Intervalles de confiance multiples pour des paramètres sélectionnés et ajustés en fonction du taux de fausse couverture dans le cadre d'expériences avec biopuces portant sur la relation dose-effet  







15:30 - 17:00 Session 2E-I3 (abstracts/résumés - page 229) **Hart House - South Dining**

Copula Theory and Methods

Théorie et méthodes liées aux copules

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Ruodu Wang (University of Waterloo)

- 15:30 - 16:00 Claudia Czado (Technische Universität München)
Vines: Building Multivariate Copulas Using Pair Copula Constructions / Vignes : Construction de copules multivariées au moyen de copules construites par paires  
- 16:00 - 16:30 Haijun Li (Washington State University)
Tail Densities of Copulas / Densités des queues de copules  
- 16:30 - 17:00 Emiliano Valdez (Michigan State University), Edward W. Frees (University of Wisconsin - Madison)
A Revisit of Hierarchical Insurance Claims Modeling / Réexamen de la modélisation hiérarchique de réclamations d'assurance  

15:30 - 17:00 Session 2E-I4 (abstracts/résumés - page 231) **Hart House - East Common**

Recent Advances in the Analysis of Complex Surveys

Récents progrès en analyse d'enquêtes complexes







Invited Paper Session / Présentations sur invitation

Chair/Présidente: Wendy Lou (University of Toronto)

Organizers/Responsables: Victoria Landsman and/et Wendy Lou (University of Toronto)

Session sponsored by the ICSA – Canada Chapter

Séance parrainée par le chapitre canadien de l'ICSA

- 15:30 - 16:00 Barry Graubard (National Institute of Health), Monroe Sirken (Retired)
Estimating Sibling Recurrence Risk in Population Sample Surveys / Estimation du risque de récurrence chez la fratrie dans des enquêtes sur la population  
- 16:00 - 16:30 Victoria Landsman, Lou Wendy & Fan Lingling (University of Toronto)
Methods to Address Bias in Web Surveys / Méthodes pour tenir compte du biais dans les enquêtes en ligne  
- 16:30 - 17:00 Sunghee Lee, Tuba Suzer Gurtekin & Michael Elliott (University of Michigan)
Sensitivity Analysis of Seed Selection in Respondent Driven Sampling / L'analyse de sensibilité de la sélection des répondants initiaux dans un échantillonnage fondé sur les répondants  

15:30 - 17:00 Session 2E-I5 (abstracts/résumés - page 233) **UC - 140**

Statistical Challenges in Modern Astronomy







Problèmes statistiques en astronomie moderne

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Jason Loeppky (University of British Columbia)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 15:30 - 16:00 Earl Lawrence & Scott Vander Wiel (Los Alamos National Laboratory), Casey Law (University of California), Geoffrey Bower (Academia Sinica)
Searching for Fast Radio Bursts with the Very Large Array / À la recherche de sursauts radioélectriques rapides à l'aide du Very Large Array  
- 16:00 - 16:30 Shane Reese & Brittany Spencer (Brigham Young University)
Hierarchical Gaussian Process Modeling of Supernova Light Curves / Modélisation gaussienne hiérarchique de courbes de lumière de supernova  
- 16:30 - 17:00 Erik Rosolowsky (University of Alberta), Jason Loeppky (University of British Columbia Okanagan)
Critiquing the Statistical Analysis of Star Formation / Critique de l'analyse statistique de la formation d'étoiles  







15:30 - 17:00 Session 2E-I6 (abstracts/résumés - page 235) **UC - 179**

Statistics of Sea Surface Winds and Non-Gaussianity in Climate Variability

Statistique des vents de surface de la mer et non-gaussianité de la variabilité climatique

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Priscilla Greenwood (University of British Columbia)

- 15:30 - 16:00 Adam Monahan (University of Victoria)
Stochastic Dynamics of Near-Surface Winds / Les données stochastiques dynamiques des vents de surface  
- 16:00 - 16:30 Philip Sura (Florida State University), Abdel Hannachi (Stockholm University)
Perspectives of Non-Gaussianity in Weather and Climate Variability / Perspectives des statistiques non gaussiennes sur la variabilité des phénomènes météorologiques et climatiques  
- 16:30 - 17:00 William Thompson (University of British Columbia), Adam H. Monahan (University of Victoria), Daan Crommelin (Centrum Wiskunde and Informatica)
Parametric Estimation of Sea Surface Wind Dynamics via the Infinitesimal Generator / Estimation paramétrique de la dynamique des vents de surface de la mer par le générateur infinitésimal  

15:30 - 16:45 Session 2E-C1 (abstracts/résumés - page 237) **UC - 152**

Actuarial Science and Finance 2

Actuariat et finance 2

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Alison Gibbs (University of Toronto)

- 15:30 - 15:45 Dezhao Han, Ken Seng Tan & Chengguo Weng (University of Waterloo)
Sparse Vine Copula with Applications to Financial Data / Applications des copules en vignes parcimonieuses à des données financières  
- 15:45 - 16:00 Zied Ben Salah (Concordia University)
On a Generalization of the Expected Discounted Penalty Function to Include Deficits at and Beyond Ruin / Généralisation de la fonction de pénalité escomptée prévue pour tenir compte des déficits au moment de la faillite et au-delà  
- 16:15 - 16:30 Jianxi Su & Edward Furman (York University)
On a Multiple-Shock Dependence Structure / Structure de dépendance multi-chocs  
- 16:30 - 16:45 Julien Trufin & Ilie-Radu Mitric (Laval University)
On a Risk Measure Inspired from the Ruin Probability and the Expected Deficit at Ruin / Une mesure du risque inspirée de la probabilité de faillite et du déficit prévu à la faillite  

15:30 - 17:00 **Session 2E-C2** (abstracts/résumés - page 239) **UC - 244**

Biostatistics - New Inferential Strategies
Biostatistique - nouvelles stratégies inférentielles







Contributed Paper Session / Séance de communications libres
 Chair/Président: Paul Peng (Queen's University)





- 15:30 - 15:45 Parisa Gavanji, Bingshu Chen & Wenyu Jiang (Queen's university)
 Residual Bootstrap for Testing the Treatment and Biomarker Interaction for Unknown Cut-off Point / Bootstrap résiduel pour tester le traitement et l'interaction des biomarqueurs pour une limite inconnue  
- 15:45 - 16:00 Nathalie Moon, Richard Cook & Leilei Zeng (University of Waterloo)
 The Impact Of Prevalent Cohort Sampling on Multistate Transition Probabilities in Chronic Disease Processes / L'impact de l'échantillonnage de cohortes prévalentes sur les probabilités de transition dans des modèles multi-états de maladies chroniques  
- 16:00 - 16:15 Qian Shi & Linglong Kong (University of Alberta)
 Variable Screening Based On Combining Quantile Regression / Filtrage de variables par régression quantile combinée  
- 16:15 - 16:30 Yukun Zhang & Xuewen Lu (University of Calgary)
 Variable Selection in Log-linear Birnbaum-Saunders Regression Model for High-dimensional Survival Data via the Elastic Net and Stochastic EM / Sélection de variables dans le modèle de régression de Birnbaum-Saunders loglinéaire pour les données de survie de grandes dimensions via le filet élastique et Stochastic EM  
- 16:30 - 16:45 Michael Wallace & Erica Moodie (McGill University)
 A New Method for Identifying Dynamic Treatment Regimes: Introducing weighted Q-learning / Nouvelle méthode d'identification des régimes de traitement dynamiques : Q-learning pondéré  
- 16:45 - 17:00 Tolulope Sajobi (University of Calgary), Lisa M. Lix (University of Manitoba), Gurbakhsh Singh & Gordon H. Fick (University of Calgary)
 Statistical Inference on Discriminant Function Coefficients in the Presence of Non-Normality and Covariance Heterogeneity / Inférence statistique sur les coefficients de fonctions discriminantes en la présence de non-normalité et d'hétérogénéité de la covariance  

15:30 - 16:45 **Session 2E-C3** (abstracts/résumés - page 243) **UC - 256**

Business and Industrial Statistics 2
Statistique industrielle et de gestion 2

Contributed Paper Session / Séance de communications libres
 Chair/Président: Derek Bingham (Simon Fraser University)

- 15:30 - 15:45 Arden Ewan Miller (University of Auckland)
 Error Rates for the Analysis of Unreplicated Experiments / Taux d'erreurs pour l'analyse de plans sans réplicats  
- 15:45 - 16:00 Paule Bodson-Clermont (Université de Montréal), François Lafleur (Institut de recherche d'Hydro-Québec), Jean-François Angers (Université de Montréal), Luc Perreault (Institut de recherche d'Hydro-Québec)
 Statistical Modeling of Cavitation Erosion within the Operation Parameters of a Hydraulic Turbine / Modélisation de l'érosion de cavitation d'une turbine hydraulique selon les paramètres d'opération  
- 16:00 - 16:15 Janie Coulombe (Université de Montréal)
 Assessment of Wind Speed Prediction Error in Wind Energy / Évaluation de l'erreur de prévision de la vitesse du vent en énergie éolienne  

- 16:15 - 16:30 Pavel Krupskiy & Harry Joe (University of British Columbia)
Structured Factor Copula Models for Multivariate Data / Modèles de copule pour des données multivariées avec facteurs structurés  
- 16:30 - 16:45 Paula Murray, Ryan Browne & Paul McNicholas (University of Guelph)
The Hidden Truncation Hyperbolic Distribution / Distribution hyperbolique d'une troncature dissimulée  

15:30 - 17:00 Session 2E-C4 (abstracts/résumés - page 246) **UC - Croft House**

Probability Theory and Finance

Théorie des probabilités et finance

Contributed Paper Session / Séance de communications libres

Chair/Président: Deniz Sezer (University of Calgary)

- 15:30 - 15:45 Katherine Davies (University of Manitoba), N. Balakrishnan (McMaster University)
Pitman Closeness Results for Type-I Censored Data from Exponential Distribution / Résultats sur le critère de proximité de Pitman pour des données censurées de type-I provenant d'une distribution exponentielle  
- 15:45 - 16:00 Deli Li (Lakehead University)
A Characterization of Chover-Type Law of Iterated Logarithm / Caractérisation de la loi du logarithme itéré de type Chover  
- 16:00 - 16:15 Éric Marchand & Djilali Ait Aoudia (Université de Sherbrooke), François Perron (Université de Montréal), Latifa Ben Hadj Slimene (Université de Sherbrooke)
On Runs, Bivariate Poisson Mixtures and Distributions That Arise in Bernoulli Arrays / À propos du nombre de succès consécutifs, de mélanges de lois de Poisson bivarié et de tableaux de Bernoulli  
- 16:15 - 16:30 Felix Camirand & Taoufik Bouezmarni (Université de Sherbrooke), Jean-François Quessy (Université du Québec à Trois-Rivières)
Empirical Copula Processes for Time Series Data / Processus de copule empirique pour des séries chronologiques  
- 16:30 - 16:45 Alexandre Scott (University of Western Ontario), Adam Metzler (Wilfrid Laurier University)
Simulating Rare Event Using a Two Stage IS Estimator / Échantillonnage d'importance en deux étapes pour simuler des événements rares  
- 16:45 - 17:00 Francois Watier & Imène Allab (Université du Québec à Montréal)
First Passage Time Estimation of Diffusion Processes Through Time-varying Boundaries / Estimation du temps de premier passage dans les processus de diffusion à travers des limites variant dans le temps  





15:30 - 16:30 Session 2E-C5 (abstracts/résumés - page 249) **UC - 144**



Robust and Shrinkage Estimation



Estimation robuste et par rétrécissement

Contributed Paper Session / Séance de communications libres

Chair/Président: Alex Leblanc (University of Manitoba)

- 15:30 - 15:45 Fuqi Chen & Sévérien Nkurunziza (The University of Windsor)
A Class of Shrinkage-Type Estimators in Multivariate Linear Regression with Change-Points / Classe d'estimateurs à rétrécissement dans un modèle de régression multivarié avec points de rupture  
- 15:45 - 16:00 Debbie Dupuis, Nicolas Papageorgiou & Bruno Rémillard (HEC Montréal)
Robust Conditional Variance and Value-at-Risk Estimation / Estimation robuste de la variance conditionnelle et de la valeur à risque  

16:00 - 16:15 David Haziza (Université de Montréal), Cyril Favre Martinoz (CREST-ENSAI), Jean-François Beaumont (Statistique Canada)
Robust Estimation for Asymmetric Infinite Populations / Estimation robuste pour des populations infinies asymétriques  

16:15 - 16:30 Marcos Sanches (Centre for Addiction and Mental Health (CAMH))
Regression Discontinuity with Age as Running Variable – Issues with Statistical Tests and Comparison with Time Series Intervention Analysis / Discontinuité par régression avec l'âge comme variable d'ordre, tests statistiques et comparaison avec l'analyse des séries chronologiques  

17:00 - 18:00 **UC 256**
NSERC Discovery Grant Application Workshop
Atelier sur les demandes de subventions à la découverte du CRSNG

17:00 - 18:00 **South Dining Room, Hart House**
Actuarial Science Section General Meeting
Groupe d'actuariat assemblée générale

17:00 - 18:00 **East Common Room, Hart House**
Biostatistics Section General Meeting
Groupe de biostatistique assemblée générale

17:00 - 18:00 **Debates Room, Hart House**
Business and Industrial Statistics Section General Meeting
Groupe de statistique industrielle et de gestion assemblée générale

17:00 - 18:00 **Music Room, Hart House**
Probability Section General Meeting
Groupe de probabilité assemblée générale

17:00 - 18:00 **UC 144**
Statistical Education Section General Meeting
Groupe d'éducation en statistique assemblée générale

17:00 - 18:00 **UC 152**
Student Research Presentation Award Committee
Comité du prix pour la présentation de recherche étudiante

17:00 - 18:00 **UC 52**
Survey Methods Section General Meeting
Groupe de méthodes d'enquête assemblée générale


18:00 - 22:00 **Board busses in Hart House Circle**
Banquet

Wednesday May 28**mercredi 28 mai**

8:45 - 9:50 Session 3A-A1 (abstract/résumé - page 251) Hart House - Music Rm
CRM-SSC Prize in Statistics Address**Allocution du récipiendaire du Prix CRM-SSC en statistique**

Special Session / Conférence spéciale



Organizer and Chair / Responsable et président: Jose Garrido (Concordia University)

8:45 - 9:50 Fang Yao (University of Toronto)
 Functional Data Analysis and Related Topics / Analyse de données fonctionnelles et sujets connexes  

8:45 - 9:50 Session 3A-A2 (abstract/résumé - page 252) Hart House - Debates Rm
SSC Impact Award Address**Allocution du récipiendaire du Prix pour impact de la SSC**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et présidente: Thérèse Stukel (Institute for Clinical Evaluative Sciences)

8:45 - 9:50 David Thomson (Queen's University)
 A Time Series Analyst's Look at Climate Data / Regard d'un analyste de séries chronologiques sur les données climatiques  

9:50 - 10:20 Great Hall
Coffee break**Pause-café**

10:20 - 11:50 Session 3B-I1 (abstracts/résumés - page 253) Hart House - East Common
Bioinformatics and Copula Models**Bioinformatique et modèles de copules**



Invited Paper Session / Présentations sur invitation



Chair/Présidente: Ying Zhang (Acadia University)



Organizer/Responsable: Laura Cowen (University of Victoria)

Session sponsored by the Committee on Women in Statistics

Séance parrainée par le Comité des femmes en statistique

10:20 - 10:50 Kelly Burkett (University of Ottawa)
 Using Gene Genealogies to Detect Rare Disease-Predisposing Genetic Variants / Détection de variantes génétiques prédisposant à des maladies rares à l'aide de la généalogie génétique  









10:50 - 11:20 Jemila Hamid & Sayantee Jana (McMaster University)
 A Test for the Growth Curve Model and its High-Dimensional Extension: Applications in Genomics / Test pour le modèle de courbe de croissance et sa généralisation de grande dimension : applications en génomique  

11:20 - 11:50 Michelle Zhou (Simon Fraser University), Shulin Zhang (Southwestern University of Finance and Economics, China), Ostap Okhrin (Humboldt University of Berlin, Germany), Peter Song (University of Michigan, USA)
 Goodness-of-fit Test for Specification of Copula-Based Semi-parametric Models / Test de validité de l'ajustement pour la spécification de modèles semi-paramétriques fondés sur les copules  

10:20 - 11:48 Session 3B-I2 (abstracts/résumés - page 255) Hart House - Debates Rm

Modern Statistical Methods for Complex Health Data
Méthodes statistiques modernes pour données sanitaires complexes







Invited Paper Session / Présentations sur invitation
 Chair/Président: Kevin He (University of Michigan)
 Organizers/Responsables: Kevin He and/et Yi Li (University of Michigan)
 Session sponsored by the Biostatistics Section
 Séance parrainée par le Groupe de biostatistique

- 10:20 - 10:42 Richard Cook (University of Waterloo)
 Statistical Challenges Arising in Transfusion Research / Défis statistiques en recherche transfusionnelle  
- 10:42 - 11:04 John Kalbfleisch, Mathieu Bray, Peter Song, Wen Wang, Alan Leichtman & Michael Rees (University of Michigan)
 Optimization in a Kidney Paired Donation Program / Optimisation d'un programme de transplantation rénale utilisant donneurs vivants jumelés par échange de bénéficiaires  
- 11:04 - 11:26 Yi Li, Ji Zhu & Kevin He (University of Michigan)
 Gateaux Differential-Based Boosting for Time-Varying Effects in Survival Analysis / Boosting de Gâteaux basé sur le différentiel pour les effets variant dans le temps en analyse de survie  
- 11:26 - 11:48 Ross Prentice (Fred Hutchison Cancer Research Center)
 Nonparametric and Semiparametric Analysis of Bivariate Failure Time Data / Analyse non paramétrique et semi-paramétrique de données de temps de défaillance bivariées  

10:20 - 11:50 Session 3B-I3 (abstracts/résumés - page 258) Hart House - Music Rm

New Challenges and Developments in Survival Analysis and its Applications
Nouveaux problèmes et progrès en analyse de survie et ses applications







Invited Paper Session / Présentations sur invitation
 Organizer and Chair / Responsable et présidente: Jiajia Zhang (University of South Carolina)

- 10:20 - 10:50 Zhezhen Jin (Columbia University)
 Nonparametric Threshold Selection with Censored Survival Data / Sélection d'un seuil non paramétrique pour des données de survie censurées  
- 10:50 - 11:20 Yingwei Peng (Queen's University), Yi Niu (Dalian University of Sciences and Technology)
 Mixture Cure Models for Clustered Survival Data / Modèles de mélange avec fraction non à risque pour des données de survie en grappes  
- 11:20 - 11:50 Grace Yi & Yang Ning (University of Waterloo), Nancy Reid (University of Toronto)
 Weighted Estimating Equations for Semiparametric Transformation Models With Missing Covariates / Équations d'estimation pondérées pour des modèles de transformation semi-paramétriques avec covariables manquantes  

10:20 - 11:50 Session 3B-I4 (abstracts/résumés - page 260) Hart House - South Dining

New Developments in Statistical Neuroscience
Récents progrès en neuroscience statistique

Invited Paper Session / Présentations sur invitation
 Organizer and Chair / Responsable et président: Shoja'eddin Chenouri (University of Waterloo)







- 10:20 - 10:50 Sam Behseta (California State University)
A Semiparametric Bayesian Model for Detecting Multiway Synchrony Among Neurons / Modèle bayésien semi-paramétrique pour la détection de la synchronie multivoies entre neurones  
- 10:50 - 11:20 Matthew Harrison (Brown University)
Robust Inference for Nonstationary Spike Trains / Inférence robuste pour les trains d'impulsion neuronaux non stationnaires  
- 11:20 - 11:50 Paul Marriott (University of Waterloo)
The Analysis of Multiple Spike Train Data / Analyse de données sur les trains d'impulsion multiple  

10:20 - 11:50 Session 3B-I5 (abstracts/résumés - page 262) **UC - Croft House**

Statistical Education in Light of Philosophy of Science (A Panel Discussion)

Philosophie de la science et éducation statistique

Invited Paper Session / Présentations sur invitation
Chair/Président: Kevin Thorpe (University of Toronto)
Organizer/Responsable: Jeff Picka (University of New Brunswick)
Session sponsored by the Statistics Education Section
Séance parrainée par le Groupe d'éducation en statistique







- 10:20 - 11:50 Maya Goldenberg (University of Guelph)
Philosophy of Science Meets Evidence-Based Medicine / La philosophie de la science à la rencontre de la médecine factuelle  
- 10:20 - 11:50 Jeff Picka (University of New Brunswick)
Discussion of Using Science Studies Resources in Statistical Education / Discussion sur l'utilisation des ressources scientifiques pour l'éducation statistique  
- 10:20 - 11:50 Nathan Young (University of Ottawa)
Statistics and the Politics of Expert Knowledge: Understanding the 'Public Life' of Statistical Knowledge, Insights from the Social Sciences / La statistique et les politiques en matière d'expertise : comprendre la 'vie publique' de la connaissance statistique ; perspectives des sciences sociales  

10:20 - 11:50 Session 3B-I6 (abstracts/résumés - page 264) **UC - 179**

Stochastic Modelling in Industry

Modélisation stochastique en industrie

Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et président: Neal Madras (York University)
Session sponsored by the Probability Section and the Actuarial Science Section
Séance parrainée par le Groupe de probabilité et le Groupe de science actuarielle

- 10:20 - 10:50 Guangyu Fu (Scotiabank)
Managing Volatility Smile in the Option Market: SABR, Local Volatility and Other Models / Gestion du «sourire de volatilité» sur le marché des options : modèle SABR, de volatilité locale et autres modèles  
- 10:50 - 11:20 Dan Wilson (Invidi Technologies Corporation)
Challenges of Centralized and Distributed Control for Addressable Advertising / Défis de la commande centralisée et répartie en matière de publicité adressable  
- 11:20 - 11:50 Ping Wu (Bank of Montreal)
Pricing Interest Rate Exotic Under LIBOR Market Model / Tarification de taux d'intérêt exotiques à l'aide du modèle de marché LIBOR  

10:20 - 11:50 Session 3B-C1 (abstracts/résumés - page 266) **UC - 161**

Actuarial Science and Finance 3

Actuariat et finance 3

Contributed Paper Session / Séance de communications libres

Chair/Président: Chengguo Weng (University of Waterloo)

- 10:20 - 10:35 Mélina Mailhot & Amanda Spiroliias (Concordia University)
[CANCELED] Retention Level for an Optimal Reinsurance Treaty / [ANNULÉ] Niveau de ré-tention pour un traité de réassurance optimal  
- 10:35 - 10:50 Peter Adamic (Laurentian University)
Cause-Deleted Life Expectancy Improvement with a Cure Distribution / Amélioration de l'évaluation de l'espérance de vie en traitant les causes de mortalité éliminées avec une distribution de guérison  
- 10:50 - 11:05 Taehan Bae & Andrei Volodin (University of Regina)
Generalization of Three Parameter Crack Distribution for Loss Severity Modeling / Généralisation d'une distribution de fissures à trois paramètres pour la modélisation de la sévérité des pertes  
- 11:05 - 11:20 Jose Garrido (Concordia University), Christian Genest & Juliana Schulz (McGill University)
Generalized Linear Models for Dependent Claim Frequency and Severity / Modèles linéaires généralisés pour fréquence et sévérité de pertes dépendantes  
- 11:20 - 11:35 Jiandong Ren (Western University)
Analysis of IBNR with Markovian Arrival Processes / Analyse des réclamations END par processus d'arrivée de Markov  
- 11:35 - 11:50 Alan Xinghua Zhou & R. Mark Reesor (Western University)
Misrepresentation and Capital Structure: Quantifying the Impact on Corporate Security Value in a Modified Merton Framework / Fausses déclarations et structure du capital : comment quantifier l'impact sur la valeur des titres de société dans un modèle de Merton modifié  







10:20 - 11:50 Session 3B-C2 (abstracts/résumés - page 269) **UC - 140**

Biostatistics - Survival, Longitudinal and Event History Analysis

Biostatistique - analyses de survie, longitudinales et d'historique d'événements

Contributed Paper Session / Séance de communications libres

Chair/Président: Joel Dubin (University of Waterloo)

- 10:20 - 10:35 Gun Ho JANG (Ontario Institute for Cancer Research)
Convergence Rate of Scan Statistics for Rare Copy Number Variation Detection / Taux de convergence des statistiques de scan pour la détection de variations du nombre de copies rares  
- 10:35 - 10:50 Elham Khodayari Moez & Elham Khodayari Moez (University of Alberta), Farid Zayeri (Shahid Beheshti University of Medical Sciences), Irina Dinu (University of Alberta)
The Accuracy of Joint Analysis of Survival and Longitudinal Data: A Simulation Study / Précision de l'analyse conjointe de données de survie et de données longitudinales : étude de simulation  
- 10:50 - 11:05 Eleanor Pullenayegum (Hospital for Sick Children)
Methods for Analysing Irregular Longitudinal Data: How do the Assumptions Match Up with Visit Processes Encountered in Practice / Méthodes d'analyse de données longitudinales irrégulières : les hypothèses correspondent-elles aux processus de visite rencontrés dans la pratique ?  

- 11:05 - 11:20 Denis Talbot, Geneviève Lefebvre & Juli Atherton (Université du Québec à Montréal)
A Cautionary Note on the Use of Stabilized Weights in Marginal Structural Models / Un avertissement concernant l'utilisation des poids stabilisés pour les modèles structureaux marginaux  
- 11:20 - 11:35 Xin Xin, Julie Horrocks & Gerarda Darlington (University of Guelph)
A Likelihood Approach to Accommodate Ties between Events and Covariate Change Times in Cox Proportional Hazards Models / Approche de vraisemblance pour tenir compte des liens entre temps d'événements et de changement de covariables dans les modèles des risques proportionnels de Cox  
- 11:35 - 11:50 Yayuan Zhu, Jerry Lawless & Cecilia Cotton (University of Waterloo)
Event History Analysis in Observational Studies with Intermittent Observation / Analyse de l'historique des événements dans les études d'observation intermittente  

10:20 - 11:50 Session 3B-C3 (abstracts/résumés - page 272)













UC - A101

Some Specific Issues in Modelling and Estimation

Quelques problématiques spécifiques en modélisation et estimation

Contributed Paper Session / Séance de communications libres

Chair/Président: Fang Yao (University of Toronto)

- 10:20 - 10:35 Elif Acar (University of Manitoba)
Conditional Copula Models with Multiple Covariates / Modèles à copules conditionnelles avec covariables multiples  
- 10:35 - 10:50 Othmane Kortbi & Eric Marchand (Université de Sherbrooke)
Estimating a Multivariate Normal Mean with a Bounded Signal to Noise Ratio / Estimation d'une moyenne normale multivariée avec un rapport signal-bruit borné  
- 10:50 - 11:05 Zhenhua Lin (University of Toronto), Jiguo Cao & Liangliang Wang (Simon Fraser University), Haonan Wang (Colorado State University)
A Smooth and Locally Sparse Estimator for Functional Linear Regression via Functional SCAD Penalty / Estimateur lisse et localement parcimonieux pour une régression linéaire fonctionnelle au moyen d'une pénalité fonctionnelle pour l'écart absolu avec coupure lisse (SCAD).  
- 11:05 - 11:20 Iraj Yadegari & Éric Marchand (Université de Sherbrooke), François Perron (Université de Montréal)
Parameter Estimation under Bounded Loss Functions / Estimation de paramètres sous fonctions de pertes limitées  
- 11:20 - 11:35 Tingting Yu & Lang Wu (University of British Columbia)
Jointly Modeling Longitudinal Process with Measurement Errors, Missing Data, and Outliers / Modélisation conjointe de processus longitudinaux avec erreurs de mesure, données manquantes et données aberrantes  
- 11:35 - 11:50 David Riegert, Aaron Springford & David Thomson (Queen's University at Kingston)
Forecasting the Likelihood of Solar Flares Using an Inferred Solar Stress Index / Comment prévoir la vraisemblance d'une éruption solaire à l'aide d'un indice de stress solaire  

12:00 - 13:30

UC 256

Accreditation Committee

Comité d'accréditation

12:00 - 13:30

UC 248

Actuarial Science Section Executive Committee 2

Comité exécutif du groupe d'actuariat 2

12:00 - 13:30 **UC 257**
Probability Section Executive Committee 2
Comité exécutif du groupe de probabilité 2

12:00 - 13:30 **UC 53**
Program Committee
Comité du programme

12:00 - 13:30 **UC 244**
Statistical Education Section Executive Committee 2
Comité exécutif du groupe d'éducation en statistique 2

12:00 - 13:30 **UC 257**
Student and Recent Graduate Committee
Comité des étudiants et diplômés récents

12:00 - 13:30 **UC 52**
Treasurers
Trésoriers

13:30 - 14:30 **Session 3C-A1** (abstract/résumé - page 276) **Hart House - Debates Rm**
CJS Award Address
Allocution du récipiendaire du Prix de la RCS
Special Session / Conférence spéciale
Chair/Président: Jiahua Chen (University of British Columbia)
Organizer/Responsable: Denis Larocque (HEC Montreal)

13:30 - 14:30 Art Owen (Stanford University)
Self-concordance for Empirical Likelihood / Auto-concordance pour la vraisemblance empirique E E

13:30 - 14:30 **Session 3C-A2** (abstract/résumé - page 277) **Hart House - Music Rm**
Isobel Loutit Invited Address
Allocution invitée Isobel Loutit
Special Session / Conférence spéciale
Organizer and Chair / Responsable et président: Thomas Loughin (Simon Fraser University)
Session sponsored by the Business and Industrial Statistics Section
Séance parrainée par le Groupe de statistique industrielle et de gestion

13:30 - 14:30 Bradley Jones (JMP)
What You Learn Depends on Where You Look / Ce que vous apprenez est fonction de ce que vous observez E E

13:30 - 14:30 Session 3C-A3 (abstract/résumé - page 278) Hart House - East Common

Pierre Robillard Award Address

Allocution du récipiendaire du Prix Pierre-Robillard

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Wenqing He (Western University)

13:30 - 14:30 Liqun Diao (University of Rochester)

Copula Models for Multi-type Life History Processes / Modèles de copules pour les processus d'historiques multiples E E

13:30 - 14:30 Session 3C-A4 (abstract/résumé - page 279) UC - 140

Survey Methods Section Presidential Invited Address

Allocution de l'invité du président du Groupe des méthodes d'enquête

Special Session / Conférence spéciale

Organizer and Chair / Responsable et présidente: Susana Rubin-Bleuer (Statistics Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

13:30 - 14:30 Claire Durand (Université de Montréal)

Surveys and Society / Sondages et société E E

14:30 - 15:00

Great Hall

Coffee break

Pause-café

15:00 - 16:30 Session 3D-I1 (abstracts/résumés - page 280) Hart House - South Dining

Bridging the Gap between Theory and Practice in Statistics

Comblant l'écart entre la théorie et la pratique en statistique

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Fernando Camacho (University of Waterloo)

Session sponsored by the Accreditation Committee

Séance parrainée par le Comité d'accréditation

15:00 - 15:30 Michal Abrahamowicz (McGill University)

Flexible Modelling of Cumulative Effects of Medications Use: Fighting with 2 Dragons (Analytical Challenges and Clinical Reviewers) / Modélisation flexible des effets cumulatifs de la consommation de médicaments : combat entre deux dragons (défis analytiques et évaluateurs cliniques) E E

15:30 - 16:00 Pina D'Angelo (Novum Pharmaceutical)

Data Analysis Concepts and Issues in Skin Adhesion and Irritation/Sensitization Studies / Concepts et problématiques d'analyse de données tirées d'études portant sur l'adhérence sur la peau, l'irritation et la sensibilisation E E

16:00 - 16:30 Janet E.A. McDougall, Huang Qianyi, Amrhein John & Chen Hong (McDougall Scientific Ltd.)

Adaptive Randomization in a Seamless Phase II/Phase III Dose-Finding and Confirmatory Trial Design / Randomisation adaptative dans le cadre d'un essai clinique ininterrompu de phases II et III visant l'établissement et la confirmation de la posologie E E







15:00 - 16:30 Session 3D-I2 (abstracts/résumés - page 282) Hart House - East Common**Business Analytics: Where does the Statistician Fit?****Méthodes analytiques de gestion : le rôle du statisticien**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Georges Monette (York University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 15:00 - 15:30 Krzysztof Dzieciolowski (Rogers)
From Small Samples to Big Data / Partir de petits échantillons de données volumineuses  
- 15:30 - 16:00 Kaiser Fung (Vimeo)
The Nature of Big Data and the Importance of Numbersense / Nature des données volumineuses et importance du sens des nombres  
- 16:00 - 16:30 Jeffrey McGill (Queen's University)
Business Analytics, Statistics, and Communication / Analytique d'affaires, statistique et communication  







15:00 - 16:30 Session 3D-I3 (abstracts/résumés - page 284) UC - 163**Evaluating the Effectiveness of our Introductory Courses****Comment évaluer l'efficacité de nos cours d'introduction ?**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Bethany White (Western University)

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

- 15:00 - 15:30 Bob delMas (University of Minnesota)
Trends in Statistical Understanding: Results From CAOS / Tendances en matière de compréhension statistique : résultats du CAOS  
- 15:30 - 16:00 Bruce Dunham & Gaitri Yapa (University of British Columbia)
Calibrating the Difficulty Level of Assessment Tools: ``Blooming" a Statistics Exam / Étalonnage du niveau de difficulté des outils d'évaluation en appliquant la taxonomie de Bloom à un examen de statistique  
- 16:00 - 16:30 Michele Millar (Mount Saint Vincent University), Bethany White (Western University), Candace Schau (CS Consultants)
Do Our Courses Improve Students' Attitudes? / Nos cours améliorent-ils l'attitude des étudiants ?  



15:00 - 16:30 Session 3D-I4 (abstracts/résumés - page 286) UC - 179**Levy Processes and Applications****Processus de Lévy et leurs applications**





Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Xiaowen Zhou (Concordia University)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

- 15:00 - 15:30 Alexey Kuznetsov & Daniel Hackmann (York University)
Numerical methods for Levy processes / Méthodes numériques pour des processus de Lévy  







- 15:30 - 16:00 Andreas Kyprianou (University of Bath, Great Britain), Juan Carlos Pardo (CIMAT, Mexico), Alex Watson (University of Zurich)
Censored Stable Processes / Processus censurés stables  
- 16:00 - 16:30 Jean-Francois Renaud (Université du Québec à Montréal), Xiaowen Zhou (Concordia University), Ronnie Loeffen (University of Manchester), Hélène Guérin (Université de Rennes 1)
Occupation Times for Spectrally Negative Lévy Processes with Applications / Temps d'occupation pour des processus de Lévy spectralement négatifs et applications  

15:00 - 16:30 Session 3D-I5 (abstracts/résumés - page 288) **UC - 161**

Predictive Density Estimation
Estimation de densité prédictive

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Eric Marchand (Université de Sherbrooke)







- 15:00 - 15:30 Fumiyasu Komaki (University of Tokyo)
Prior Construction Based on the Predictive Metric / Construction de lois a priori basées sur des mesures prédictives  
- 15:30 - 16:00 William Strawderman (Rutgers University), Tatsuya Kubokawa (University of Tokyo), Eric Marchand (Université de Sherbrooke)
Predictive Density Estimation Under Integrated L1 and L2 Loss / Estimation de densité prédictive sous des pertes intégrées L1 et L2  
- 16:00 - 16:30 Xinyin Xu (Ohio State University), Edward George (University of Pennsylvania), Feng Liang (University of Illinois at Urbana-Champaign)
From Minimax Shrinkage Estimation to Minimax Shrinkage Prediction / De l'estimation minimax par rétrécissement à la prévision minimax par rétrécissement  

15:00 - 16:30 Session 3D-I6 (abstracts/résumés - page 290) **Hart House - Music Rm**

Statistical Analysis for Infectious Diseases
Analyse statistique des maladies infectieuses

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Juxin Liu (University of Saskatchewan)

- 15:00 - 15:30 Rob Deardon, Scott Hunt, Nadia Bifulchi & Longyao Cai (University of Guelph)
Optimal Experimental and Study Design for Infectious Disease Systems of Animals / Plan d'étude et d'expérience optimal pour des systèmes de pathologies infectieuses animales  
- 15:30 - 16:00 Nathaniel Osgood, Juxin Liu & Seth Dueck (University of Saskatchewan), David Vickers (Alberta Health Services)
Combining MCMC and Compartmental Modeling to Enhance Understanding of Chlamydia Control in Saskatchewan / Combinaison de MCMC et de modélisation compartimentale en vue d'améliorer la compréhension du contrôle de la chlamydia en Saskatchewan  
- 16:00 - 16:30 Erin Schliep (Duke University), Daniel Gear (United States Department of Agriculture), Colleen Webb (Colorado State University)
Modeling Prevalence of Bovine Tuberculosis Using Multiple-Phase Testing / Modélisation de la prévalence de tuberculose bovine à l'aide de tests à phases multiples  

15:00 - 16:30 Session 3D-C1 (abstracts/résumés - page 292) **UC - A101**

Actuarial Science and Finance 4
Actuariat et finance 4

Contributed Paper Session / Séance de communications libres

Chair/Président: Adam Metzler (Wilfrid Laurier University)

- 15:00 - 15:15 Ruodu Wang (University of Waterloo), Valeria Bignozzi (ETH Zurich), Andreas Tsanakas (City University London)
How Superadditive Can a Risk Measure Be? / À quel point une mesure de risque peut-elle être suradditive ?  
- 15:15 - 15:30 Sungsoo Kim & Steve Drekic (University of Waterloo)
Ruin Analysis in a Discrete-time Sparre Andersen Model with External Financial Activities and Random Dividends / Analyse de ruine dans un modèle à temps discret de Sparre Andersen comprenant des activités financières et des dividendes aléatoires  
- 15:30 - 15:45 Jean-Philippe Boucher & Rofick Inoussa (Université du Québec à Montréal)
A Posteriori Ratemaking with Panel Data / Tarification a posteriori avec données de panel  
- 15:45 - 16:00 Nicolas Essis-Breton & Patrice Gaillardetz (Concordia University)
Parallel Computing Methods for Variable Annuities / Méthodes de calcul parallèle pour les rentes variables  
- 16:00 - 16:15 Geneviève Gauthier, Debbie Dupuis & Frédéric Godin (HEC Montréal)
Short-term Hedging for an Electricity Retailer / Stratégie de couverture pour un revendeur d'électricité  
- 16:15 - 16:30 David Edward Wilson (University of Waterloo)
Volatility and Intensity / Volatilité et intensité  











15:00 - 16:15 **Session 3D-C2** (abstracts/résumés - page 295) **Hart House - Debates Rm**

Biostatistics - Variable Selection

Biostatistique - sélection de variables

Contributed Paper Session / Séance de communications libres

Chair/Président: Martin Lysy (University of Waterloo)

- 15:00 - 15:15 Yeying Zhu (University of Waterloo), Donna Coffman & Debashis Ghosh (Pennsylvania State University)
A Boosting Algorithm for Estimating Generalized Propensity Scores / Algorithme d'augmentation pour l'estimation des scores de propension généralisés  
- 15:15 - 15:30 Halimah Asiri & Jianan Peng (Acadia University)
Simultaneous Inference for Low Dose Risk Estimation with Abbott-adjusted Weibull Model in Benchmark Analysis / Inférence simultanée pour l'estimation du risque à faible dose avec modèle de Weibull ajusté selon la formule d'Abbott pour l'analyse d'un banc d'essais  
- 15:30 - 15:45 Yang Ning & Grace Yi (University of Waterloo), Baojiang Chen (University of Nebraska Medical Center), Nancy Reid (University of Toronto)
Variable Selection and Inference for Ultra-High Dimensional Survival Data With Missing Covariates Under Proportional Hazards Models / Sélection de variable et inférence pour des données de survie en très grande dimension avec des covariables manquantes dans des modèles à risques proportionnels  
- 15:45 - 16:00 Saadia Pinky & Giseon Heo (University of Alberta)
Hypothesis Testing with Power Calculation and Variable Selection with Sparse Dirichlet-Multinomial in Clostridium Difficile data / Tests d'hypothèse avec calcul de puissance et sélection de variable pour une régression multinomiale de Dirichlet parcimonieuse avec des données portant sur le Clostridium difficile  
- 16:00 - 16:15 Farhad Shokoohi, Abbas Khalili & Masoud Asgharian (McGill University)
Variable Selection in Mixture of Survival Models / Sélection de variables dans les modèles de mélange de survie  

16:45 - 18:00**Board of Directors 2
Conseil d'administration 2****UC - Croft Chapter House**

18:00 - 19:00**Executive Committee 2
Comité exécutif 2****UC - Croft Chapter House**

Abstracts • Résumés

1A-A2: SSC Presidential Invited Address
Allocution de l'invité du président de la SSC

Organizer and Chair / Responsable et président: Mike Evans (University of Toronto)

Room/Salle: MedSci Auditorium

Abstract/Résumé

MICHAEL JORDAN (University of California Berkeley), **VENKAT CHANDRASEKARAN** (Caltech), **JOHN DUCHI**, **MARTIN WAINWRIGHT & YUCHEN ZHANG** (University of California, Berkeley)

[Monday May 26 / lundi 26 mai, 8:45]

On the Computational and Statistical Interface and "Big Data"

Statistique, informatique et l'inférence à grande échelle

The rapid growth in the size and scope of datasets in science and technology has created a need for novel foundational perspectives on data analysis that blend the statistical and computational sciences. A particularly critical need is that of obtaining guarantees on statistical risk that incorporate notions of computational complexity, and in particular of runtime. Taking statistical decision theory as a point of departure, we treat "externalities" such as computation, communication and privacy as constraints on estimation procedures, and explore the tradeoffs that result.

La croissance rapide de la taille et de l'étendue des collections de données en science et technologie a créé un besoin pour des perspectives fondamentales nouvelles qui mélangent les sciences statistiques et les sciences informatiques. Un besoin particulièrement crucial est celui d'obtenir des garanties sur le risque statistique qui prennent en compte des notions de complexité algorithmique, et en particulier le temps d'exécution. Prenant comme point de départ la théorie de la décision statistique, on traite des «externalités» telles que le calcul, la communication et la vie privée comme contraintes sur les procédures d'estimation, et nous étudions les compromis qui résultent.

1B-I1: Big Data Données volumineuses

Organizer and Chair / Responsable et présidente: Susana Rubin-Bleuer (Statistics Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: Hart House - Debates Rm

Abstracts/Résumés

HUGH CHIPMAN (Acadia University), **ROBERT MCCULLOCH** (University of Chicago), **MATT PRATOLA** (Ohio State University), **DAVE HIGDON** (Los Alamos National Laboratory), **STEVE SCOTT** (Google), **EDWARD GEORGE** (University of Pennsylvania)

[Monday May 26 / lundi 26 mai, 10:20]

Ensemble Learning for Big Data

Apprentissage par ensembles pour données volumineuses

Ensembles are powerful supervised learning methods that combine together many individual models in an adaptive manner. Bayesian Additive Regression Trees (BART) are a successful statistical version of ensemble models, offering full inference. However, computation can be expensive for big data, especially for MCMC. We discuss two different approaches to parallelizing BART, using either multiple processes within each MCMC step, or parallel MCMC runs on independent machines.

Les méthodes d'ensembles pour l'apprentissage supervisé combinent bon nombre de modèles individuels de façon adaptative et sont ainsi très efficaces. Les arbres additifs de régression bayésienne (BART) sont une version statistique des modèles d'ensembles qui permettent une inférence complète. Par contre, les calculs nécessaires peuvent être lourds pour des données volumineuses, notamment en ce qui a trait aux MCMC. Nous discutons deux approches différentes pour paralléliser BART qui utilisent soit des processus multiples à chaque étape MCMC ou des itérations parallèles de MCMC sur des noeuds indépendants.

STEVEN HEERINGA (University of Michigan), **ROBERTO EDMUNDO MELLIPILAN** (University of Michigan, Program in Survey Methods)

[Monday May 26 / lundi 26 mai, 10:50]

Survey Research and Big Data: Estrangement or a Long-term Relationship?

Recherche-enquête et données volumineuses : éloignement ou relation à long terme?

Contemporary advances in large scale data acquisition, compilation and analysis have many statisticians questioning what the future role for traditional surveys will be (Groves,2011; Couper, 2014). This talk argues that survey designs and methods are in a long term partnership with large scale data resources. Relative to the past where "big data" often served to calibrate complex survey measurements, future surveys may increasingly be designed to fill in information on unmeasured relationships and address the errors in our big data

Face aux avancées contemporaines en matière d'acquisition de données à grande échelle, de compilation et d'analyse, bon nombre de statisticiens se questionnent sur le rôle futur des enquêtes traditionnelles (Groves, 2011 ; Couper, 2014). Cet exposé montre que le concept d'enquête et les méthodes utilisées sont compatibles à long terme avec les ressources conçues pour les données volumineuses. En comparaison avec les données volumineuses qui, dans le passé, servaient souvent à calibrer des mesures d'études complexes, les études futures pourraient être de mieux en mieux conçues pour combler l'information manquante sur

1B-I1: Big Data Données volumineuses

models of social, economic and public health outcomes. Four case studies conducted over the past 40 years will be used to demonstrate that this trend toward big data driving survey design, measurement and analysis is by no means entirely new.

des relations non mesurées et corriger des résultats erronés de nos modèles de données volumineuses dans les domaines social, économique et de santé publique. Quatre études de cas réalisées au cours des 40 dernières années serviront à démontrer que cette tendance à utiliser les données volumineuses dans la conception, les mesures et l'analyse d'enquête n'a rien de nouveau.

JEAN PIGNAL (Statistics Canada), **FERHANA ANSARI**, **LILY MA** (Statistics Canada)

[Monday May 26 / lundi 26 mai, 11:20]

Big Data and Official Statistics

Les données volumineuses et les statistiques officielles

Big Data are an emerging source of business analytics, but many national statistical organizations are still engaged in discovering what it means for them. This presentation will discuss some of the "Big Data" initiatives currently underway at Statistics Canada as well as some of the methodological, technological, human resource and legal challenges inherent in Big Data analytics for official statistics.

Les données volumineuses sont une source analytique importante pour les entreprises, mais plusieurs organismes statistiques nationaux sont toujours en cours de découvrir ce que cela signifie pour eux. Cette présentation discutera quelques-unes des initiatives basées sur les «données volumineuses» mises en œuvre par Statistique Canada avec un œil visé sur les défis méthodologique, technique, de main d'œuvre et légale qui sont spécifiques aux données volumineuses.

Organizer and Chair / Responsable et président: Ruslan Salakhutdinov (University of Toronto)

Room/Salle: Hart House - East Common

Abstracts/Résumés

ALI GHODSI, MICHAEL BIGGS, STEPHEN VAVASIS & TAMEEM HESHAM (University of Waterloo)

[Monday May 26 / lundi 26 mai, 10:20]

Nonnegative Matrix Factorization via Rank-One Dwndate and its Application in Sum-Product Networks Structure Learning

Factorisation de matrices non négatives par l'algorithme de recul de rang un et ses applications dans l'apprentissage de la structure de réseaux sommes et produits

Nonnegative matrix factorization (NMF) attempts to approximate a matrix with nonnegative entries by a product of two low-rank matrices, also with nonnegative entries. In this talk, I introduce an algorithm called rank-one dwndate (R1D) for computing an NMF, as well as its application in Sum-Product Networks (SPN) structure learning.

This algorithm computes the dominant singular values and vectors of adaptively determined subma-

trices of a matrix. On each iteration, R1D extracts a rank-one submatrix from the original matrix according to an objective function. I establish a theoretical result that maximizing this objective function corresponds to correctly classifying articles in a nearly separable corpus. I also provide computational experiments showing the success of this method in identifying features in realistic datasets.

La factorisation non négative de matrices (FNM) tente d'approximer une matrice dont les entrées sont non négatives par un produit de deux matrices de faible rang, possédant elles aussi des entrées négatives. Dans cet exposé, je présente un nouvel algorithme appelé recul de rang un (RR1) pour le calcul d'une FNM, ainsi que son utilisation dans l'apprentissage de la structure de réseaux sommes et

produits (RSP).

Cet algorithme calcule les valeurs et les vecteurs singuliers dominants de sous-matrices adaptatives d'une matrice. À chaque itération, RR1 extrait une sous-matrice de rang un de la matrice originale selon une fonction objective. J'établis un résultat théorique à l'effet que la maximisation de cette fonction objective correspond à classier correctement les articles d'un corpus quasi-séparable. Je fournis aussi des expériences numériques montrant le succès de cette méthode pour l'identification de caractéristiques dans un jeu de données réaliste.

PETER ORBANZ (Columbia University)

[Monday May 26 / lundi 26 mai, 10:50]

Random Walk Models of Graphs and Networks

Modèles de marche aléatoire de graphes et de réseaux

The best-understood statistical models of graphs and networks are models based on exchangeable graphs. These are tractable by the standards of random graph models, but inherently (and provably) misspecified for network data -- a graph generated from such a model would bear no resemblance to most networks arising in applications.

Les modèles statistiques de graphes et de réseaux les mieux compris sont ceux basés sur des graphes échangeables. Ceux-ci s'expriment sous une forme analytique simple selon les standards des modèles de graphes aléatoires, mais ils sont essentiellement inadaptés aux données en réseaux -- un graphe généré à partir d'un tel modèle ne ressemblerait en rien à la plupart des réseaux provenant des applica-

1B-I2: Machine Learning Apprentissage machine

Developing non-exchangeable proves much more difficult, both mathematically and in terms of inference. In my talk, I will review why we are interested in non-exchangeable models and why inference in such models is so difficult. I will explain why random walks play a key role in the mathematical theory of non-exchangeable models, and present a class of models based on random walks for which the likelihood is analytically tractable.

tions. L'élaboration de modèles non échangeables s'avère beaucoup plus difficile, tant d'un point de vue mathématique qu'en matière d'inférence. Dans cet exposé, je présente les raisons qui motivent notre intérêt à l'égard des modèles non échangeables et j'explique pourquoi l'inférence est si difficile à réaliser dans ces modèles. J'explique le rôle clé des marches aléatoires dans la théorie mathématique des modèles non échangeables et je présente une classe de modèles basés sur les marches aléatoires pour lesquels la vraisemblance s'exprime sous une forme analytique simple.

RUSLAN SALAKHUTDINOV & NITISH SRIVASTAVA (University of Toronto)

[Monday May 26 / lundi 26 mai, 11:20]

Multimodal Learning with Deep Boltzmann Machines

Apprentissage multimodal profond à l'aide de machines de Boltzmann

Data often consists of multiple diverse modalities: images are tagged with textual information and videos are accompanied by audio. Each modality is characterized by having distinct statistical properties. We propose a Deep Boltzmann Machine for learning a generative model of such multimodal data. The model can be used to create fused representations by combining features across modalities. These learned representations are useful for classification and information retrieval. By sampling from the conditional distributions over each data modality, it is possible to create these representations even when some data modalities are missing. We conduct experiments on bi-modal image-text and audio-video data. The fused representation achieves state-of-the-art classification results on the MIR-Flickr dataset outperforming SVMs that use Multiple Kernel Learning.

Les données comportent souvent diverses modalités multiples : les images sont étiquetées par de l'information textuelle et les vidéos sont accompagnées d'un enregistrement audio. Chaque modalité se caractérise par des propriétés statistiques distinctes. Nous proposons une machine de Boltzmann pour l'apprentissage profond d'un modèle génératif pour de telles données multimodales. Le modèle peut servir à créer des représentations fusionnées en combinant les caractéristiques des modalités. Ces représentations apprises sont utiles pour la classification et la recherche d'informations. En échantillonnant les distributions conditionnelles sur chacune des modalités de données, il est possible de créer ces représentations même lorsqu'il manque certaines modalités de données. Nous réalisons des expériences sur des données bimodales image-texte et audio-vidéo. La représentation fusionnée produit des résultats de classification exceptionnels avec le jeu de données MIR-Flickr, surpassant les machines à vecteurs de support (SVM) qui utilisent l'apprentissage par noyaux multiples.

1B-I3: Recent Developments in Finite Mixture Models Récents progrès en modèles de mélanges finis

Organizer and Chair / Responsable et président: Pengfei Li (University of Waterloo)

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

JIAHUA CHEN (University of British Columbia), **PENGFEI LI** (University of Waterloo), **YUKUN LIU** (East China Normal University)

[Monday May 26 / lundi 26 mai, 10:20]

Sample Size Calculation for Homogeneity Test

Calcul de la taille de l'échantillon pour un test d'homogénéité

Mixture models are widely used to explain excessive variations in observations not captured by parametric models, and lead to suggestive latent structures. The hypothetical latent structure often needs critical examination based on experimental data. Knowing the sample size needed to ensure a reasonable chance of success is therefore an important problem. We investigate this issue for the EM-test and $C(\alpha)$ test. They are shown to be asymptotically equivalent and have simple limiting distributions under two sets of local alternatives for commonly used mixture models. A simple sample-size formula is obtained and its usefulness is validated via data examples and simulation studies.

Les modèles de mélange sont largement utilisés pour expliquer des variations excessives d'observations non saisies par des modèles paramétriques, et ils donnent lieu à des structures latentes suggestives. La structure latente hypothétique nécessite souvent un examen critique basé sur les données expérimentales. La connaissance de la taille d'échantillon requise pour assurer une probabilité raisonnable de réussite est, par conséquent, un problème important. Nous examinons cette question pour les tests EM et $C(\alpha)$. Ils se révèlent être asymptotiquement équivalents, et pour des modèles de mélange courants, ils montrent des distributions limitatives sous deux ensembles de contre-hypothèses locales. Nous obtenons une formule de taille d'échantillon simple et nous validons son utilité au moyen d'exemples de données et d'études de simulation.

HAJO HOLZMANN (Philipp University of Marburg, Germany), **FLORIAN SCHWAIGER** (Marburg University, Germany)

[Monday May 26 / lundi 26 mai, 10:50]

Hidden Markov Models with State-Dependent Mixtures: Minimal Representation, Model Testing and Applications to Clustering

Les modèles à chaîne de Markov cachée avec des mélanges dépendants de l'état : représentation minimale, test du modèle et application à l'analyse de regroupement.

We consider hidden Markov models (HMMs) in which the state-dependent distributions are finite mixtures. The parametrization is then not unique, but we determine a unique (up to label switching) representation in which the Markov chain has a minimal number of states, and propose a likelihood-ratio test for the hypothesis that the number of states in the Markov chain can be reduced without changing the distribution of the time-series model. Further, we propose a two-step clustering

Nous examinons des modèles à chaîne de Markov cachée (CMC) dont les distributions dépendant de l'état sont des mélanges finis. Dans un tel cas, la paramétrisation n'est pas unique, mais nous déterminons une représentation unique (à une permutation des étiquettes près) dans laquelle la chaîne de Markov a un nombre minimal d'états, et nous proposons un test de rapport de vraisemblance, dans l'hypothèse que le nombre d'états dans la chaîne de Markov peut être réduit sans changer la distribution du modèle de séries chronologiques. En outre, nous proposons un algo-

1B-I3: Recent Developments in Finite Mixture Models Récents progrès en modèles de mélanges finis

algorithm. Starting with a BIC choice for a standard HMM, in the first step we determine the minimal representation of the HMM by testing, and in the second step we merge components in the resulting state-dependent finite mixtures by using a local entropy criterion.

rithme de mise en grappes en deux étapes. Avec un choix initial de critères d'information bayésiens (BIC) pour un CMC ordinaire, nous effectuons un test en première étape pour déterminer la représentation minimale du CMC, et dans la deuxième étape, nous fusionnons les composantes en des mélanges finis dépendants de l'état qui en résultent, en utilisant un critère d'entropie locale.

HIROYUKI KASAHARA (University of British Columbia), **KATSUMI SHIMOTSU** (University of Tokyo)

[Monday May 26 / lundi 26 mai, 11:20]

Testing the Number of Components in Heteroscedastic Normal Mixtures

Test du nombre de composantes dans des mélanges normaux hétéroscédastiques

Testing the number of components in finite normal mixture models is a long-standing challenge because of its non-regularity. This paper studies likelihood-based testing of the number of components in heteroscedastic finite normal mixture models. We construct a likelihood-based test of the null hypothesis of m_0 components against the alternative hypothesis of $m_0 + 1$ components for any m_0 . The null asymptotic distribution of the proposed modified EM test statistic is the maximum of m_0 chi-squared random variables with two degrees of freedom, which can be easily simulated. The simulations show that the proposed test has very good finite sample size and power properties.

Un test du nombre de composantes dans des modèles de mélanges normaux finis constitue un défi de longue date en raison de sa non-régularité. Cet article examine un test basé sur la vraisemblance pour le nombre de composantes dans des modèles de mélanges normaux finis hétéroscédastiques. Nous construisons un test basé sur la vraisemblance pour l'hypothèse nulle à m_0 composante versus une contre-hypothèse à $m_0 + 1$ composantes pour tout m_0 . La distribution asymptotique de la statistique EM modifiée proposée sous l'hypothèse nulle est le maximum de m_0 variables aléatoires chi-deux avec deux degrés de liberté, ce qui peut facilement être simulé. Les simulations révèlent que le test proposé montre d'excellentes propriétés pour des tailles d'échantillon fini et une bonne puissance.

1B-14: Statistical Genetic and Genomic Modeling and Analysis of Complex Traits
Modélisation génétique et génomique statistique et analyse des caractères complexes

Organizer and Chair / Responsable et président: Brad McNeney (Simon Fraser University)

Room/Salle: UC - 140

Abstracts/Résumés

JOSEPH BEYENE (McMaster University)

[Monday May 26 / lundi 26 mai, 10:20]

Integrative Analysis of High-Dimensional Biological Data

Analyse intégrative des données biologiques de grande dimension

There is an increasing interest in integrating diverse data sets in the biological sciences. For instance, integrative genomics studies attempt at combining different types of high-dimensional molecular data (e.g., genotype, expression and methylation data) with the aim of extracting useful information for understanding basic biological processes and improve predictive accuracy for outcomes of interest. Novel integrative statistical and computational methods are emerging. However, there are several statistical and methodological challenges associated with data integration. In this talk, I will share lessons that we have learnt from attempting to integrate high-dimensional genomic data. I will present results from comprehensive simulation studies as well as illustrative real data examples. I will highlight strengths and limitations of integrative approaches and highlight challenges.

L'intégration de divers ensembles de données en biologie fait l'objet d'un intérêt grandissant. À titre d'exemple, des études génomiques intégratives procèdent à la combinaison de différents types de données moléculaires de grande dimension (i.e. le génotype, les données d'expression et de méthylation) ont pour objectifs d'extraire l'information utile à la compréhension des processus biologiques fondamentaux et d'améliorer la précision prévisionnelle des résultats d'intérêt. Il y a une émergence de méthodes novatrices d'intégration statistique et informatique. Par contre, elles présentent plusieurs défis statistiques et méthodologiques associés à l'intégration de données. Dans cet exposé, je partage des leçons tirées des essais d'intégration de données génomiques de grande dimension. Je présente les résultats d'études de simulation approfondies, ainsi que des exemples de données réelles. Enfin, je souligne les forces et les limites des approches intégratives ainsi que les défis qu'elles posent.

ANGELO CANTY (McMaster University), **NABIN SRESTHA** (McMaster University), **MARIE-PIERRE SYLVESTRE** (Université de Montréal), **JOHN BRUNZELL** (University of Washington), **ANDREW BORIGHT** (University of Toronto), **ANDREW PATERSON** (Hospital for Sick Children, Toronto), **SHELLEY BULL** (Mount Sinai Hospital, Toronto)

[Monday May 26 / lundi 26 mai, 10:50]

GWAS for Lipoprotein Profiles Using an Empirically Fitted Null Distribution

Études d'association pangénomique pour les profils de lipoprotéines à l'aide d'une distribution nulle ajustée empiriquement

Density gradient ultracentrifugation data is an ordered profile measuring the abundance of fractions of lipoproteins. We model these DGUC profiles using penalized splines fitted by linear mixed models. The effect of a SNP is modelled using fixed effects for genotype and genotype-fraction interaction. We test for a genetic effect using likelihood

Les données d'ultracentrifugation à gradient de densité constituent un profil ordonné qui mesure l'abondance des fractions de lipoprotéines. Nous avons modélisé ces profils d'ultracentrifugation à l'aide de fonctions splines pénalisées et ajustées par des modèles linéaires mixtes. L'effet d'un SNP est modélisé à l'aide d'effets fixes pour l'interaction entre le génotype et la fraction du génome. Nous testons l'effet

1B-14: Statistical Genetic and Genomic Modeling and Analysis of Complex Traits Modélisation génétique et génomique statistique et analyse des caractères complexes

ratios. Simulations show that the null distribution of the test statistics has much higher variability than the asymptotic chi-squared distribution. We propose to fit a three-parameter distribution related to the non-central chi-squared to model this null distribution. We use the observed GWAS test statistics to empirically estimate the parameters. We apply our method to real DGUC data and use simulations to examine its properties and compare it to alternatives.

génétique à l'aide d'un test du rapport de vraisemblance. Les simulations montrent que la distribution nulle des statistiques présente une variabilité plus élevée que sa distribution asymptotique du chi-deux. Nous proposons d'ajuster une distribution à trois paramètres liée au chi-deux non centré afin de modéliser cette distribution nulle. Nous utilisons les données statistiques observées des études d'association pangénomique pour estimer empiriquement les paramètres. Nous appliquons notre méthode aux données d'ultracentrifugation réelles et nous utilisons des simulations pour examiner ses propriétés et les comparer à celles des autres méthodes.

JO KNIGHT (University of Toronto), **SARAH GAGLIANO** (Centre for Addiction and Mental Health), **MICHAEL BARNES** (Queen Mary University of London), **MICHAEL WEALE** (King's College London)

[Monday May 26 / lundi 26 mai, 11:20]

Data Integration to Boost Association Signals

Intégration de données pour amplifier des signaux d'association

The increasing quantity and quality of functional genomic information motivate the integration of these data with association data, including data from genome-wide association studies (GWAS). We used previously described GWAS signals ("hits") to train a regularized logistic model to predict genetic variant causality on the basis of a large multivariate functional dataset. We show how this model can be used to derive Bayes factors for integrating functional and association data into a combined Bayesian analysis. Functional characteristics were obtained from the Encyclopedia of DNA Elements (ENCODE) and from other sources of genome-wide characteristics. We trained the model using all GWAS signals combined, and also using phenotype specific signals for autoimmune, brain-related, cardiovascular disorders, and cancer.

La quantité et la qualité croissantes de l'information génomique fonctionnelle favorisent l'intégration de ces données avec des données d'association, y compris des données issues des études d'association pangénomique (GWAS). Nous avons utilisé des signaux GWAS déjà découverts comme données d'entraînement pour un modèle logistique régularisé afin de prédire la causalité de variante génétique en fonction d'un vaste ensemble de données fonctionnelles multidimensionnelles. Nous montrons que le présent modèle peut être utilisé pour dériver des facteurs bayésiens pour l'intégration de données fonctionnelles et associatives en une analyse bayésienne combinée. Nous avons obtenu des caractéristiques fonctionnelles de l'Encyclopédie des éléments d'ADN (ENCODE) et d'autres sources de caractéristiques pangénomiques. Nous avons ajusté le modèle en utilisant tous les signaux GWAS combinés, ainsi que des signaux de phénotype spécifiques aux maladies autoimmunes, cérébrales ou cardiovasculaires, ainsi qu'au cancer.

1B-15: Statistical Issues in Health Economics and Health Technology Assessment Problèmes statistiques en économie de la santé et évaluation des technologies de la santé

Chair/Président: Nickolas Mitsakakis (University of Toronto)

Organizers/Responsables: Nickolas Mitsakakis (University of Toronto) and/et Eleanor Pullenayegum (University of Toronto & Hospital for Sick Children)

Room/Salle: UC - 163

Abstracts/Résumés

HEEJUNG BANG (University of California at Davis)

[Monday May 26 / lundi 26 mai, 10:20]

Analysis of "Censored" Medical Costs for the Last 15 Years: Something Old, Something New and Something Resurrected

Analyse des coûts médicaux «censurés» pour les 15 dernières années : du vieux, du neuf et du ressuscité

Medical cost is a health outcome data often collected in clinical trials and observational studies. As in survival data, censoring commonly occurs in cost data but censoring mechanism is different. It has been about 15 years since it was shown that use of standard statistical techniques can be invalid. Most analyze censored survival data correctly nowadays, whereas some still find censoring in costs tricky to handle despite computing codes and educational papers. We will review valid analysis methods that have been developed and provide some advice for practitioners. Particularly, the mathematical relationships/equivalencies among some cost estimators, and among Kaplan-Meier, inverse-probability-weighting, redistribute-to-the-right and replace-from-the-right could be useful when one tries to understand how different estimators work and which estimator to choose.

Les coûts médicaux font partie des données sur la santé souvent recueillies dans des essais cliniques et des études d'observations. Comme pour les données de survie, la censure est couramment effectuée sur les données de coûts, même si le mécanisme appliqué est différent. Il y a environ une quinzaine d'années, la validité des techniques statistiques traditionnelles a été remise en question. Aujourd'hui, les données de survie censurées sont analysées correctement en général, même si la censure des coûts pose toujours un défi pour certains, malgré les codes de calcul et les articles d'ordre éducatif. Nous passons en revue des méthodes d'analyse valides, et nous offrons quelques conseils utiles aux praticiens. Nous soulignons en particulier des relations ou équivalences mathématiques entre certains estimateurs de coûts, et signalons notamment l'estimateur de Kaplan-Meier, la pondération selon la probabilité inverse, la redistribution vers la droite et le remplacement de la droite, qui pourraient être utiles pour comprendre le mode de fonctionnement des divers estimateurs et pour déterminer le meilleur choix.

KRISTIAN THORLUND (McMaster University)

[Monday May 26 / lundi 26 mai, 10:50]

Integration of Network Meta-Analysis and Health Economic Evaluations -- Reducing Uncertainty and Avoiding Biases
Intégration de la méta-analyse en réseau et des évaluations économiques sur la santé afin de réduire l'incertitude et d'éviter les biais

Network meta-analysis (NMAs) allow for comparisons between interventions that have not all been compared head-to-head in randomized clinical trials (RCTs), and are increasingly being used to in-

La méta-analyse en réseau (NMA) permet la comparaison entre les interventions qui n'ont pas été comparées directement dans des essais cliniques randomisés et qui sont utilisées de plus en plus pour prévoir des paramètres d'effi-

1B-15: Statistical Issues in Health Economics and Health Technology Assessment Problèmes statistiques en économie de la santé et évaluation des technologies de la santé

form efficacy and safety parameters in health economic models. The integration of NMAs and health economic models is often challenging. When sub-optimal integration occurs, the health economic model loses out on valuable precision, thus making decision-making increasingly difficult.

This talk will provide conceptual introduction to what constitutes full integration of network meta-analysis and health economic analysis with a brief outline of key technical challenges will be provided. Conceptual and technical examples on select major diseases will be provided.

capacité et de sécurité dans les modèles d'économie de la santé. L'intégration des NMA et des modèles d'économie de la santé pose souvent des défis. Lorsqu'une intégration est sous-optimale, le modèle d'économie de la santé perd une précision significative, ce qui complique de plus en plus le processus décisionnel.

Cet exposé décrit une introduction conceptuelle de ce qui constitue une pleine intégration à la méta-analyse en réseau. Il propose une analyse d'économie de la santé et un bref aperçu des défis techniques clés, ainsi que des exemples conceptuels et techniques de certaines des principales maladies.

ANDY WILLAN (University of Toronto)

[Monday May 26 / lundi 26 mai, 11:20]

Should Stochastic Uncertainty Play a Role in the Pricing of New Health Care Interventions?

L'incertitude stochastique devrait-elle jouer un rôle dans l'établissement du prix des nouvelles interventions en matière de soins de santé?

A societal decision maker must consider the evidence in support of a new health care intervention prior to approving it for reimbursement at some price. Positive mean incremental net benefit (INB) has traditionally been used as the criterion for approval. More recently it has been argued that the uncertainty regarding INB should also be considered. Decision theory identifies the trade-off between the expected opportunity loss of approval with current evidence and the expected opportunity cost of delaying approval while additional evidence is acquired. There exists a threshold price for which the reduction in the expected opportunity loss from additional evidence equals the opportunity cost of delay. This is the maximum price acceptable to the decision maker.

Un décideur en matière d'affaires sociétales doit considérer les preuves à l'appui d'une nouvelle intervention en matière de soins de santé avant d'en autoriser le remboursement. Une moyenne positive de bénéfice net incrémentiel (BNI) a traditionnellement été utilisée comme critère d'approbation. Certains ont souligné plus récemment que l'incertitude en matière de BNI devrait être prise en considération. Les théories décisionnelles situent le compromis entre la perte d'opportunité attendue d'une approbation en fonction des preuves en mains et le coût d'opportunité d'un retard d'approbation en attente de preuves supplémentaires. Pour un décideur, le prix maximum acceptable est défini comme le seuil de prix pour lequel la réduction de perte d'opportunité liée à l'acquisition de preuves supplémentaires équivaut au coût d'opportunité d'un délai.

1B-I6: Statistical Methods in Actuarial Science Méthodes statistiques en science actuarielle

Organizer and Chair / Responsable et présidente: Yi Lu (Simon Fraser University)

Session sponsored by the Actuarial Science Section
Séance parrainée par le Groupe de science actuarielle

Room/Salle: UC - 179

Abstracts/Résumés

ARTHUR CHARPENTIER (Université du Québec à Montréal), **EMMANUEL FLACHAIRE** (Université d'Aix-Marseille, GREQAM-EHESS,)

[Monday May 26 / lundi 26 mai, 10:20]

Quantiles Estimation from Heavy-Tailed Distribution

Estimation de quantiles à partir de distributions à ailes relevées

In this work, we estimate quantiles (and other risk measures) from a nonparametric density estimation based on transformed data. A parametric cumulative distribution function is initially used to transform the data into values over the unit interval, from which a nonparametric density estimation is obtained. Finally, an estimation of the density of the original sample is obtained by back-transformation. This approach may be particularly useful to estimate heavy-tailed distributions. We discuss its implementation and its finite sample properties for density estimation, and for estimation and inference with quantiles.

Dans cet article, nous estimons les quantiles (et d'autres mesures du risque) à partir d'une estimation de la densité non paramétrique fondée sur des données transformées. Nous utilisons initialement une fonction de répartition paramétrique pour transformer les données en valeurs sur l'intervalle unitaire, desquelles nous obtenons une estimation de la densité non paramétrique. Enfin, nous obtenons une estimation de la densité de l'échantillon original par transformation inverse. Cette approche est particulièrement utile pour estimer des distributions à ailes relevées. Nous discutons son implémentation et examinons les propriétés pour des échantillons de taille finie pour l'estimation de la densité, ainsi que l'estimation et l'inférence de quantiles.

EDWARD FURMAN & JIANXI SU (York University), **RICARDAS ZITIKIS** (University of Western Ontario)

[Monday May 26 / lundi 26 mai, 10:50]

Background Risk Models: Theory and Applications

Modèles de risque rémanent: théorie et applications

Let the coordinates of the non-negative $m(\in \mathbf{N})$ -variate r.v. $\mathbf{Y} = (Y_1 \dots, Y_m)'$ be interpreted as possibly dependent risk factors (r.f.'s), and denote by $\mathbf{X} = (X_1, \dots, X_n)'$ risk portfolio (r.p.), whose risk components X_i are exposed to (sub)sets $\mathcal{S}_i \subseteq \{1, \dots, m\}$, $i = 1, \dots, n$ of the r.f.'s Y_j , $j = 1, \dots, m$. Let \mathcal{Y} and \mathcal{X} be collections of the r.f.'s and the r.p.'s, respectively. I will discuss maps $T : \mathcal{Y} \rightarrow \mathcal{X}$ and hence r.p.'s $X = T(Y)$ that yield additive and multiplicative background risk models. More specifically, I will examine connections be-

Supposons que les coordonnées de la variable $m(\in \mathbf{N})$ non négative $\mathbf{Y} = (Y_1 \dots, Y_m)'$ sont interprétées comme des facteurs de risque possiblement dépendants, et notons le portefeuille de risque par $\mathbf{X} = (X_1, \dots, X_n)'$, dont les composantes de risque X_i sont exposées aux (sous-)ensembles $\mathcal{S}_i \subseteq \{1, \dots, m\}$, $i = 1, \dots, n$ des facteurs de risque Y_j , $j = 1, \dots, m$. Supposons que \mathcal{Y} et \mathcal{X} sont les ensembles des facteurs de risque et du portefeuille de risque, respectivement. Je discute les cartes $T : \mathcal{Y} \rightarrow \mathcal{X}$, ainsi que le portefeuille de risque $X = T(Y)$ qui produit des modèles de risque rémanent additifs et multiplicatifs. Plus précisé-

1B-I6: Statistical Methods in Actuarial Science Méthodes statistiques en science actuarielle

tween the two, derive the underlying dependencies formally and explore the implications, state characteristic results and touch on some well-known special cases, formulate admissible classes of distributions of the r.f.'s. I will conclude with a number of applications to actuarial pricing and risk measurement.

ment, j'analyse les liens entre les deux, j'établis les dépendances sous-jacentes de façon formelle et j'en examine les conséquences, j'énonce les résultats des caractéristiques et j'aborde certains cas particuliers bien connus, et enfin je formule des classes admissibles de distributions des facteurs de risque. Je présente également quelques applications en matière de tarification actuarielle et de mesure du risque.

NATALIA NOLDE (University of British Columbia)

[Monday May 26 / lundi 26 mai, 11:20]

The Effect of Aggregation on Extremes from Asymptotically Independent Light-Tailed Risks

Effet de l'agrégation sur les valeurs extrêmes de distributions de risques à ailes plates asymptotiquement indépendantes

Portfolio risk diversification is a well-established concept in finance and insurance. While aggregation of several risky assets generally reduces the overall investment risk, the effectiveness of diversification depends on the stochastic properties of the assets comprising the portfolio. A new approach to quantifying the effect of portfolio tail diversification is proposed under the assumption of existence of a limit set. This property is satisfied by a number of distributions commonly used in financial applications. Several analytical examples are given to illustrate the proposed asymptotic diversification index as a measure of the effect of risk aggregation on extremes as well as to quantify the impact of dimension on diversification and as a tool in optimal portfolio selection.

La diversification du risque du portefeuille est un concept bien établi dans le domaine de la finance et de l'assurance. Même si l'agrégation de plusieurs actifs risqués réduit généralement le risque global d'investissement, l'efficacité de la diversification dépend des propriétés stochastiques des actifs qui composent le portefeuille. Nous proposons une nouvelle approche pour quantifier l'effet de la diversification des ailes de la distribution du portefeuille sous l'hypothèse de l'existence d'un ensemble limite. Bon nombre de distributions utilisées couramment en finances satisfont à cette propriété. Nous présentons plusieurs exemples analytiques pour illustrer l'indice de diversification asymptotique proposé à titre de mesure de l'effet de l'agrégation du risque sur les valeurs extrêmes et pour quantifier l'impact de la dimension sur la diversification, ainsi que comme outil pour la sélection du portefeuille optimal.

1B-C1: Biostatistics - New Estimation Methods Biostatistique - nouvelles méthodes d'estimation

Chair/Présidente: Lei Sun (University of Toronto)

Room/Salle: UC - 161

Abstracts/Résumés

CHRISTIAN BOUDREAU (University of Waterloo), **HUGO VARET** (Université Paris Descartes), **KUAN LIU** (London Health Sciences Centre), **SANDRINE KATSAHIAN** (Hôpital Henri Mondor)

[Monday May 26 / lundi 26 mai, 10:20]

Frailty Models for Competing Risks

Modèles à effets aléatoires pour risques compétitifs

In this talk, we extend the Fine & Gray model for the subdistribution of a competing risk to clustered time-to-event data by including random center effects or frailties. We first assume that these frailties are normally distributed, and use a penalized partial likelihood approach to fit our proposed model. This has the advantage that the majority of the computation can be done using standard survival analysis software. We then adapt our model and estimation method to gamma distributed frailties. We illustrate the proposed methodology with multi-center data from the European Blood and Marrow Transplant (EBMT) Group, and with data from a clinical trial on the efficacy of B-type natriuretic peptide (BNP) in patients in Intensive Care Units.

Dans cette présentation, nous étendons le modèle pour risques compétitifs de Fine & Gray lorsque les temps d'événements sont corrélés en incluant des termes de fragilité. Nous nous intéressons d'abord au cas où les fragilités suivent une loi normale et proposons une procédure d'estimation par vraisemblance partielle pénalisée couplée à une approximation de Laplace. L'avantage de cette méthode est que la majorité des calculs peuvent être faits avec des logiciels standards. Par la suite, nous étendons notre méthode aux fragilités de loi gamma. Finalement, nous illustrons notre méthodologie sur des données de greffes de moelle pour des patients atteints de leucémie et un essai clinique sur l'efficacité du peptide natriurétique pour les patients en soins intensifs.

KUN LIANG (University of Waterloo)

[Monday May 26 / lundi 26 mai, 10:35]

False Discovery Rate Estimation for Large-scale Homogeneous Discrete P -values

Estimation du taux de fausses découvertes pour des valeurs p discrètes homogènes à grande échelle

Large-scale homogeneous discrete p -values are encountered frequently in high-throughput genomics studies, and the related multiple testing problems become challenging because most existing methods for the false discovery rate (FDR) control and estimation assume continuous p -values. We show that the existing FDR control methods offer unsatisfactory solutions while the FDR estimation approach can provide a direct and practical answer. In the finite sample setting, we propose a novel class of conservative FDR estimators. Furthermore, we show that a broad class of FDR estima-

Des valeurs p discrètes homogènes à grande échelle sont courantes dans des études génomiques à très haut débit, et les multiples problèmes de tests qui en résultent posent un défi réel, étant donné que la plupart des méthodes existantes pour le contrôle et l'estimation du taux de fausses découvertes (TFD) doit présumer de la continuité des valeurs p . Nous démontrons que les méthodes de contrôle du TFD offrent des solutions insatisfaisantes, alors que l'approche d'estimation du TFD peut fournir directement une réponse pratique. Dans un contexte d'échantillon fini, nous proposons une nouvelle classe d'estimateurs du TFD conservateurs. En outre, nous expliquons qu'une grande classe

1B-C1: Biostatistics - New Estimation Methods Biostatistique - nouvelles méthodes d'estimation

tors is simultaneously conservative over all support points under some weak dependence condition in the asymptotic setting.

d'estimateurs du TFD est conservatrice simultanément pour tous les points de support dans des conditions de faible dépendance du contexte asymptotique.

LASANTHA PREMARATHNA (Simon Fraser University), **CARL JAMES SCHWARZ** (Simon Fraser University.)

[Monday May 26 / lundi 26 mai, 10:50]

Partial Stratification in Two-Sample Capture-Recapture Experiments

Stratification partielle dans les expériences de capture-recapture à deux échantillons

In many fishery experiments, it is more difficult, for example to sex all captured fish. In these cases, a sub-sample of the captured fish at each sampling occasion is selected and further, more costly, measurements (e.g. sex determination through sacrificing) are made. In this talk we develop and apply new methods for these types of experiments. Furthermore, given the relative costs of sampling for a simple capture and for processing the sub-sample, optimal allocation of effort for a given cost can be determined. We also develop methods to account for additional information such as prior information about sex ratio, continuous covariates (e.g. length). These methods are applied to a problem of estimating the size of the walleye population in MilleLacs, MN.

Dans bon nombre d'expériences de pêche, il est difficile de déterminer le sexe de tous les poissons capturés. Dans ces situations, on prélève un sous-échantillon des poissons capturés lors de chaque échantillonnage et on procède à d'autres mesures plus coûteuses (p. ex., sexage par sacrifice). Nous développons et appliquons de nouvelles méthodes pour ces types d'expériences. Par ailleurs, étant donné le coût relatif de l'échantillonnage pour chaque capture et du traitement du sous-échantillon, nous déterminons la répartition optimale de l'effort pour un coût donné. Nous développons aussi des méthodes pour traiter d'autres données comme l'information préalable concernant la répartition par sexe et des covariables continues (longueur, etc.). Nous appliquons ces méthodes à un problème d'estimation de la taille de la population des dorés à MilleLacs, MN.

KRISHNA SAHA (Central CT State University)

[Monday May 26 / lundi 26 mai, 11:05]

Statistical Analysis of Means and Dispersions from Several Treatment Groups for Extra-Dispersed Count Data from Toxicology Studies

Analyse statistique des moyennes et dispersions de plusieurs groupes de traitement pour des données de dénombrement surdispersées tirées d'études en toxicologie

This study presents inference procedures on means and dispersions for extra-dispersed count data from several negative binomial distributions as well as semi-parametric distributions that belong to the family of distributions of counts with the same means and variances of the negative binomial distributions. We derive several tests, including ones based on the adjusted counts introduced by Rao and Scott and Neyman's score functions, to test the equality of means of several treatment groups of counts. We also derive several $C(\alpha)$ tests for testing the validity of equal or unequal dispersions assumption. We illustrate the methodology with an example from toxicology studies and compare the

Cette étude présente des procédures d'inférence sur les moyennes et les dispersions pour des données de dénombrement surdispersées de plusieurs distributions binomiales négatives et de distributions semiparamétriques qui appartiennent à la famille des distributions de dénombrements avec les mêmes moyennes et variances que les distributions binomiales négatives. Nous développons plusieurs tests, dont un basé sur les dénombrements ajustés introduits par Rao et Scott et les fonctions de score de Neyman, afin de tester l'égalité des moyennes de plusieurs groupes de traitement de dénombrements. Nous dérivons également plusieurs tests $C(\alpha)$ pour tester la validité d'une hypothèse de dispersions égales ou inégales. Nous illustrons ces méthodes par un exemple tiré d'études en toxicologie et comparons le

1B-C1: Biostatistics - New Estimation Methods Biostatistique - nouvelles méthodes d'estimation

procedures in terms of their empirical levels and powers with a Monte Carlo simulation study.

niveau empirique et la puissance de ces procédures via une étude par simulation Monte Carlo.

DONEAL THOMAS & ANDREA BENEDETTI (McGill University)

[Monday May 26 / lundi 26 mai, 11:20]

Estimating Generalized Linear Mixed Models with Binary Outcomes for the Analysis of Individual Patient Data Meta-Analyses: Which method is best?

Estimation des modèles mixtes linéaires généralisés avec résultats binaires pour l'analyse de méta-analyses sur données individuelles : quelle est la meilleure méthode?

Individual patient data meta-analysis (IPD-MA) is as regarded the gold standard, having several advantages over conventional MA. Penalized quasi-likelihood (PQL) and Adaptive Gauss-Hermite quadrature (AGHQ) are estimation procedures for the generalized linear mixed models. Few comparisons between the methods are reported in the context of IPD-MA with binary outcomes \hat{u} that is, few clusters, imbalance cluster sizes, large inter-study heterogeneity, small effect sizes, and an interest in the variance parameters. Preliminary results show the biases of fixed effects estimates by PQL and AGHQ are systematically related under the random coefficient models (true models) as well as misspecified models. PQL, however, is superior for data with large inter-study variability. Inappropriate methods and misspecified models in IPD-meta-analyses could lead to biased parameter estimates and erroneous conclusion.

La méta-analyse sur données individuelles (MADI) constitue la référence car elle présente plusieurs avantages par rapport à la MA traditionnelle. La quasi-vraisemblance pénalisée (QVP) et la quadrature adaptative de Gauss-Hermite (QAGH) sont des procédures d'estimation pour les modèles mixtes linéaires généralisés. Il n'existe que peu de comparaisons entre les méthodes dans le contexte de la MADI avec résultats binaires : c'est-à-dire peu de grappes, des tailles de grappes déséquilibrées, une grande hétérogénéité entre études, de petites tailles d'effets et un intérêt pour les paramètres de variance. Les résultats préliminaires montrent que les biais des estimations des effets fixes via QVP et QAGH sont systématiquement connectés dans les modèles à coefficients aléatoires (modèles réels) et les modèles avec erreurs de spécification. La QVP est supérieure pour les données qui présentent une grande variabilité entre études. Dans les méta-analyses sur données individuelles, des méthodes inappropriées et des modèles avec erreurs de spécification risquent de produire des estimations de paramètres biaisées et des conclusions erronées.

SAMAN MUTHUKUMARANA (University of Manitoba), **MICHAEL EVANS** (University of Toronto)

[Monday May 26 / lundi 26 mai, 11:35]

Bayesian Hypothesis Assessment in Two-arm Trials Using Relative Belief Ratios

Évaluation bayésienne d'hypothèses utilisant les rapports de confiance relative dans des essais à deux groupes

We describe a Bayesian approach for assessing equivalence and non-inferiority hypotheses in two-arm trials using relative belief ratios. A relative belief ratio is a measure of statistical evidence and can indicate evidence either for or against a hypothesis. In addition to the relative belief ratio, we also compute a measure of the strength of this evidence as a calibration of the relative belief ratio. Prior elicitation, model checking and checking for prior-data conflict procedures are developed to ensure that the choices of model and prior made are

Nous décrivons une approche bayésienne utilisant des rapports de confiance relative pour évaluer les hypothèses d'équivalence et de non infériorité dans un essai à deux groupes. Un rapport de confiance relative constitue une mesure statistique pouvant indiquer une préférence soit en faveur ou contre une hypothèse. En plus du rapport de confiance relative, nous avons calculé une mesure de la force de cette information afin de calibrer le ratio de confiance relative. Nous avons procédé à l'élicitation de lois a priori, à la vérification du modèle et à la détection de conflits entre les données et la loi a priori afin de nous assurer que le

1B-C1: Biostatistics - New Estimation Methods **Biostatistique - nouvelles méthodes d'estimation**

relevant to the specific application. We highlight the applicability of the approach and illustrate the proposed method by applying it to a data set obtained from a two-arm clinical trial.

choix de modèle et de loi a priori sont pertinents à l'application visée. Nous soulignons l'applicabilité de notre approche en l'utilisant pour analyser les données d'un essai à deux groupes.

1B-C2: Probability Theory and Applications Théorie des probabilités et applications

Chair/Président: Serge Provost (Western University)

Room/Salle: UC - A101

Abstracts/Résumés

FARID EL KTAIBI & GAIL IVANOFF (University of Ottawa)

[Monday May 26 / lundi 26 mai, 10:20]

Blockwise Bootstrap for Empirical Processes and Change-Point Detection

Méthodes de rééchantillonnage pour les processus empiriques et détection de changement de structure dans la loi marginale

The asymptotic behaviour of the sequential empirical process of a stationary sequence is a classic problem which has been approached in many statistical studies. Almost always, conditions on mixing or association are imposed. In a linear model, this can be avoided by using a combination of conditions on the moments of the innovations and the summability of the coefficients.

We will propose a nonparametric approach for testing for a change in the marginal distribution function

for a linear model. We consider here two consistent tests based on the Kolmogorov-Smirnov and the Cramér-von Mises statistics.

Lastly, we discuss Moving Block Bootstrap techniques and their application to our test procedure.

la théorie des comportements asymptotiques des processus empiriques est d'un intérêt majeur dans plusieurs applications. Cependant, la plupart des travaux générés antérieurement sur les sommes partielles des séries stationnaires se basent sur des conditions de mélange ou d'association. Ces dernières conditions peuvent être évitées, dans le cas des modèles linéaires, en optant pour une approche fondée sur des arguments liés aux moments des erreurs et aux sommes

des coefficients.

On propose ici une procédure non paramétrique pour tester des changements structuraux de la loi marginale d'un modèle linéaire en utilisant les statistiques de Kolmogorov-Smirnov et Cramér-von Mises.

Finalement, on étudie l'apport des méthodes de rééchantillonnage «bootstrap de bloc mobile» à notre approche.

SAMIRA SADEGHI (University of Alberta)

[Monday May 26 / lundi 26 mai, 10:35]

Outer-product of Linear Processes with Heavy-tail and Long-range Dependence

Produit extérieur de processus linéaires à ailes lourdes et dépendance à long terme

The Marcinkiewicz Strong Law, $\lim_{n \rightarrow \infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^n (D_k - D) = 0$ a.s. with $p \in (1, 2)$, is studied for outer products $D_k = X_k \bar{X}_k^T$, where $\{X_k\}, \{\bar{X}_k\}$ are both two-sided (multivariate) linear processes (with coefficient matrices $(C_l), (\bar{C}_l)$ and i.i.d. zero-mean innovations $\{\Xi\}, \{\bar{\Xi}\}$). Matrix sequences C_l and \bar{C}_l can decay slowly enough (as $|l| \rightarrow \infty$) that $\{X_k, \bar{X}_k\}$ have long-range dependence while $\{D_k\}$ can have heavy tails. In particular, the heavy-tail and long-

La loi forte de Marcinkiewicz, $\lim_{n \rightarrow \infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^n (D_k - D) = 0$ a.s.

avec $p \in (1, 2)$, est étudiée pour les produits extérieurs $D_k = X_k \bar{X}_k^T$, où $\{X_k\}, \{\bar{X}_k\}$ sont des processus linéaires multivariés (avec des matrices de coefficients $(C_l), (\bar{C}_l)$ et innovations i.i.d. de moyenne nulle $\{\Xi\}, \{\bar{\Xi}\}$). Les séquences de matrices C_l et \bar{C}_l peuvent se dégrader si lentement (quand $|l| \rightarrow \infty$) que $\{X_k, \bar{X}_k\}$ présentent une dépendance à long terme tandis que $\{D_k\}$ peuvent présenter des ailes lourdes. En particulier, les phénomènes d'ailes lourdes et de dépen-

1B-C2: Probability Theory and Applications Théorie des probabilités et applications

range-dependence phenomena for $\{D_k\}$ are handled simultaneously and a new decoupling property is proved that shows the convergence rate is determined by the worst of the heavy-tails or the long-range dependence, but not the combination. The main result is applied to obtain Marcinkiewicz Strong Law of Large Numbers for stochastic approximation, non-linear functions forms and autocovariances.

dance à long terme pour $\{D_k\}$ sont traités simultanément et une nouvelle propriété de découplage est prouvée, qui indique que le taux de convergence est déterminé par le pire élément des ailes lourdes ou de la dépendance à long terme, mais non par une combinaison des deux. Le résultat principal est appliqué pour obtenir une loi forte des grands nombres de Marcinkiewicz pour l'approximation stochastique, les formes de fonctions non linéaires et les autocovariances.

IBRAHIM ABDELRAZEQ, GAIL IVANOFF & RAFAL KULIK (University of Ottawa)

[Monday May 26 / lundi 26 mai, 10:50]

Goodness of Fit Test: Recovered Noise for CAR(1) Processes

Test d'adéquation : récupération du bruit dans des processus CAR(1)

When an Ornstein-Uhlenbeck (or CAR(1)) process is observed at discrete times $0, h, 2h, \dots, [T/h]h$, the unobserved driving process (noise) can be approximated from the observed process. Abdelrazeq, Ivanoff, and Kulik (2013) have used the approximated increments of the driving process to test the assumption that the process is Lévy-driven. If it can be concluded that the driving process is Lévy, the empirical process of the approximated increments can then be used to carry out more precise tests of goodness-of-fit. For example, one can test whether the driving process can be modeled as a Brownian motion or a gamma process.

Lorsqu'un processus d'Ornstein-Uhlenbeck (ou CAR(1)) est observé à des temps discrets $0, h, 2h, \dots, [T/h]h$, le processus sous-jacent non observé (bruit) peut être approximé à partir du processus observé. Abdelrazeq, Ivanoff et Kulik (2013) ont utilisé des incréments approximatés du processus sous-jacent pour mettre à l'épreuve l'hypothèse que le processus est gouverné par des processus de Lévy. Si cette hypothèse peut être prouvée, le processus empirique d'incrémentés approximatés peut alors être utilisé pour effectuer des tests d'adéquation plus précis. À titre d'exemple, il est possible de tester si le processus sous-jacent peut être modélisé en tant que mouvement brownien ou processus gamma.

SEAN JEWELL, NEIL SPENCER & ALEXANDRE BOUCHARD-CÔTÉ (University of British Columbia)

[Monday May 26 / lundi 26 mai, 11:05]

Modelling Graffiti Prevalence in Metro Vancouver under a Hierarchical Bayesian Non-parametric Framework

Modélisation de la prévalence de graffitis dans la région métropolitaine de Vancouver dans un contexte bayésien hiérarchique non paramétrique

Poisson process models are commonly used for analyzing spatial patterns in urban crime data. However, there are instances when the continuous assumption on the Poisson process intensity measure is invalid. This happens in datasets where exact locations are anonymized, or aggregated to the nearest civic address for practical purposes. In these instances, the intensity measure is more accurately represented as a collection of discrete atoms. We address this problem using a hierarchical Bayesian non-parametric framework with discrete random measures to encapsulate the aggre-

Les modèles de processus de Poisson sont couramment utilisés pour l'analyse des répartitions spatiales des données sur la criminalité urbaine. Par contre, il y a des cas où l'hypothèse de continuité de la mesure d'intensité du processus de Poisson est erronée, notamment dans des ensembles de données où, pour des raisons pratiques, les emplacements exacts sont anonymisés, ou agrégés à l'adresse réelle la plus près. Dans de tels cas, la mesure d'intensité est représentée plus précisément comme une collection d'atomes discrets. Nous abordons ce problème en utilisant un contexte bayésien hiérarchique non paramétrique avec des mesures aléatoires discrètes pour encapsuler l'effet d'a-

1B-C2: Probability Theory and Applications Théorie des probabilités et applications

gation effect. Throughout this talk, we demonstrate our method using graffiti prevalence in Metro Vancouver as a running example.

grégation. L'exposé décrit la méthode expérimentale à l'aide des données sur la prédominance des graffitis dans la région métropolitaine de Vancouver.

SIMON MAK (Georgia Institute of Technology), **DEREK BINGHAM & YI LU** (Simon Fraser University)

[Monday May 26 / lundi 26 mai, 11:20]

A Spatial Compound Poisson Process for Hurricane Damage

Processus spatial de Poisson composé pour évaluer les dommages résultant d'un ouragan

The compound Poisson process is often used for modelling aggregate claims and pricing premiums for insuring against economic damage incurred by hurricanes and tropical storms. This process models total economic damage by taking into account both (independent) fluctuations in the event frequency and individual damage associated. However, it does not provide a mechanism for predicting claims and pricing premiums within different spatial locations. In this presentation, a Bayesian spatial-augmented compound Poisson process is proposed, which uses CAR models (proposed by Besag, 1974) to incorporate a storm path model as well as spatial effects for damage. The proposed model is motivated and applied to historical hurricane data. Other results related to model selection, prediction, and premium/reserve pricing are also presented.

Le processus de Poisson combiné est souvent utilisé pour modéliser des demandes d'indemnisation totales et des primes de tarification pour l'assurance contre les dommages résultant des ouragans et des tempêtes tropicales. Ce processus modélise les dommages économiques totaux en tenant compte à la fois des fluctuations (indépendantes) en matière de fréquence d'événement et de dommages individuels associés. Toutefois, il ne prévoit aucun mécanisme de prédiction des demandes d'indemnisation ni de primes de tarification pour divers emplacements. Nous proposons dans cette présentation un processus de Poisson combiné dans un modèle bayésien augmenté d'un paramètre spatial, lequel utilise les modèles CAR (proposés par Besag, 1974) afin d'intégrer un modèle de trajectoire de tempête ainsi que des effets spatiaux pour les dommages. Le modèle proposé est appliqué à des données historiques sur les ouragans. Nous présentons aussi d'autres résultats liés à la sélection et à la prédiction de modèles, ainsi qu'à la tarification en fonction des primes et de la réserve.

1B-C3: Statistical Inference and Applications 1 Inférence statistique et applications 1

Chair/Président: Tolulope Sajobi (University of Calgary)

Room/Salle: UC - Croft House

Abstracts/Résumés

MALAK ALMUTAIRI & JIANAN PENG (Acadia University)

[Monday May 26 / lundi 26 mai, 10:20]

Simultaneous Inference for Low Dose Risk Estimation with Abbott-adjusted Log-Logistic Response Model Benchmark Analysis

Inférence simultanée pour l'estimation du risque à faible dose avec le modèle log-logistique ajusté selon la formule d'Abbott

Risk evaluation studies are imperative in that they are enacted to secure small dose levels on human beings and animals that may emerge from a noxious agent. This is especially challenging since the study approach is obstructed to large dose levels regarding the urgent. According to Buckley and Piegorsch (Statistical Methodology, 2008), their hyperbolic band that was to depict low dose inference with regard to Abbott-adjusted Logistic response approach is desirable. The shape of the invariable width is more desirable than the hyperbolic band. Risk analysis in some cases should be narrower especially towards the lower end of little doses. We will explore through presentation that width band is likely to perform better than the hyperbolic.

Les études d'évaluation du risque sont primordiales du fait qu'elles sont mises en oeuvre pour établir des niveaux sécuritaires de faibles doses pour les humains et les animaux pouvant provenir d'un agent toxique. Ce qui est particulièrement problématique, c'est que l'approche de l'étude est restreinte à des niveaux de doses élevées de l'agent toxique. Selon Buckley et Piegorsch (Statistical Methodology, 2008), la bande hyperbolique développée pour l'inférence à faible dose selon une approche logistique ajustée selon la formule d'Abbott présente un intérêt. Cependant, il semblerait qu'une forme de bande à largeur constante semble plus appropriée que la bande hyperbolique. Dans cet exposé, nous montrons qu'une bande à largeur constante devrait être plus efficace qu'une bande hyperbolique.

MARTI ANDERSON (Massey University, Auckland, New Zealand)

[Monday May 26 / lundi 26 mai, 10:35]

A Comparison of Permutation, Bootstrap and Monte Carlo Techniques Applied to the Multivariate Behrens-Fisher Problem: Behaviour and Consequences for Statistical Inference

Comparaison des techniques de permutation, bootstrap et Monte Carlo appliquées à un problème multivarié de Behrens-Fisher : comportement et conséquences pour l'inférence statistique

The Behrens-Fisher problem (BFP), one of the oldest puzzles in statistics, asks: how may we validly compare means (centroids) between two or more populations when variances (dispersions) differ? Some solutions to the multivariate BFP do exist, but they assume multivariate normality. More robust solutions for non-normal and high-dimensional cases are needed.

We compared some empirical solutions to the multivariate BFP that rely on permutation, bootstrap

L'une des plus anciennes énigmes statistiques, le problème de Behrens-Fisher (BFP) soulève la question suivante : comment peut-on valablement comparer des moyennes (centres de masse) entre deux populations ou plus lorsque les variances (dispersions) diffèrent ? Il existe certaines solutions au BFP multivarié, mais elles supposent une normalité multivariée. Des solutions plus robustes sont nécessaires pour des cas de grande dimension.

Dans une étude de simulation, nous comparons des solutions empiriques au BFP multivarié reposant sur les méth-

1B-C3: Statistical Inference and Applications 1

Inférence statistique et applications 1

or Monte Carlo (re-)sampling methods in a simulation study. The permutation approach was found to be liberal, while the bootstrap approach, even with bias-corrections, was conservative. A Monte Carlo approach may provide a more exact test in some cases, but at what cost in terms of additional underlying assumptions?

odes d'échantillonnage de permutation, de bootstrap ou de Monte Carlo. Nous avons constaté que l'approche de permutation est libérale, tandis que l'approche bootstrap est conservatrice, même avec ses corrections de biais. Une approche de Monte Carlo peut produire un test plus précis dans certains cas, mais au coût de combien d'hypothèses sous-jacentes additionnelles ?

MOHAMED BELALIA & TAOUFIK BOUEZMARNI (Université de Sherbrooke), **ALEXANDRE LEBLANC** (University of Manitoba)

[Monday May 26 / lundi 26 mai, 10:50]

Bernstein Polynomials for Estimating a Conditional Distribution Function

Estimation de la fonction de répartition conditionnelle par les polynômes de Bernstein.

In a variety of statistical problems, estimation of the conditional distribution function is always a challenge. In this talk we propose a new nonparametric method estimation for the conditional distribution function. We use an adjusted form of the Nadaraya-Watson estimator based on the Bernstein polynomial as smooth estimator for the conditional distribution. We derive the asymptotic properties of the proposed estimator. The asymptotic variance of the Nadaraya-Watson estimator is reduced through the use of smoothing. Simulation results shows the performance of the proposed estimator.

L'estimation de la fonction de répartition conditionnelle est toujours un défi pour plusieurs problèmes en statistique. Dans cet exposé, nous proposons une nouvelle méthode non paramétrique pour l'estimation de la fonction de répartition conditionnelle. Nous utilisons les polynômes de Bernstein pour lisser l'estimateur de Nadaraya-Watson de la fonction de distribution conditionnelle. Nous étudions les propriétés asymptotiques de cet estimateur. Des résultats de simulations montrent la performances de cet estimateur.

JINJUN TONG (memorial university of NL), **J CONCEPCIÓN LOREDO-OSTI** (Memorial University of Newfoundland)

[Monday May 26 / lundi 26 mai, 11:05]

Estimation of the State-Space Model with Stable Noise

Estimation d'un modèle à espace d'états avec un bruit stable

In stead of Gaussian noise, a positive sub-family in stable distribution which is named as Lévy-type stable distribution was considered in a state-space model with generalized extreme value (GEV) marginals. A quantile regression would be used to estimate the marginal heavy-tailed GEV parameters while parameters associated with the dynamics of the model could be estimated by a modified Baum-Welch procedure with the initial values deduced from the function of the empirical distribution of the GEV sequence. The model filtering, smoothing will be based on these estimators.

Au lieu d'un bruit gaussien, nous avons analysé une sous-famille positive à distribution stable, appelée distribution stable de type Lévy, dans un modèle à espace d'états avec des marges de valeur extrême généralisée (GEV). Une régression quantile serait utilisée pour estimer les paramètres GEV à queue lourde des marges, tandis que les paramètres associés à la dynamique du modèle pourraient être estimés à l'aide d'une procédure modifiée de Baum-Welch avec les valeurs initiales déduites de la fonction de distribution empirique de la séquence GEV. Le filtrage et le lissage du modèle seront fondés sur ces estimateurs.

1B-C3: Statistical Inference and Applications 1

Inférence statistique et applications 1

SEONG-HWAN JUN (University of British Columbia)

[Monday May 26 / lundi 26 mai, 11:20]

A Beginner's Perspective on Causal Inference: Potential Outcomes Framework, Graphical Models and Structural Causal Model

Perspective d'un novice sur l'inférence causale : cadre pour les résultats potentiels, modèles graphiques et modèle causal structurel

As a PhD student in Statistics I have primarily focussed on the potential outcomes framework for inferring causality. However, it came to my attention recently that we are not the only ones developing theories regarding causal inference. There have been frameworks and inference methods developed in parallel (mostly from computer science) based on what is known as the structural causal

model, which makes heavy use of graphical models.

I will present my perspectives on causal inference by comparing the potential outcomes framework to the structural causal model in analyzing the observational data and on the use of graphical models. Summary of data analysis carried out on the American time use survey data will be presented.

En qualité d'étudiant au doctorat en statistique, j'ai concentré ma recherche sur le cadre des résultats potentiels d'une inférence de causalité. Toutefois, j'ai constaté récemment que nous ne sommes pas les seuls à élaborer des théories en matière d'inférence causale. Des méthodes et des cadres d'inférence ont été parallèlement mis au point (principalement en science informatique) sur la base du modèle causal structurel qui fait largement appel aux modèles graphiques. Je présente donc mes perspectives sur l'inférence causale en comparant le cadre des résultats potentiels au modèle causal structurel dans l'analyse de données d'observation et

dans l'utilisation de modèles graphiques. Je présente aussi un sommaire de l'analyse des données issues de l'enquête sur l'emploi du temps aux États-Unis.

LUCAS THORPE & MEI LING HUANG (Brock University)

[Monday May 26 / lundi 26 mai, 11:35]

On Nonparametric Methods of Extreme Value Distributions

Méthodes non paramétriques de distributions de valeurs extrêmes

The Pareto distribution is a heavy-tailed distribution with applications to extreme events in the real world. Many existing estimation methods for the Pareto distribution are based on estimating the tail index. There are theoretical difficulties with this estimation since the Pareto distribution often has infinite moments. Also, there are some bias issues with estimating the tail index. In this paper, we propose a nonparametric estimation method to overcome these difficulties. The results of Monte Carlo simulations show good properties of the proposed method. We also study an example of real-world applications to extreme values using the proposed method.

La distribution de Pareto est une distribution à ailes lourdes avec des applications aux événements extrêmes du monde réel. Bon nombre de méthodes d'estimation de la distribution de Pareto se fondent sur une estimation de l'indice caudal. Mais cette estimation présente des difficultés théoriques car la distribution de Pareto a souvent des moments infinis. Il existe également des problèmes de biais en ce qui concerne l'estimation de l'indice caudal. Dans cette présentation, nous proposons une méthode d'estimation non paramétrique pour surmonter ces difficultés. Les résultats des simulations de Monte Carlo montrent les bonnes propriétés de la méthode proposée. Nous étudions également un exemple d'application réelle de cette méthode sur les valeurs extrêmes.

1B-C4: Survey Methodology 2 Méthodologie d'enquête 2

Chair/Président: David Haziza (University of Montreal)

Room/Salle: Hart House - South Dining

Abstracts/Résumés

TOUNKARA FODE & RIVEST LOUIS-PAUL (Université Laval)

[Monday May 26 / lundi 26 mai, 10:20]

Copula Model-based Approach and Empirical Bayes Techniques for Small Area Proportions

Approche basée sur le modèle de copules et les techniques empiriques bayésiennes sur les proportions pour les petits domaines

Survey data are used to obtain information regarding characteristics such as mean and proportion relating not only to a population, but also to areas. However, the sample size in some areas may be small, leading to direct estimators of poor quality. Indirect estimators based on auxiliary information and constructed from implicitly and explicitly models-based methods are recommended. On this point the methodology is very rich, However for the case of binary data the theory is limited. Here, we propose a copula-based approach of small area estimation for binary data. Empirical Bayes techniques are used to obtain explicit form for estimates of proportions. Using simulation study, we evaluated the performance of estimators. Finally, we present a numerical example.

Les données de sondage sont utilisées pour obtenir de l'information sur des caractéristiques telles qu'une moyenne et une proportion relative à non seulement une population mais aussi à des domaines. Par contre, la taille de l'échantillon dans certains domaines peut être petite, ce qui mène à des estimateurs de mauvaise qualité. Des estimateurs indirects fondés sur des renseignements auxiliaires et construits à partir de méthodes implicitement et explicitement basées sur des modèles sont recommandés. Sur ce point la méthodologie est très abondante. Cependant, dans le cas de données binaires, la théorie est limitée. Ici, nous proposons une approche d'estimation sur petits domaines pour les données binaires basée sur des copules. Des techniques empiriques bayésiennes sont utilisées pour obtenir des formes explicites pour l'estimation de proportions. Nous évaluons la performance des estimateurs en utilisant une étude de simulation. Finalement, nous présentons un exemple numérique.

NAIMA GOUZI (Statistique Canada)

[Monday May 26 / lundi 26 mai, 10:35]

Integrating Establishment Characteristics When Calibrating a Business Survey

Intégrer les caractéristiques des établissements lors du calage d'une enquête-entreprise

In the context of business surveys, an enterprise can have several production entities, called establishments. Each establishment works in an area of activity in a geographic region. The purpose of a survey is to estimate the total of the variables of interest associated with specific areas of activity. Surveys use the same population of enterprises; an enterprise can therefore be in the scope of several surveys. First, a large sample of enterprises is selected for all surveys to correct auxiliary informa-

Dans le contexte des enquêtes économiques, une entreprise peut avoir plusieurs entités de production appelées établissements. Chaque entité exploite un secteur d'activité dans une région géographique. L'objectif d'une enquête est d'estimer le total des variables d'intérêt associées à un ensemble spécifique de secteurs d'activité. Les enquêtes utilisent la même population d'entreprises; une entreprise peut être dans le champ de plusieurs enquêtes. Un grand échantillon d'entreprises est d'abord sélectionné pour toutes les enquêtes afin de corriger l'information auxiliaire comme la mau-

1B-C4: Survey Methodology 2 Méthodologie d'enquête 2

tion, such as misclassification. Each survey then selects its own subsample; subsamples can overlap. We present the issues surrounding the use of auxiliary information on establishments to improve each survey estimates.

vaise classification. Ensuite, chaque enquête sélectionne son propre sous-échantillon. Les sous-échantillons peuvent se chevaucher. Nous présentons les enjeux de l'utilisation de l'information auxiliaire de l'établissement pour améliorer les estimations de chaque enquête.

AMANDA HALLADAY (Statistics Canada)

[Monday May 26 / lundi 26 mai, 10:50]

Methodological Challenges with the 2015 Canadian Community Health Survey - Nutrition

Défis méthodologiques du volet de nutrition de l'Enquête sur la santé dans les collectivités canadiennes de 2015

The 2015 Canadian Community Health Survey (CCHS) - Nutrition presents several interesting methodological challenges. The aims of this survey are to gather information regarding dietary intake and consumption of dietary supplements, to estimate distributions for usual dietary intake, to obtain accurate height and weight measurements, to collect data on certain health conditions, and to evaluate changes from the 2004 Nutrition Survey. The purpose of this talk is to discuss the challenges that Statistics Canada faces, including: the number of second interviews required to get an accurate picture of the within-person variability and the optimal methods of sample allocation and estimation of dietary intake given budgetary constraints. We will address these issues and others as we discuss the complexities of nutritional data.

Le volet de nutrition de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) de 2015 présente bon nombre de défis méthodologiques intéressants. Cette étude a pour objectifs de recueillir des renseignements sur l'apport diététique et la consommation de suppléments diététiques afin d'estimer les distributions de l'apport diététique ordinaire, d'obtenir des mesures précises de taille et de poids, de recueillir des données sur certains états de santé, et d'évaluer les changements par rapport à l'enquête sur la nutrition de 2004. L'objectif de cet exposé consiste à discuter les défis que Statistiques Canada doit relever, et notamment les suivants : le nombre de deuxièmes entrevues nécessaires pour obtenir une représentation précise de la variabilité pour une même personne, et les méthodes optimales de répartition de l'échantillon et d'estimation d'apport diététique, tout en tenant compte des contraintes budgétaires. Nous abordons ces questions et d'autres encore dans une discussion sur les complexités des données nutritionnelles.

ERIC PELLETIER & PIERRE-OLIVIER JULIEN (Statistique Canada)

[Monday May 26 / lundi 26 mai, 11:05]

The Volatility and the Rarity of Scientific Research and Development in Canada -- How to "Search" for It?

La volatilité et la rareté de la recherche et développement scientifique au Canada -- Comment la « chercher »?

Due to its transition into the Integrated Business Statistics Program (IBSP), the Research and Development in Canadian Industry (RDCI) survey will undergo major methodological changes. The target population for the RDCI survey -- Canadian enterprises performing or funding scientific research and development -- is rare and volatile, which makes the challenges even more considerable. This presentation provides a general overview of the survey and discusses its inherent complexities, the improvements that will be introduced under IBSP, as well as the difficulties arising from

Suite à son intégration au Programme intégré de la statistique sur les entreprises (PISE), l'enquête sur la recherche et développement dans l'industrie canadienne (RDIC) subira de grands bouleversements méthodologiques. La population ciblée par l'enquête RDIC -- les entreprises canadiennes exécutant ou finançant des dépenses en recherche et développement scientifique -- étant rare et volatile, les défis sont d'autant plus importants. Cette présentation donnera un aperçu de l'enquête et traitera des complexités qui y sont inhérentes, des améliorations qui y seront apportées sous le PISE ainsi que des difficultés rencontrées en identifiant la population ciblée.

1B-C4: Survey Methodology 2 Méthodologie d'enquête 2

identifying the target population.

MICHAEL ROTONDI (York University)

[Monday May 26 / lundi 26 mai, 11:20]

Towards the Estimation of Effect Measures in Studies Using Respondent-Driven Sampling

Vers une estimation des mesures d'effet dans les études avec échantillonnage fondé sur les répondants (EFR)

Respondent-Driven Sampling (RDS) is an increasingly common sampling technique to recruit hidden populations. Statistical methods for RDS are not straightforward due to the correlation between individual outcomes and subject weighting, thus analyses are typically limited to estimation of population proportions. This presentation discusses the Method Of Variance Estimates Recovery (MOVER) to construct confidence intervals (CIs) for effect measures such as the risk difference or relative risk in studies using RDS. To illustrate the approach, MOVER is used to examine differences in the prevalence of demographic characteristics between an RDS study and convenience study of injection drug users. MOVER is then applied to obtain CIs for the relative risk between education levels and HIV-seropositivity, respectively.

L'échantillonnage fondé sur les répondants (EFR) est une technique d'échantillonnage de plus en plus courante pour recruter les populations cachées. Les méthodes statistiques employées pour l'EFR sont complexes en raison de la corrélation entre les résultats individuels et la pondération du sujet, si bien que les analyses sont généralement limitées à l'estimation des proportions de populations. Cette présentation discute de la méthode MOVER (Method Of Variance Estimates Recovery) pour établir des intervalles de confiance (IC) pour les mesures d'effets comme la différence de risque ou le risque relatif dans les études employant l'EFR. Pour illustrer cette approche, la méthode MOVER est utilisée pour examiner les différences de prévalence de caractéristiques démographiques entre une étude par EFR et une étude de commodité d'utilisateurs de drogues injectables. La méthode MOVER est alors appliquée pour obtenir des IC pour le risque relatif entre le niveau d'éducation et la séropositivité, respectivement.

KYLE VINCENT (Bank of Canada), **SAMAN MUTHUKUMARANA & JENNA TICHON** (University of Manitoba)

[Monday May 26 / lundi 26 mai, 11:35]

Bayesian Item Response Analysis of Method-of-Payment Habits in Banking Surveys

Analyse bayésienne des réponses par items dans les sondages bancaires sur les habitudes de moyens de paiement

Customers have a wide variety of choices in selecting a method of payment in modern society due to advancements in technology. We investigate the method of payment habits of banking customers using item response models. We consider three binary item response models used in the literature within an empirical Bayesian framework. These models capture the heterogeneity and complexity of customer perception on methods of payment in different capacities, with different features. For this reason, model assessment methods need to be developed for better inferential purposes. We introduce an assessment criterion based on predictive simulations and illustrate the approach using graphical summary measures. The approach is further highlighted using empirical financial data.

Dans notre société moderne, grâce aux progrès en technologie, les clients ont le choix entre de nombreuses méthodes de paiement. Nous étudions les habitudes de moyens de paiement de clients bancaires à l'aide de modèles de réponses par items. Nous considérons trois modèles de réponses binaires par items utilisés dans la littérature dans un cadre bayésien empirique. Ces modèles permettent de saisir l'hétérogénéité et la complexité de la perception des clients concernant les méthodes de paiement à divers titres, avec des caractéristiques différentes. Pour cette raison, il est nécessaire de développer des méthodes d'évaluation de modèles afin d'améliorer l'inférence. Nous introduisons un critère d'évaluation fondé sur des simulations prédictives et illustrons notre approche à l'aide de mesures sommaires graphiques, ainsi que des données financières empiriques.

1D-I1: Opportunities and Challenges in the First Years of an Academic Career (A Panel Discussion)

Possibilités et défis en début de carrière universitaire (table ronde)

Organizer and Chair / Responsable et présidente: Anne-Sophie Charest (Laval University)

Session sponsored by the New Investigator Committee

Séance parrainée par le Comité des nouveaux chercheurs

Room/Salle: UC - 163

Description

In this panel session we will discuss what to expect in the first years of your academic career. It is a period filled with new opportunities as you have the chance to define your own research program, build new collaborations, advise and mentor students, develop teaching material,... But these opportunities come with a whole new set of challenges, and it may be stressful to successfully conduct research, teaching and service activities, especially while focusing on your own career goals, and trying to keep a good work-life balance. During the session, three academics in different stages of their career will share their experience and advice regarding the first few years in academia. Participants in attendance will be invited to ask questions to the panellists and participate in the discussion.

Dans ce panel, nous échangerons sur ce à quoi vous pouvez vous attendre dans les premières années de votre carrière universitaire. C'est une période remplie de nouvelles possibilités : vous avez l'occasion de définir votre propre programme de recherche, de développer de nouvelles collaborations, de conseiller et encadrer des étudiants, de concevoir du matériel pédagogique, ... Évidemment, ces possibilités s'accompagnent d'un ensemble de nouveaux défis. Il peut alors être stressant de mener à bien à la fois vos activités de recherche, d'enseignement et de service, tout en gardant en tête vos propres objectifs de carrière, et en essayant de garder un bon équilibre travail-vie. Dans ce panel, trois professeurs à différents stades de leur carrière partageront avec vous leur expérience et quelques conseils pour les premières années d'une carrière universitaire. Les participants présents seront invités à poser des questions aux intervenants et à participer à la discussion.

1D-I2: Optimal Design of Experiments Plans d'expérience optimaux

Organizer and Chair / Responsable et présidente: Julie Zhou (University of Victoria)

Room/Salle: UC - 140

Abstracts/Résumés

SAUMEN MANDAL & MONSUR CHOWDHURY (University of Manitoba)

[Monday May 26 / lundi 26 mai, 13:30]

Optimal Experimental Design for Maximum Likelihood Estimation

Plan d'expérience optimal pour l'estimation du maximum de vraisemblance

We present a quite flexible methodology to solve a maximum likelihood estimation problem using optimal design theory and simultaneous optimization techniques. We consider a problem of determining maximum likelihood estimates under a hypothesis of marginal homogeneity for data in square contingency tables. This is an optimization problem with respect to variables that satisfy several constraints based on the marginal homogeneity conditions. We first formulate the Lagrangian function and then transform the problem to that of maximizing some functions of the cell probabilities simultaneously. We apply the methodologies in some data sets for which the hypothesis of marginal homogeneity is of interest. The methodologies could be applied to a wide class of optimization problems where constraints are imposed on the parameters.

Nous présentons une méthodologie relativement flexible pour résoudre un problème d'estimation du maximum de vraisemblance à l'aide de la théorie du plan optimal et de techniques d'optimisation simultanée. Nous tentons de déterminer les estimateurs du maximum de vraisemblance sous une hypothèse d'homogénéité des marges pour des données dans des tableaux de contingence carrés. Il s'agit d'un problème d'optimisation des variables qui satisfont à plusieurs contraintes fondées sur les conditions de l'homogénéité des marges. Nous formulons d'abord la fonction lagrangienne, puis nous transformons la question en un problème de maximisation simultanément certaines fonctions des probabilités de cellule. Nous appliquons cette méthodologie à des jeux de données pour lesquels l'hypothèse d'homogénéité des marges est à l'étude. Les méthodes proposées peuvent être appliquées à une vaste gamme de problèmes d'optimisation qui comportent des contraintes imposées sur les paramètres.

PETER QIAN (University of Wisconsin-Madison)

[Monday May 26 / lundi 26 mai, 14:00]

Samurai Sudoku-Based Space-Filling Designs

Plans de grilles Samurai inspirés du Sudoku

The game-board of Samurai Sudoku consists of five overlapping Sudoku grids, for each of which several entries are provided and the remaining entries must be filled subject to no row, column and three-by-three subsquare containing duplicate numbers. By exploiting these uniformity properties, we construct a new type of design, called a Samurai Sudoku-based space-filling design. This design has several attractive properties: (1) the complete design achieves attractive uniformity in

Le jeu de Sudoku Samurai comporte cinq grilles qui se recoupent, pour chacune desquelles plusieurs chiffres sont fournis, le reste étant des cases à remplir de sorte qu'aucune rangée, ni colonne, ni bloc de trois cases sur trois ne contiennent deux fois le même chiffre. En tirant parti de ces propriétés d'uniformité, nous construisons un nouveau plan d'expérience appelé plan de grilles Samurai inspiré du Sudoku. Ce plan possède plusieurs propriétés intéressantes : (1) le plan complet réalise une uniformité intéressante à la fois dans les marges univariées et bivariées ; (2) il peut être

1D-I2: Optimal Design of Experiments Plans d'expérience optimaux

both univariate and bivariate margins; (2) it can be divided into groups of subdesigns with overlaps such that each subdesign achieves maximum uniformity in both univariate and bivariate margins; (3) each of the overlaps achieves maximum uniformity in both univariate and bivariate margins. These designs are appealing for meta-analysis and cross-validation.

divisé en groupes de sous-plans avec des recouvrements tels que chacun des sous-plans réalise une uniformité maximale à la fois dans les marges univariées et bivariées; (3) chacun des recouvrements réalise une uniformité maximale à la fois dans les marges univariées et bivariées. Ces plans sont attrayants pour la méta-analyse et la validation croisée.

WENG KEE WONG (University of California Los Angeles), **JIAHENG QIU** (UCLA), **RAY-BING CHEN** (National Cheng Kung University), **WEICHUNG WANG** (National Taiwan University)

[Monday May 26 / lundi 26 mai, 14:30]

Using Animal Instincts to Find Efficient Experimental Designs

Utilisation de l'instinct animal pour trouver des plans d'expériences efficaces

I first present a brief overview of optimal design methodology. Particle swarm optimization (PSO) is then introduced to find optimal designs for potentially any model and any design criterion. The method works quite magically and frequently finds the optimal design or a nearly optimal design in an efficient way. There is virtually no explicit assumption required for the method to perform well and the user only needs to input a few easy tuning param-

eters in the PSO algorithm.

Using models from the biopharmaceutical sciences as examples, I demonstrate how PSO searches for different types of optimal experimental designs in dose response studies, including mini-max types of optimal designs where effective algorithms to find such designs have remained elusive until now.

Je présente d'abord un bref aperçu d'une méthodologie de plan optimal, puis je présente l'optimisation en banc de sable (PSO) pour trouver des plans optimaux pour presque n'importe quel modèle ou critère. La méthode fonctionne à merveille et trouve fréquemment et efficacement le plan optimal ou quasi optimal. Il n'y a pour ainsi dire aucune hypothèse explicite nécessaire à l'efficacité de la méthode et l'utilisateur doit simplement spécifier pour l'algorithme PSO quelques paramètres de réglage simples.

En utilisant comme exemples des modèles de science biopharmaceutique, je démontre le processus de recherche du

PSO pour différents types de plans expérimentaux optimaux dans des études dose-effet, y compris des plans optimaux de type minimax, pour lesquels des algorithmes efficaces permettant de trouver de tels plans n'avaient pas encore pu être élaborés.

1D-I3: Recent Advances in Spatial Statistics Récents progrès en statistique spatiale

Chair/Président: Ye Li (University of Toronto)

Organizer/Responsable: Mahmoud Torabi (University of Manitoba)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: Hart House - Debates Rm

Abstracts/Résumés

YING MACNAB (University of British Columbia)

[Monday May 26 / lundi 26 mai, 13:30]

Bayesian Hierarchical Spatial Mixture for Modeling Sparse Data on a Discontinuous and Modestly Large Irregular Lattice

Modélisation bayésienne hiérarchique de mélanges pour données clairsemées sur un treillis irrégulier discontinu d'envergure modeste

A class of Bayesian hierarchical Bernoulli-Binomial spatial mixture zero-inflated Binomial models is developed for modeling noisy and mostly low small area crude rates. Three spatial prior formulations, the intrinsic conditional autoregressive or iCAR, the BYM, and the modified BYM models, were explored for their performance on modelling sparse data on a modestly large (N in the thousands) and discontinuous irregular lattice. The methods are illustrated via an in-depth Bayesian analysis of postcode-level response rates from a recent recruitment for a physical exercise intervention in Sheffield, UK. With increasing availability of spatial data referenced at fine spatial scales, the sparse-data situation and the Bayesian models and methods discussed herein should have considerable relevance to disease mapping and spatial regression.

Nous présentons une classe de modèles bayésiens hiérarchiques binomiaux de mélanges spatiaux à surreprésentation de zéros de type Bernoulli-binomial pour la modélisation de taux bruts de petits domaines bruyants et principalement faibles. Nous avons examiné la performance de trois formules de loi a priori spatiale et des modèles d'autorégression conditionnelle intrinsèque (iCAR), de BYM et de BYM modifié pour la modélisation de données clairsemées sur un treillis irrégulier discontinu d'envergure modeste (pour un N dans les milliers). Les méthodes sont illustrées par le biais d'une analyse bayésienne approfondie des taux de réponses par code postal à un recrutement récent pour une proposition d'exercice physique à Sheffield au Royaume-Uni. En présence d'une disponibilité croissante de données spatiales référencées à des petites échelles spatiales, une situation avec données clairsemées et les méthodes et modèles bayésiens abordés dans cet exposé devraient présenter un grand intérêt en matière de cartographie des maladies et de régression spatiale.

RHONDA ROSYCHUK (University of Alberta), **HSING-MING CHANG** (University of British Columbia Okanagan)

[Monday May 26 / lundi 26 mai, 14:00]

A Spatial Scan Statistic for Compound Poisson Data, Using Negative Binomial Distribution and Accounting for Population Stratification

Statistique de balayage spatial pour des données de Poisson composées selon une distribution binomiale négative en tenant compte de la stratification de la population

1D-I3: Recent Advances in Spatial Statistics Récents progrès en statistique spatiale

Much effort has been devoted to the development of spatial data analysis methods, including the detection of spatial clusters of cases and events in the biological sciences and epidemiology. Recently, research has examined detecting clusters of correlated count data associated with health conditions of individuals. Such a method allows researchers to examine spatial relationships of disease-related events rather than just incident or prevalent cases. We introduce a spatial scan test that identifies clusters of (correlated) events. We base the test on a special compound Poisson model (representation of a negative binomial distribution) that realizes advantages in computation over the general compound Poisson model. We illustrate our method on emergency department visits data, adjusted by key population characteristics such as age.

Un effort considérable a été consacré au développement de méthodes d'analyse de données spatiales, y compris la détection de grappes spatiales de cas et d'événements en sciences biologiques et en épidémiologie. Récemment, des travaux de recherche ont porté sur la détection de grappes de données de dénombrement corrélées associées à l'état de santé des individus. Une telle méthode permet d'examiner les relations spatiales entre des événements liés aux maladies plutôt qu'à de simples cas incidents ou prévalents. Nous proposons un test de balayage spatial qui identifie des grappes d'événements (corrélés). Le test est basé sur un modèle de Poisson composé spécial (représentant une distribution binomiale négative) qui présente des avantages en matière de calculs par rapport au modèle de Poisson composé classique. Nous illustrons notre méthode à l'aide de données de visites au service de l'urgence ajustées selon des caractéristiques clés de la population comme l'âge.

MAHMOUD TORABI & MAHMOUD TORABI (University of Manitoba)

[Monday May 26 / lundi 26 mai, 14:30]

Spatial Generalized Linear Mixed Models with Multivariate CAR Models for Areal Data

Modèles linéaires mixtes généralisés spatiaux avec modèles autorégressifs conditionnels multivariés pour des données géospatiales

Disease mapping studies have been widely performed with considering only one disease in the estimated models. Simultaneous modeling of different diseases can also be a valuable tool both from the epidemiological and also from the statistical point of view. In particular, when we have several measurements recorded at each spatial location, we need to consider multivariate models in order to handle the dependence among the multivariate components as well as the spatial dependence between locations. We use an approach, which yields to maximum likelihood estimation, to conduct frequentist analysis of spatial generalized linear mixed models with multivariate conditional autoregressive (CAR) models for areal data. The proposed approach is evaluated through a simulation study and also by a real dataset.

De nombreuses études de cartographie de maladies ont été réalisées en tenant compte d'une seule maladie dans les modèles estimés. La modélisation simultanée de différentes maladies peut s'avérer un outil précieux à la fois d'un point de vue épidémiologique et statistique. En particulier, lorsque plusieurs mesures sont enregistrées à chaque emplacement spatial, nous devons considérer les modèles multivariés afin de traiter la dépendance parmi les composantes multivariées, ainsi que la dépendance spatiale entre les emplacements. L'approche proposée, de laquelle est obtenue l'estimation du maximum de vraisemblance, nous permet d'effectuer une analyse fréquentiste des modèles linéaires mixtes généralisés spatiaux avec modèles autorégressifs conditionnels multivariés pour des données géospatiales. Elle est évaluée au moyen d'une étude de simulation et d'un jeu de données réelles.

1D-I4: Statistical Analysis with Data Complications Analyse statistique et complications liées aux données

Chair/Présidente: Wei Liu (York University)

Organizer/Responsable: Lang Wu (University of British Columbia)

Room/Salle: Hart House - East Common

Abstracts/Résumés

JOAN HU (Simon Fraser University)

[Monday May 26 / lundi 26 mai, 13:30]

Challenges and Strategies for Statistical Inference with Administrative Databases

Défis et stratégies en matière d'inférence statistique sur des bases de données administratives

A typical administrative database contains information over the calendar time from a target population starting on a particular date. However, often a scientifically meaningful analysis considers the individual time as the individual's age or the elapsed time since an individual-specific event. Plus many situations demand inference regarding a population larger than the one from which the data are collected. This talk begins with a description of two real administrative databases to introduce the inherent challenges in analyzing the data, and then presents a uniform framework to investigate several incomplete data structures and to assess the merits of their associated inference methods. We explore usefulness of different supplementary information to account for incompleteness mechanism, improve inference efficiency, and reduce computational intensity.

Une base de données administratives classique comprend des renseignements sur une population cible recueillis sur une période de temps à partir d'une date précise. Toutefois, il arrive souvent qu'une analyse scientifique porte sur un temps comme l'âge d'un individu, ou comme le temps écoulé depuis un événement spécifique à l'individu. En outre, bon nombre de situations nécessitent une inférence relativement à une population plus vaste que celle qui a fait l'objet de la collecte. Dans cet exposé, nous décrivons d'abord deux bases de données administratives réelles pour aborder les défis inhérents à l'analyse des données, puis nous présentons un cadre de référence uniforme pour analyser plusieurs structures de données incomplètes et pour évaluer les mérites des méthodes d'inférence qui leur sont associées. Nous explorons l'utilité de diverses données supplémentaires pour remédier au mécanisme d'incomplétude, améliorer l'efficacité d'inférence et réduire l'intensité des calculs.

SANJOY SINHA (Carleton University), **ABDUS SATTAR** (Case Western Reserve University)

[Monday May 26 / lundi 26 mai, 14:00]

A Robust Approach to Analyzing Longitudinal Data with Informative Dropouts and Outliers

Approche robuste pour l'analyse de données longitudinales comportant des cessations de suivi et des données aberrantes informatives

In this talk, I will introduce a robust method for analyzing incomplete longitudinal data with informative dropouts. The method is developed in the framework of the maximum likelihood, and is useful for bounding the influence of potential outliers in the data when estimating the model parameters. The empirical properties of the estimators will be investigated based on simulations. An application will

Dans cet exposé, je présente une méthode robuste d'analyse de données longitudinales incomplètes comportant des données aberrantes informatives. Élaborée dans un contexte de vraisemblance maximale, cette méthode est utile pour limiter l'influence des données aberrantes potentielles dans l'estimation des paramètres. J'examine les propriétés empiriques des estimateurs sur la base de simulations. Je présente une application utilisant des mesures longitudi-

1D-I4: Statistical Analysis with Data Complications Analyse statistique et complications liées aux données

be provided using longitudinal measurements on a biomarker of sepsis being considered for investigating the progression of pneumonia to severe sepsis in a group of patients.

nales d'un biomarqueur de septicémie pour investiguer la progression de la pneumonie vers la septicémie grave dans un groupe de patients.

LEILEI ZENG, LAN WEN & RICHARD COOK (University of Waterloo)

[Monday May 26 / lundi 26 mai, 14:30]

Model Misspecification in Progression-Free Survival Analysis Due to Intermittent Assessment

Erreurs de spécification de modèles dues à une évaluation intermittente en analyse de survie sans progression

Cancer clinical trials are routinely designed to assess the effect of treatment on disease progression and death, often in terms of a composite endpoint called progression-free survival. Progression status is only known at periodic assessment times and hence the time of progression is interval censored. Despite the considerable advances in methods for dealing with interval censored data, naive methods such as right-end point imputation are used routinely. We derive the asymptotic bias of estimators of treatment effect and the impact on power for tests of treatment effect, when analyses are based on the time progression is first detected. This is carried out in the context of a three-state illness-death model. Empirical studies corroborate the asymptotic findings.

Les essais cliniques sur le cancer sont couramment conçus pour évaluer l'effet d'un traitement sur la progression d'une maladie et la mortalité, souvent en fonction d'un critère composite appelé durée de survie sans progression. L'état de progression est connu seulement à des moments d'évaluation périodique et, par conséquent, le moment de progression est censuré par intervalle. Malgré des avancées considérables dans les techniques de traitement des données censurées par intervalles, des méthodes naïves comme l'imputation à droite sont couramment utilisées. Nous dérivons le biais asymptotique des estimateurs de l'effet du traitement et l'impact sur la puissance des tests basés sur ces estimateurs, lorsque les analyses sont effectuées en fonction du moment initial d'une détection de la progression. L'analyse est appliquée en fonction du modèle de maladie-mortalité à trois états. Des études empiriques corroborent les résultats asymptotiques.

1D-I5: Stochastic Spatial Processes Processus spatiaux stochastiques

Organizer and Chair / Responsable et président: Tom Salisbury (York University)

Session sponsored by the Probability Section
Séance parrainée par le Groupe de probabilité

Room/Salle: UC - 179

Abstracts/Résumés

LOUIGI ADDARIO-BERRY (McGill University)

[Monday May 26 / lundi 26 mai, 13:30]

Growing Random Maps

Comment s'enfuir de l'Hexagone

We use a growth procedure for binary trees due to Luczak and Winkler, a bijection between binary trees and irreducible quadrangulations of the hexagon due to Fusy, Poulalhon and Schaeffer, and the classical angular mapping between quadrangulations and maps, to define a growth procedure for maps. The growth procedure is local, in that every map is obtained from its predecessor by an operation that only modifies vertices lying on a common face with some fixed vertex. As n tends to infinity, the probability that the n th map in the sequence is 3-connected tends to $2^8/3^6$. The sequence of maps has an almost sure limit G ; we show that G is the distributional local limit of large, uniformly random 3-connected graphs.

On définit un processus stochastique $(M_n, n \geq 1)$ tels que M_n a la loi uniforme sur les quadrangulations enracinées et irréductibles de l'hexagone. La quadrangulation M_{n+1} est obtenue de M_n à partir d'une règle *locale*, c'est-à-dire que M_{n+1} et M_n sont identiques en dehors de l'ensemble de faces voisines d'un seul sommet. Une bijection classique associe à chaque quadrangulation M_n une carte G_n à n faces. On établit la convergence presque sûre des M_n et des G_n vers des limites M_∞ et G_∞ . Finalement, on démontre que si $(\hat{G}_n, n \geq 1)$ est telle que \hat{G}_n a la loi uniforme sur l'ensemble de cartes enracinées 3-connexes avec n arêtes, alors \hat{G}_n converge en loi vers G_∞ dans le sens «Benjamini-Schramm».

DENIZ SEZER (University of Calgary)

[Monday May 26 / lundi 26 mai, 14:00]

Exit Densities and the Martin Boundary of Super-Brownian Motion

Densités de passage et frontière de Martin du supermouvement brownien

The relationship between the equation $1/2\Delta u = 2u^2$ and super-Brownian motion (SBM) is analogous to the relationship between Brownian motion and Laplace's equation. Substantial progress on this semi-linear p.d.e. has been made by exploring this analogy. An area that remains to be explored is the Martin boundary theory, which in the semi-linear case is defined as the extreme elements of the convex set of X -harmonic functions.

La relation entre l'équation $1/2\Delta u = 2u^2$ et le supermouvement brownien est analogue à la relation entre l'équation de Laplace et le mouvement brownien. L'exploration de cette analogie a permis la réalisation de progrès substantiels sur cette équation aux dérivées partielles. Cependant, la théorie de la frontière de Martin n'a pas encore été explorée. Dans le cas semi-linéaire, elle est définie par les éléments extrêmes de l'ensemble convexe des fonctions harmoniques X . Je décris une importante famille de fonctions harmoniques X , H^ν , indexées par les mesures finies ν sur ∂D pour un

1D-I5: Stochastic Spatial Processes Processus spatiaux stochastiques

I will describe an important family of X -harmonic functions, H^ν , indexed by finite measures ν on ∂D for a smooth and bounded domain D , and their connection to the Martin boundary of SBM. These functions can be thought as the X -harmonic functions corresponding to conditioning SBM on $X_D = \nu$.

domaine lisse et borné D , ainsi que leurs liens avec la frontière de Martin du supermouvement brownien. Ces fonctions peuvent être considérées comme les fonctions harmoniques X permettant de conditionner le supermouvement brownien sur $X_D = \nu$.

JANOSCH ORTMANN & JEREMY QUASTEL (University of Toronto), **DANIEL REMENIK** (Universidad de Chile)
[Monday May 26 / lundi 26 mai, 14:30]

Exact Formulas for Random Growth off a Flat Interface

Formules exactes pour la croissance aléatoire à partir d'une interface plane

We will describe formulas for the asymmetric simple exclusion process (ASEP) starting from half-flat and flat initial data. The formulas are for the exponential moments of the height function associated with ASEP. They lead to explicit formulas for certain generating functions of the model, which in the flat case can be expressed as a Fredholm Pfaffian. The limiting fluctuations are conjectured to fall in the Kardar-Parisi-Zhang (KPZ) universality class. We will explain how these formulas can be used to provide formal derivations of these limiting fluctuations, and discuss some related models.

On va décrire des formules pour le processus d'exclusion simple asymétrique (PESA) à partir de données initiales demi-planes et planes. Les formules sont pour les moments exponentiels de la fonction de hauteur associée au PESA. Elles mènent à des formules explicites pour certaines fonctions génératrices du modèle, qui dans le cas plan peut être exprimé par un pfaffien de Fredholm. On conjecture que les fluctuations limites sont dans la classe d'universalité Kardar-Parisi-Zhang (KPZ). On va expliquer comment ces formules peuvent être utilisées pour donner des dérivations formelles des fluctuations limites, et on discute de modèles connexes.

1D-16: Teaching Introductory Statistics with R Enseigner la statistique de base avec R

Chair/Président: Duncan Murdoch (Western University)
Organizer/Responsable: John Braun (Western University)

Session sponsored by the Statistics Education Section
Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: UC - 161

Abstracts/Résumés

JOEL DUBIN (University of Waterloo)

[Monday May 26 / lundi 26 mai, 13:30]

Teaching R to Introductory Biostatistics Students in a Professional Distance-Education Masters Program

Enseignement de R dans un cours d'introduction à la biostatistique aux étudiants d'un programme de maîtrise professionnelle à distance

Teaching biostatistics to those not enrolled/majoring in statistics/biostatistics can be difficult in its own right. Doing so in a distance format creates additional challenges. In this talk, I will focus on experiences teaching R software in a distance education biostatistics course in the University of Waterloo Masters in Public Health program. We will discuss current strategies as well as some future plans. The hope is to convey some lessons learned that could apply to both distance and traditional lecture-based introductory statistics courses for students that generally do not have a quantitative or computational background.

L'enseignement de la biostatistique à des étudiants inscrits ou spécialisés dans un domaine autre que la statistique ou la biostatistique peut s'avérer difficile en soi. L'enseignement à distance dans ce contexte peut ajouter des défis supplémentaires. Cet exposé portera sur des expériences d'enseignement du logiciel R dans le cadre d'un cours à distance de biostatistique du programme de maîtrise en santé publique de l'Université de Waterloo. Nous discutons de stratégies courantes et de divers projets à réaliser. Nous espérons ainsi tirer des leçons applicables à la fois à des cours d'introduction à la statistique offert à distance et sous la forme d'exposés magistraux traditionnels à des étudiants qui ne possèdent généralement pas de formation quantitative ou informatique.

KEVIN THORPE (University of Toronto)

[Monday May 26 / lundi 26 mai, 14:00]

Using R in Introductory Biostatistics Courses

Utilisation du logiciel R dans des cours d'introduction à la biostatistique

It is virtually impossible to teach applied biostatistics without teaching something about the use of statistical software. One of the risks of teaching statistics software within a statistics course is that learning a package can obscure the statistical education. This talk will present my approach and experience with integrating R into the fabric of the introductory biostatistics courses I teach at the Dalla

Il est pratiquement impossible d'enseigner la biostatistique appliquée sans enseigner quelques rudiments sur les logiciels statistiques. L'un des risques d'intégrer cet enseignement dans un cours de statistique réside dans le fait que l'apprentissage d'un progiciel peut occulter l'éducation statistique. Dans cet exposé, je présente mon approche et mon expérience en matière d'intégration du logiciel R dans l'élaboration des cours d'introduction à la biostatistique que j'en-

1D-16: Teaching Introductory Statistics with R Enseigner la statistique de base avec R

Lana School of Public Health, University of Toronto.

seigne à la Dalla Lana School of Public Health de l'Université de Toronto.

ALISON GIBBS, DISCUSSANT, University of Toronto
[Monday May 26 / lundi 26 mai, 14:30]

1D-C1: Actuarial Science and Finance 1 Actuariat et finance 1

Chair/Président: Paul Marriott (University of Waterloo)

Room/Salle: UC - A101

Abstracts/Résumés

MATHIEU BOUDREAU (Université du Québec à Montréal), **GENEVIÈVE GAUTHIER** (HEC Montréal), **TOMMY THOMASSIN** (Standard Life)

[Monday May 26 / lundi 26 mai, 13:30]

Estimation of Asset Correlations with Transient Shocks in a Multi-name Credit Risk Model

Estimation de corrélations d'actifs avec des chocs transitoires dans un modèle sans-nom de risque de crédit

It is well documented (Aït-Sahalia (2004), Duan and Fulop (2009), Johannes et al. (2009)) that asset prices are contaminated by trading noise which make it difficult to distinguish permanent from transient shocks. In this presentation, we will show how transient shocks impact the estimation of asset correlations in a general multi-name credit risk model. An estimation technique is presented and the statistical properties of this estimator are compared with other approaches. We conclude with an empirical example where we find that ignoring trading noise can seriously underestimate asset correlations and consequently, credit risk measures used for risk management.

Il est bien documenté (Aït-Sahalia (2004), Duan and Fulop (2009), Johannes et al. (2009)) que le prix des actifs est contaminé par le bruit d'échange ce qui rend difficile de distinguer les chocs permanents des chocs transitoires. Au cours de cette présentation, nous allons démontrer comment les chocs transitoires influencent l'estimation de la corrélation des actifs dans un modèle général sans-nom de risque de crédit. Une technique d'estimation est présentée et les propriétés statistiques de cet estimateur sont comparées avec d'autres approches. Nous concluons avec un exemple empirique pour lequel nous trouvons qu'ignorer le bruit d'échange peut conduire à une sous-estimation importante des corrélations d'actifs et ainsi, des mesures de risque de crédit utilisées en gestion de risque.

MARIE-PIER CÔTÉ & CHRISTIAN GENEST (McGill University)

[Monday May 26 / lundi 26 mai, 13:45]

Copula-based Model for Risk Aggregation

Modèle d'agrégation de risques basé sur des copules

A flexible approach is proposed for risk aggregation. The model consists of a tree structure, bivariate copulas, and marginal distributions. The construction relies on a conditional independence assumption whose implications are studied. A procedure for selecting the tree structure is developed using hierarchical clustering techniques, along with a distance metric based on Kendall's tau. Estimation, simulation, and model validation are also discussed. The approach is illustrated using data from a Canadian property and casualty insurance company.

On propose une approche flexible pour l'agrégation de risques. Le modèle est constitué d'une arborescence, de copules bivariées et de lois marginales. La construction s'appuie sur un postulat d'indépendance conditionnelle dont les ramifications sont étudiées. On montre comment choisir l'arborescence au moyen de techniques de classification et d'une métrique définie à partir du tau de Kendall. L'estimation, la simulation et l'adéquation du modèle sont aussi abordées. L'approche est illustrée à l'aide de données d'une compagnie canadienne en assurance IARD.

1D-C1: Actuarial Science and Finance 1

Actuariat et finance 1

HASSAN OMIDI FIROUZI (University of Montreal), **ANDREW LUONG** (Laval University)

[Monday May 26 / lundi 26 mai, 14:00]

Optimal Portfolio Problem for Entropic Value at Risk: When the Underlying Distribution is Non Elliptical

Problème de portefeuille optimal pour la valeur à risque entropique : quand la distribution sous-jacente est non elliptique

In modern portfolio theory, we typically find asset returns that are modeled by a random variable with an elliptical distribution and the notion of portfolio risk is described by an appropriate risk measure. In this joint work, we propose a new stochastic model for the asset returns that are based on Jumps-Diffusion (J-D) distributions. On the other hand, we also propose to use a new coherent risk measure, so-called, Entropic Value at Risk (EVaR), in the optimization problem. For certain models, including a jump-diffusion distribution, this risk measure yields an explicit formula for the objective function so that the optimization problem can be solved without resorting to numerical approximations.

Dans la théorie moderne du portefeuille, nous trouvons généralement des rendements d'actifs qui sont modélisés par une variable aléatoire avec une distribution elliptique et la notion de risque de portefeuille est définie par une mesure de risque appropriée. Dans ce travail réalisé conjointement, nous proposons un nouveau modèle stochastique pour le rendement des actifs qui repose sur des distributions de diffusion avec sauts. D'autre part, nous recommandons aussi d'utiliser une nouvelle mesure cohérente du risque appelée valeur à risque entropique (VaRE) pour le problème d'optimisation. Pour certains modèles, comprenant une distribution de diffusion avec sauts, cette mesure du risque donne une formule explicite pour la fonction objective de manière à ce que le problème d'optimisation puisse être résolu sans avoir recours à des approximations numériques.

ETIENNE MARCEAU, HÉLÈNE COSSETTE, ETIENNE LARRIVÉE-HARDY & JULIEN TRUFINH (Université Laval)

[Monday May 26 / lundi 26 mai, 14:15]

Infinite-time Ruin Measures for Compound Renewal Risk Models with Dependence

Mesures de ruine pour les modèles de risque basés sur les processus de renouvellement composés avec dépendance

We study infinite-time ruin measures within risk models assuming dependence between interclaim times and claim amounts. We use change of measure techniques to obtain exact expressions for both the infinite-time ruin probability and the Gerber-Shiu discounted penalty function. Two different approaches are considered in the application of these techniques. Based on the expressions for the infinite-time ruin probability and the Gerber-Shiu function, we also derive Lundberg-type inequalities and asymptotic expressions for these two quantities. We investigate an important sampling method based on change of measure techniques. For specific bivariate distributions for the interclaim time and the claim amount, we derive their corresponding bivariate distributions resulting from the change of measure.

On étudie les mesures de ruine dans les modèles de risque qui supposent la dépendance entre les temps inter-sinistres et les montants des sinistres. On utilise les méthodes de changement de mesure afin d'obtenir des expressions exactes pour la probabilité de ruine et pour la fonction Gerber-Shiu. En se basant sur les expressions de la probabilité de ruine et de la fonction Gerber-Shiu, on développe aussi des inégalités de type Lundberg et des expressions asymptotiques pour ces deux quantités. De plus, on étudie une méthode d'échantillonnage préférentiel basée sur la technique de changement de mesure. Pour les distributions bivariées du couple temps inter-sinistre et montant de sinistre, on développe les lois bivariées correspondantes qui découlent du changement de mesure.

1D-C1: Actuarial Science and Finance 1

Actuariat et finance 1

SHANOJA NAIK & TAEHAN BAE (University of Regina)

[Monday May 26 / lundi 26 mai, 14:30]

Calibration of Wishart Stochastic Correlation Model for Sovereign Default Risk

Calage du modèle de corrélation stochastique de Wishart pour le risque de faillite souveraine

In the context of structural credit risk modeling, we consider a stochastic correlation model with Wishart process for sovereign credit risk. Different from corporate counterparties, there is no reliable proxy for the market value of a sovereign credit asset, which makes the model estimation problem challenging. We apply the extended Kalman Filter to calibrate the model, based on the term structure of survival probabilities which are implied by the market values of sovereign credit default swaps.

Dans un contexte de modélisation du risque de crédit structurel, nous étudions un modèle de corrélation stochastique en appliquant le processus de Wishart pour déterminer le risque de faillite souveraine. Contrairement aux contreparties de l'entreprise, il n'y a aucune approximation fiable de la valeur marchande de biens créditeurs souverains, ce qui complique le problème d'estimation du modèle. Nous appliquons le filtre étendu de Kalman pour calibrer le modèle en fonction de la structure des probabilités de survie implicites aux valeurs de marché des swaps sur défaillance souveraine.

CHENGGUO WENG (University of Waterloo)

[Monday May 26 / lundi 26 mai, 14:45]

CPPI under Regime Switching and Transaction Cost

CPPI en cas de changement de régime et coûts de transaction

Portfolio insurance has been extensively applied in the financial industry, typically designed to protect portfolio value from dropping below certain predetermined floor. It allows participation in risky investments for potential upside profits and meanwhile control downside risk. As one of the main portfolio insurance strategies, the CPPI adopts a simplified self-financing strategy to dynamically rebalance portfolio between a risky asset and a reserve asset. In this talk, I will introduce some results I recently obtained on CPPI in a discrete-time regime switching setting with transaction cost. Explicit formulas are derived for a variety of portfolio performance measures and can be easily implemented for sensitivity analysis. A numerical example with a real dataset will be given to demonstrate those obtained results.

L'assurance de portefeuille s'utilise de plus en plus dans l'industrie financière, notamment pour protéger la valeur du portefeuille contre une chute en dessous d'un certain seuil prédéterminé. Cette assurance permet la participation dans des investissements risqués pour les profits potentiels, tout en contrôlant le risque négatif. Le CPPI (assurance de portefeuille à proportion constante) est l'une des principales stratégies : il adopte une stratégie d'autofinancement simplifiée qui permet de rééquilibrer le portefeuille de manière dynamique entre un actif risqué et un actif de réserve. Dans cette présentation, j'introduis des résultats récemment obtenus sur le CPPI dans un environnement de changement de régime à moments discrets avec coûts de transaction. Je dérive des formules explicites pour une variété de mesures de performance du portefeuille, dont l'analyse de la sensibilité est facile à mettre en œuvre. Je donne un exemple numérique sur un ensemble de données réelles pour prouver les résultats obtenus.

1D-C2: Biostatistics - Applications 1 Biostatistique - applications 1

Chair/Président: Joseph Beyene (McMaster University)

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

FARNOOSH ABBAS AGHABABAZADEH, MAYER ALVO & DAVID BICKEL (University of Ottawa)

[Monday May 26 / lundi 26 mai, 13:30]

Optimal Reference Set for Estimating Local False Discovery Rate

Ensemble de référence optimal pour l'estimation du taux local de fausses découvertes

In genome-wide association studies, markers will be scanned across the complete set of genomes of individuals to find genetic variations associated with disease. A restriction of current methods used to express significance in GWAS is that they do not provide clear information about the probability that GWAS findings are true or false. The lack of some information is provided to consider a posterior probability that the marker is not associated with disease (LFDR). We estimate LFDR via improved estimates of the proportion of null hypothesis and the distribution of data. Therefore instead of considering all markers to estimate the distribution of data, we try to find optimal reference set as a group of markers to estimate unknown parameters and distributions.

Dans les études d'association pangénomique (GWAS), les marqueurs font l'objet d'un repérage à travers un ensemble complet de génomes d'individus pour détecter des variations génétiques associées à la maladie. Une des restrictions des méthodes courantes utilisées pour exprimer le seuil de signification des GWAS est qu'elles ne produisent pas une information claire sur la probabilité que les résultats des GWAS soient vrais ou faux. L'absence de certains éléments d'information est prévue pour examiner la probabilité a posteriori que le marqueur ne soit pas associé avec la maladie (LFDR). Nous estimons le LFDR au moyen d'estimations améliorées de la proportion de l'hypothèse nulle et de la distribution de données. Par conséquent, au lieu d'examiner tous les marqueurs pour estimer la distribution des données, nous essayons de trouver un ensemble de référence optimal comme groupe de marqueurs pour estimer des paramètres et des distributions inconnues.

KIAN JAHROMI & DIANLIANG DENG (University of Regina)

[Monday May 26 / lundi 26 mai, 13:45]

Stable Trend Detection for Gene Expression Under Multiple Biological Conditions

Détection d'une tendance stable pour l'expression génique dans des conditions biologiques multiples.

Temporal gene expression data contains rich information in characterization of gene function. However, the current temporal gene expressions usually have few measuring time series levels. Sometimes, the number of time points in gene expression data set is not large enough to use regular high-dimensional methods in this case it is better to reduce the size of data set to use classical statistical classification methods. For this reason relative change rate is used to reduce the dimension of time points. Moreover, several classical and high-

Les données temporelles d'expression génique contiennent des renseignements précieux sur la caractérisation des fonctions génétiques. Par contre, les expressions géniques temporelles courantes ont en général peu de niveaux de mesure de séries chronologiques. Dans les données d'expression génique, le nombre de points est parfois insuffisant pour utiliser des méthodes de grande dimension, auquel cas, il vaut mieux réduire la taille du jeu de données afin d'utiliser des méthodes de classification statistique classiques. Pour cette raison, un taux de changement relatif est utilisé pour réduire le nombre de points dans le temps. En outre,

1D-C2: Biostatistics - Applications 1

Biostatistique - applications 1

dimensional pairwise tests are compared with each other to find how reducing the size of genes can affect classification results.

Keywords: Relative change rate, Temporal gene expression, Classification, Behrens-Fisher Problem, Longitudinal data analysis

l'étude compare bon nombre de tests classiques et de tests de grande dimension deux à deux afin d'établir l'effet d'une réduction de taille des gènes sur les résultats de classification.

VANDA MILHEIRO LOURENÇO (FCT and CMA - NOVA University of Lisbon), **PAULO C. RODRIGUES** (Federal University of Bahia), **ANA M. PIRES** (IST - Universidade Técnica de Lisboa)

[Monday May 26 / lundi 26 mai, 14:00]

A Robust Coefficient of Determination for Heritability Estimation in Genetic Association Studies

Coefficient robuste de détermination pour l'estimation de l'héritabilité dans les études d'association génétique

Heritability is key in plant studies to help achieve better yield and other agronomic traits of interest. In candidate gene studies regression models are used to test for associations between phenotype and candidate SNPs. SNP imputation guarantees that marker information is complete and the data are balanced. So both the coefficient of determination, R^2 , and broad-sense heritability are equivalent. However, when the normality assumption is violated, the classical R^2 may be seriously affected. Recently two R^2 alternatives with good properties were proposed for the linear mixed model. We evaluate their performance under contamination and step forward a robust version of these coefficients assessing their adequacy for heritability estimation via simulation. An example of application is also presented.

La connaissance de l'héritabilité est essentielle pour les études sur les plantes afin d'obtenir un meilleur rendement et d'autres caractéristiques d'intérêt agronomique. Dans les études de gènes candidats, des modèles de régression sont utilisés pour tester les associations entre phénotypes et plusieurs SNP candidats. Après l'imputation des SNP, les données sont équilibrées. Donc le coefficient de détermination, R^2 , et l'héritabilité au sens large sont équivalentes. Toutefois, lorsque l'hypothèse de normalité est violée, le R^2 classique sera biaisé et non fiable. Récemment, deux alternatives à R^2 ayant de bonnes propriétés ont été proposées pour le modèle linéaire mixte. Nous évaluons par simulation leur adéquation pour l'estimation de l'héritabilité sous contamination, et nous présentons une version robuste de ces coefficients avec un exemple d'application.

NARGES NAZERI RAD & JERALD F LAWLESS (University of Waterloo)

[Monday May 26 / lundi 26 mai, 14:15]

Estimation of Prevalence Probabilities and Assessment of Multistate Models

Estimation des probabilités de prévalence et évaluation des modèles multi-états

Individuals in observational studies are often observed intermittently over time. When there is considerable variability in observation times across individuals, there is a strong possibility that these times are related to previous observations. The estimation of failure time distributions, prevalence probabilities, and other quantities can be biased in these situations. In this talk, we discuss this issue in the context of multistate modelling and analysis. Then, we propose a method to correct estima-

Dans les études d'observation, les individus sont souvent observés par intermittence dans le temps. Lorsque les temps d'observation varient considérablement d'un individu à l'autre, il est fort probable que ces temps soient liés aux observations précédentes. Dans de telles situations, l'estimation de la distribution des temps de défaillance, des probabilités de prévalence et d'autres quantités peut être biaisée. Nous discutons de ce problème dans le contexte de la modélisation et de l'analyse multi-états. Nous proposons ensuite une méthode pour corriger le biais d'esti-

1D-C2: Biostatistics - Applications 1

Biostatistique - applications 1

tion bias in the presence of dependent observation times. Finally, we present an application involving an analysis of time to viral failure (viral rebound) in HIV-positive persons whose virus has been suppressed by combination antiretroviral therapy.

mation en la présence de temps d'observation dépendants. Enfin, nous présentons une application avec une analyse du temps avant défaillance virale (ou rebond viral) chez des personnes séropositives dont le virus a été réprimé par un traitement antirétroviral d'association.

MARYLÈNE TROUPÉ & JEAN VAILLANT (Université des Antilles et de la Guyane), **LÉONIDE CÉLINI** (Université de Paris Est - Créteil)

[Monday May 26 / lundi 26 mai, 14:30]

Loglinear Models in the Presence of Overdispersion with Application to Aphis Gossypii Count Data

Modèles loglinéaires en présence de surdispersion. Application aux données de comptage d'Aphis Gossypii.

We considered overdispersed ecological individual count data. We first verified whether the overdispersion hypothesis was valid, and then, we tested environmental effects on the variability of individual numbers. Therefore, we used the loglinear model on *Aphis Gossypii* abundance data. Those data consist of information on occupancy leaf strata, storm events and developing stages of aphids, and are represented in a multiway contingency table. The results show a significant influence of each studied factor and a high percentage of variability explained by the model.

Nous nous sommes intéressés à des données de comptage d'individus écologiques sur dispersés. Après avoir vérifié le bien-fondé de l'hypothèse de surdispersion, nous avons testé les effets environnementaux sur la variation des effectifs. Pour cela, le modèle log-linéaire a été utilisé sur des données d'abondance du puceron du cotonnier *Aphis Gossypii*. Ces données concernent les strates foliaires d'occupation, des événements orageux et des étapes du mécanisme de développement du puceron. Elles sont présentées sous forme d'un tableau de contingence à plusieurs entrées. Les résultats obtenus indiquent une influence significative des facteurs étudiés et un fort pourcentage de variabilité expliquée par le modèle.

YISHU WANG & MICHAL ABRAHAMOWICZ (McGill University)

[Monday May 26 / lundi 26 mai, 14:45]

Assessing the Impact of Exposure Measurement Errors on the Results of Time-to-Event Analyses of Adverse Effects of Medications: a Simulation Study

Évaluation de l'impact des erreurs de mesures de l'exposition sur les résultats des analyses du temps avant événement des effets négatifs de médicaments : une étude par simulation

Administrative databases record prescriptions filled by the patients rather than their actual drug use. This induces exposure measurement errors, or discrepancies between actual drug use versus recorded drug exposure, due to, for example, partial or complete treatment non-adherence. We use simulations to assess the impact of ignoring different magnitudes of exposure measurement errors on the time-to-event analyses of adverse effects of medications. We generate event times for 5 different models linking drug exposure with adverse effects: (i) current use, (ii) current dose, (iii) cumulative use, (iv) cumulative dose, (v) cumulative dose in last 60 days. We then analyze the data ig-

Les bases de données administratives enregistrent les ordonnances remplies par les patients plutôt que leur consommation réelle de médicaments. Cela mène à des erreurs de mesure de l'exposition, c'est-à-dire des écarts entre la consommation réelle et l'exposition enregistrée, en raison notamment d'un non-respect partiel ou complet du traitement. Nous utilisons des simulations pour évaluer l'impact de la non prise en compte de diverses ampleurs de l'erreur de mesure de l'exposition sur l'analyse du temps avant événement des effets négatifs des médicaments. Nous générons des temps d'événement pour différents modèles qui relient l'exposition au médicament et les effets négatifs : (i) utilisation actuelle, (ii) dose actuelle, (iii) utilisation cumulée, (iv) dose cumulée, (v) dose cumulée sur 60 jours, puis nous analysons

1D-C2: Biostatistics - Applications 1

Biostatistique - applications 1

noring measurement errors. The impact is apparent as the magnitude of errors increases and has greater effects on cumulative drug linkage models.

les données sans tenir compte des erreurs de mesure. L'impact est clairement visible à mesure que l'ampleur des erreurs augmente et produit davantage d'effets sur les modèles de relation des médicaments cumulés.

1D-C3: Biostatistics - Bayesian Analysis Biostatistique - analyse bayésienne

Chair/Présidente: Kristian Thorlund (McMaster University)

Room/Salle: UC - Croft House

Abstracts/Résumés

AHALYA SIVATHAYALAN & PATRICK FARRELL (Carleton University), **DANIEL KREWSKI** (University of Ottawa)

[Monday May 26 / lundi 26 mai, 13:30]

Introduction and Assessment of a New PoD for Human Exposure Guidelines

Introduction et évaluation d'un nouveau point de départ pour les recommandations sur l'exposition humaine

The Hill model is commonly used in dose-response applications. Using this model, we introduce a new point of departure, that is based on the signal-to-noise crossover dose (SNCD) suggested by Sand et. al. (2011). We propose analytical, parametric bootstrap and non-parametric bootstrap methods to estimate this quantity. A Taylor series approximation is used to estimate the analytical bias and variability in the estimator. Subsequent bias correction to the confidence intervals significantly improve their coverage.

Le modèle de Hill est communément employé dans les applications de réponse au dosage. Sur la base de ce modèle, nous introduisons un nouveau point de départ fondé sur la dose de croisement signal-bruit (signal-to-noise crossover dose, ou SNCD) suggérée par Sand et. al. (2011). Nous proposons des méthodes analytiques, de bootstrap paramétrique et de bootstrap non paramétrique pour estimer cette quantité. Une approximation par séries de Taylor est utilisée pour estimer le biais analytique et la variabilité de l'estimateur. Une correction subséquente du biais des intervalles de confiance permet d'améliorer nettement leur couverture.

SAYANTEE JANA, JEMILA HAMID & NARAYANASWAMY BALAKRISHNAN (McMaster University)

[Monday May 26 / lundi 26 mai, 13:45]

A Bayesian Growth Curve Model for High Dimensional Data

Modèle bayésien de courbe de croissance pour des données de grande dimension

Traditional approaches of testing linear hypothesis on the mean of the Growth Curve Model fail in high-dimensional scenarios ($n < p$) since such setup leads to singular sample covariance matrix. Previously, we suggested a simple approach to moderation: the Moore-Penrose generalized-inverse. Although simulation studies showed that it worked well, but its performance declined in near singularity zone ($n \approx p$). Here, we propose a Bayesian growth, assuming Wishart prior for the unknown covariance matrix. The posterior distribution is then used to provide a shrinkage estimator for the covariance matrix, which was then used in our test statistic. Simulation studies show that the test is unbiased, powerful, symmetric, monotone with respect to n and parameter matrix and has level close

Les approches traditionnelles de tests d'hypothèse linéaire sur la moyenne d'un modèle de courbe de croissance échouent dans des scénarios de grande dimension ($n < p$), étant donné qu'un tel contexte mène à une matrice des covariances expérimentales singulières. Antérieurement, nous avons proposé une approche simple de modulation : l'inverse généralisé de Moore-Penrose. Même si des études de simulation ont montré qu'elle est efficace, son niveau de performance diminue près de la zone de singularité ($n \approx p$). Nous proposons ici une approche bayésienne, avec une loi a priori de Wishart pour une matrice de covariance inconnue. La distribution a posteriori est ensuite utilisée afin de produire un estimateur de rétrécissement pour la matrice de covariance, qui a ensuite servi dans notre analyse statistique. Des études de simulation montrent que le test est non biaisé, puissant, symétrique et monotone par rapport à n et à

1D-C3: Biostatistics - Bayesian Analysis Biostatistique - analyse bayésienne

to the nominal value.

la matrice des paramètres, avec un niveau près de la valeur nominale.

SARAH GAGLIANO & JO KNIGHT (Centre for Addiction and Mental Health), **MICHAEL WEALE** (King's College London), **MICHAEL BARNES** (Queen Mary University of London)

[Monday May 26 / lundi 26 mai, 14:00]

Investigation of Predictive Accuracy Measures for Genetic Models

Exploration des mesures de précision des modèles génétiques

The increasing quantity and quality of functional genomic information motivate the integration of these data with association data, including data from genome-wide association studies (GWAS). We used previously described GWAS signals ("hits") to train a regularized logistic model to predict genetic variant causality on the basis of a large multivariate functional dataset. We assessed the accuracy of the models for predicting genetic variant causality using various methodologies: receiver operating characteristic (ROC) curves, histograms of classification probability, and positive predictive values, among others. These have variable interpretability depending on factors such as whether the dataset is balanced in terms of numbers of genetic variants classified as hits versus those that are not. We investigate the validity of these techniques.

L'augmentation de la qualité et la quantité des éléments fonctionnels dans le génome encourage l'intégration de ces données avec l'information d'association telles que les données des études d'association en génomique. Nous avons utilisé des variations génétiques courantes liées à des maladies (« des signaux ») pour déceler des modèles logistiques qui prédisent la causalité des variations génétiques en utilisant plusieurs variables. Nous évaluons la précision des modèles avec plusieurs méthodes : la courbe ROC, les histogrammes et les valeurs prédictives positives, parmi d'autres. L'interprétation de ces méthodes varie selon des facteurs tels que la quantité des variations génétiques qui sont classifiées comme des signaux et celles qui n'ont pas cette classification. Nous examinons la validité de ces techniques.

KHADER KHADRAOUI (Université Laval)

[Monday May 26 / lundi 26 mai, 14:15]

Nonparametric Bayesian Regression under Combinations of Local Shape Constraints

Régression bayésienne non paramétrique sous combinaisons de contraintes locales de forme

A nonparametric Bayesian method for regression under combinations of local shape constraints is proposed. The shape constraints considered include monotonicity, concavity (or convexity), unimodality and in particular combinations of several types of range-restricted constraints. By using a B-spline basis, the support of the prior distribution is included in the set of piecewise polynomial functions. It is shown that the choice of number and position of interior spline knots have a profound effect on the smoothness of the estimated regression function. Furthermore, thanks to the local support property of B-splines, many combinations of constraints can easily be considered by identifying

Une méthode de régression bayésienne sous combinaisons de contraintes locales de forme est proposée. Les contraintes de forme considérées incluent la monotonie, la convexité (concavité), l'unimodalité et en particulier la combinaison de plusieurs contraintes localisées. En utilisant la base B-spline, le support de la loi a priori est inclus dans le support des fonctions polynomiales par morceaux. Le choix de la dimension et la position des nœuds intérieurs a un effet important sur le degré de lissage de la fonction de régression. Clairement, grâce à la propriété du support local des fonctions B-splines, plusieurs combinaisons de contraintes de forme localisées peuvent être considérées. Des simulations suivant la loi a posteriori sont réalisées par un algorithme de type MCMC.

1D-C3: Biostatistics - Bayesian Analysis Biostatistique - analyse bayésienne

B-splines whose support intersects with each constrained region. Simulations from the posterior distribution are obtained by an MCMC sampling algorithm.

ELAHEH TORKASHVAND, MOHAMMAD JAFARI JOZANI & MAHMOUD TORABI (University of Manitoba)

[Monday May 26 / lundi 26 mai, 14:30]

Pseudo-Empirical Bayes Estimation of Small Area Means Based on the James-Stein Estimation in Linear Regression Models with Functional Measurement Error

Estimation bayésienne pseudo-empirique des moyennes pour petites régions fondée sur l'estimation de James-Stein dans les modèles de régression linéaire avec erreur de mesure fonctionnelle

The maximum likelihood and method of moments techniques have been used in the literature for small area estimation problems with covariates subject to functional measurement error. In this paper, we first obtain the James-Stein estimator of the true covariate, subject to the functional measurement error. Then, we propose a new pseudo-empirical Bayes (*PEB*) predictor of small area means based on the James-Stein estimator. In addition, we show the *PEB* predictor is asymptotically optimal. The weighted and unweighted jackknife estimators of mean squared prediction error of the *PEB* predictor are also derived. Simulation studies are conducted to evaluate the performance of the proposed approach.

Le maximum de vraisemblance et la méthode des moments sont des techniques utilisées dans la littérature pour les problèmes d'estimation dans les petites régions où les covariables sont sujettes à une erreur de mesure fonctionnelle. Dans cette présentation, nous commençons par obtenir l'estimateur de James-Stein de la vraie covariable sujette à l'erreur de mesure fonctionnelle. Ensuite, nous proposons un nouveau prédicteur bayésien pseudo-empirique (BPE) des moyennes pour petites régions fondée sur l'estimateur de James-Stein. Puis nous montrons que le prédicteur BPE est asymptotiquement optimal. Nous dérivons également les estimateurs jackknife pondérés et non pondérés de l'erreur quadratique moyenne du prédicteur BPE. Nous effectuons des études par simulation pour évaluer la performance de l'approche proposée.

MARIA ESTHER PEREZ TREJO & ROBERT PLATT (McGill University)

[Monday May 26 / lundi 26 mai, 14:45]

Bayesian Modeling of Clustered Measured Binary Outcomes

Modélisation bayésienne des résultats binaires mesurés regroupés

In cluster trials, a source of variability that cannot be taken into account at the design stage occurs when the measurement of the outcome is inherently clustered within the same groups used as randomization units. The worse of these situations happens when there is a single observer per cluster measuring the outcomes. The variation due to clustering cannot then be separated from the variation due to the observer unless additional observations are available. A Bayesian hierarchical model is used to estimate the treatment effect for clustered measured binary responses. Simulation-based results are compared with their equivalent for Normal outcomes. Challenges arising from the anal-

Dans les essais en grappes, une source de variabilité qui ne peut être prise en considération lors de la conception de l'étude se présente lorsque la mesure du résultat est intrinsèquement regroupée au sein des groupes utilisés comme unités de randomisation. Dans le pire des cas, un seul observateur dans chaque groupe mesure les résultats. La variation due au regroupement ne peut alors plus être séparée de la variation due à l'observateur à moins que d'autres observations ne soient disponibles. Un modèle hiérarchique bayésien est utilisé pour estimer l'effet de traitement des réponses binaires mesurées regroupées. Les résultats de simulations sont comparés à leurs équivalents pour des résultats normaux. Les problèmes résultant de l'analyse des réponses binaires sont décrits. Les résultats de l'anal-

1D-C3: Biostatistics - Bayesian Analysis **Biostatistique - analyse bayésienne**

ysis of binary responses are described. Results of the statistical analysis from PROBIT experiment outcomes are presented.

yse statistique des résultats de l'expérience PROBIT sont présentés.

1D-C4: Statistical Inference and Applications 2 Inférence statistique et applications 2

Chair/Président: Zhou Zhou (University of Toronto)

Room/Salle: UC - 144

Abstracts/Résumés

SHIRIN GOLCHI & DAVID CAMPBELL (Simon Fraser University)

[Monday May 26 / lundi 26 mai, 13:30]

Sequentially-constrained Monte Carlo

Monte Carlo avec contraintes séquentielles

Imposition of constraints can be a source of challenge in Bayesian modelling. Constraint can be interpreted in a broad sense as any kind of explicit restriction over the model such as positivity of parameters or monotonicity of functions, adherence of the model to a deterministic system or a conservative selection criteria in approximate Bayesian computation. We propose a variant of sequential Monte Carlo algorithm for posterior sampling in presence of constraints. The specific parametrization of the constraints in the model is used to define a filtering sequence of distributions. Particles generated from an unconstrained or mildly constrained distribution are filtered and moved through sampling and re-sampling steps to obtain a sample from the fully constrained target distribution.

L'imposition de contraintes peut être une source de défi dans la modélisation bayésienne. Le terme contrainte peut être interprété au sens large comme étant n'importe quelle sorte de restriction explicite imposée à un modèle, tel que la positivité des paramètres ou la monotonie des fonctions, l'adhérence du modèle à un système déterministe ou un critère de sélection conservateur dans un calcul bayésien approximatif. Nous proposons une variante de l'algorithme séquentiel Monte Carlo pour l'échantillonnage a posteriori en présence de contraintes. Le paramétrage spécifique des contraintes dans le modèle est utilisé pour définir une séquence de filtrage des distributions. Les particules générées par une distribution sans contrainte ou avec de faibles contraintes sont filtrées et passent par les étapes d'échantillonnage et de rééchantillonnage pour obtenir un échantillon de la distribution cible entièrement sous contraintes.

AVIK HALDER & GLEN TAKAHARA (Queen's University)

[Monday May 26 / lundi 26 mai, 13:45]

On Simulation of the Beta-Dirichlet Process

Sur la simulation du processus beta-Dirichlet

The beta-Dirichlet process was introduced by Kim, James and Weissbach (2012). They have used this process as a prior for Bayesian analysis of cumulative hazard functions associated with any multistate event history data analysis. We will propose two new methods for simulation from beta-Dirichlet process and use them for estimation of the cumulative hazard functions associated with multistate event history data analysis. The first method is based on the Indian Buffet Process and the second method is based on an approximation proposed by Al Labadi (2012). We will compare these two sim-

Le processus beta-Dirichlet a été proposé par Kim, James et Weissbach (2012). Ils ont utilisé ce processus comme un préalable pour l'analyse bayésienne de fonctions cumulatives de risque (*hazard*) associées à toute analyse multi-état de données de l'historique d'événements. Nous proposerons deux nouvelles méthodes de simulation à partir du processus beta-Dirichlet et nous les utiliserons pour estimer les fonctions cumulatives de risque associées à l'analyse multi-état de données de l'historique d'événements. La première méthode est basée sur le processus du Buffet Indien et la seconde méthode se base sur une approximation proposée par Al Labadi (2012). Nous comparerons ces deux systèmes

1D-C4: Statistical Inference and Applications 2 Inférence statistique et applications 2

ulation schemes with one proposed by Kim, James and Weissbach (2012). The first method based on the Indian Buffet Process produces exact samples while the other methods produce approximate samples.

de simulation avec celui proposé par Kim, James et Weissbach (2012). La première méthode basée sur le processus du buffet indien produit des échantillons précis tandis que l'autre méthode produit des échantillons approximatifs.

WAN-CHEN LEE (University of Manitoba)

[Monday May 26 / lundi 26 mai, 14:00]

On Curved Exponential Imbedding and its Applications

À propos de l'imbrication de lois exponentielles courbes et de ses applications

In the regression model, the mean and variance of the response variable may be a function of the parameter such as $E(Y | X) = X\theta$ and $Var(Y | X) = \gamma(\theta)$. Under this case, the usual weighted least squares estimator is often an inconsistent estimator. The quasi likelihood estimator is an inefficient estimator, however, may be a consistent estimator. We propose an estimator based on curved exponential imbedding procedure. The estimator derived by this procedure has several optimal properties such as consistency and exponential rate of convergence. An example will be given to illustrate the procedure and results.

Dans un modèle de régression, la moyenne et la variance de la variable de réponse peut être une fonction du paramètre, comme $E(Y | X) = X\theta$ et $Var(Y | X) = \gamma(\theta)$. Dans un tel cas, l'estimateur aux moindres carrés est souvent incohérent. L'estimateur de quasi vraisemblance est peu efficace, toutefois, il peut être convergent. Nous proposons un estimateur fondé sur une procédure d'imbrication de lois exponentielles courbes. L'estimateur dérivé de cette procédure présente bon nombre de propriétés optimales et notamment, la convergence et le taux exponentiel de convergence. Un exemple illustre la procédure et les résultats.

FRANCOIS PERRON (University of Montreal), **AYI AJAVON** (Université de Montréal)

[Monday May 26 / lundi 26 mai, 14:15]

On the Extensions of Subcopulas

Sur les possibles prolongements d'une sous-copule

The Sklar theorem says that it is always possible to extend a subcopula to a copula. Particular extensions are known. We want to find all of the possible extensions. The solutions to this problem are known for the subcopula associated to discrete random variables. In the most general setting some solutions have been proposed but we found major errors in the proposed solutions. In this talk we will present our solution and we will explain why the other solutions are not fully general. We will also discuss the problem of finding the maximal copula extending a subcopula. The standard results, for copula related to discrete random variables involve series expansions. We will derive a fairly simple formula.

Le théorème de Sklar dit que toute sous-copule admet un prolongement en une copule. Des prolongements simples sont connus. On cherche à obtenir une façon de décrire tous les prolongements possibles. Notre problème est résolu dans le cas où la sous-copule est associée à des variables aléatoires discrètes. Pour des sous-copules quelconques des solutions décrivant tous les prolongements possibles existent mais les méthodologies proposées reposent sur des résultats erronnés. Dans cet exposé nous allons présenter notre solution. Nous allons aussi aborder le problème de trouver, parmi tous les prolongements possibles, celui qui atteint le maximum. Encore une fois, pour les sous-copules associées à des variables discrètes la solution est compliquée. Notre solution est autrement plus simple.

1D-C4: Statistical Inference and Applications 2

Inférence statistique et applications 2

SERGE PROVOST (The University of Western Ontario)

[Monday May 26 / lundi 26 mai, 14:30]

Improved Saddlepoint Density Approximations through Polynomial Adjustments
Approximations en point de selle de la densité par ajustements par polynôme

Improved density approximations are obtained by means of moment-based polynomial adjustments applied to the widely used saddlepoint approximation, the initial distributional support being determined from the Lugannani-Rice formula. Approximate percentiles, as evaluated from the original saddlepoint formula and its adjusted counterpart, are compared both numerically and graphically to their exact values in several illustrative examples. The bivariate case is addressed by applying a polynomial adjustment to the product of the approximated marginal densities of the standardized variables. Furthermore, extensions to the context of density estimation are formulated and applied to several univariate and bivariate data sets. Interestingly, the proposed methodology for approximating bivariate distributions gives rise to copula density functions that prove much more flexible than the conventional functional type.

Des approximations en point de selle de la densité sont améliorées grâce à un ajustement polynomial basé sur les moments appliqué à l'approximation en point de selle largement utilisée, la distribution initiale étant déterminée à partir de la formule de Lugannani-Rice. Des percentiles approximatifs, évalués à partir de la formule en point de selle originale et de sa contrepartie ajustée, sont comparés numériquement et graphiquement à leurs valeurs exactes dans plusieurs exemples descriptifs. Le cas bivarié est traité par application d'un ajustement polynomial au produit des densités marginales approximatives des variables standardisées. Par ailleurs, des généralisations au contexte de l'estimation de la densité sont formulées et appliquées à divers ensembles de données univariées et bivariées. Il est intéressant de noter que la méthodologie proposée pour l'approximation des distributions bivariées donne des fonctions de densité de copules beaucoup plus souples que le type fonctionnel courant.

AARON SPRINGFORD & DAVID THOMSON (Queen's University at Kingston)

[Monday May 26 / lundi 26 mai, 14:45]

Analysis of Time series with Unknown Times Using a Hierarchical Chronology Model
L'analyse des séries chronologiques avec temps latents employant un modèle hiérarchique

Time series analysis is well developed when sampling times are known, but latent times pose an additional challenge. For example, paleoenvironmental core data consist of samples collected at sequential depths in an extracted core (e.g. sediment, ice). The age of samples is unknown partly due to variation in deposition rate of the core material. A chronology model that relates depth to age is required in order for time series analysis to proceed. Early approaches included curve fitting and piecewise linear chronology models. More recent approaches model the deposition process but allow unrealistic deposition rates. I present a Bayesian hierarchical approach that regularises deposition rates at the prior level. The model is then incorporated into the overall time series analysis framework.

L'analyse des séries chronologiques est bien développée, mais les temps latents posent un défi supplémentaire. Par exemple, les carottes paléoenvironnementales (ex. de sédiments ou de glace) se composent d'échantillons obtenus à des profondeurs successives. L'âge des échantillons est inconnu parce que la déposition de la carotte n'est pas uniforme -- une modèle liant l'âge et la profondeur est nécessaire. Deux approches classiques sont les ajustements de courbes et les modèles linéaires par morceaux. Les approches modernes sont basés sur le processus de déposition, mais permettent des vitesses de déposition peu réalistes. Je présente un modèle bayésien hiérarchique qui régularise les vitesses de déposition au niveau a priori. Le modèle est ensuite incorporé dans la structure de l'analyse des séries chronologiques.

1D-C5: Survey Methodology 1 Méthodologie d'enquête 1

Chair/Président: Christian Boudreau (University of Waterloo)

Room/Salle: Hart House - South Dining

Abstracts/Résumés

CLAUDE GIRARD (Statistics Canada)

[Monday May 26 / lundi 26 mai, 13:30]

Can the Approximation to the Design-based Variance of Model Assisted Estimators Provided by Linearization be Trusted under Model Misspecification?

Peut-on toujours se fier à l'approximation obtenue par linéarisation de la vraie variance par rapport au plan de sondage d'estimateurs s'appuyant sur un modèle?

In a given survey context, one may have to choose between several competing estimators - one of which being the Generalized REGression estimator (GREG) - by comparing their design-based variances. Since GREG is model assisted but not model dependent, its true variance is valid regardless of the fit of the underlying linear model to the data. However, because GREG is a nonlinear estimator, such comparisons are made in practice based on an approximation of the variance obtained by linearization. We discuss an example showing that the quality of this approximation is model dependent: a misspecified model can lead to a very poor approximation of the variance of the GREG estimator, even when the sample size is large.

Dans un contexte d'enquête donné, on peut avoir à choisir entre plusieurs estimateurs -- l'estimateur par la régression généralisée (GREG) pouvant être l'un d'eux -- en comparant leurs variances par rapport au plan. Puisque GREG s'appuie sur un modèle mais n'en dépend pas, cette variance est valide que le choix du modèle qui le sous-tend soit judicieux ou non. Cependant, en raison de la non-linéarité de GREG, ces comparaisons se font en pratique sur la base d'une approximation de sa variance obtenue par linéarisation. Nous présentons un exemple montrant que la qualité de cette approximation dépend du modèle sous-jacent : un modèle mal avisé peut mener à une très pauvre approximation de sa variance même si la taille d'échantillon est grande.

CAREN HASLER & YVES TILLÉ (University of Neuchâtel)

[Monday May 26 / lundi 26 mai, 13:45]

Fast Balanced Sampling for Highly Stratified Population

Échantillonnage équilibré dans des populations hautement stratifiées

Balanced sampling is a very efficient sampling design when the variable of interest is correlated to the auxiliary variables on which the sample is balanced. A procedure to select balanced samples in stratified populations has previously been proposed. Unfortunately, this procedure becomes very slow as the number of strata increases and it even fails to select samples for some large numbers of strata. A new procedure to select balanced samples in stratified populations is proposed. This

L'échantillonnage équilibré s'avère être un plan d'échantillonnage très efficace lorsque la variable d'intérêt est corrélée aux variables auxiliaires sur lesquelles on équilibre l'échantillon. Une procédure pour sélectionner des échantillons équilibrés dans des populations stratifiées a auparavant été proposée. Cette procédure devient malheureusement très lente lorsque le nombre de strates augmente et ne permet plus de sélectionner des échantillons lorsque celui-ci dépasse une certaine limite. Un nouvel algorithme de sélection d'échantillons dans des populations stratifiées est proposé. Lorsque

1D-C5: Survey Methodology 1 Méthodologie d'enquête 1

procedure becomes much faster than the existing one when the number of strata increases and it makes it possible to select samples for some large numbers of strata, which was impossible with the existing method.

le nombre de strates est important, cette nouvelle procédure est plus rapide que celle existante. De plus, elle permet de sélectionner des échantillons dans des populations comportant de nombreuses strates et pour lesquelles la procédure existante ne le permettait pas.

WILSON LU (Acadia University)

[Monday May 26 / lundi 26 mai, 14:00]

Algorithms for Unequal Probability Sampling Without Replacement

Algorithmes pour l'échantillonnage sans remise à probabilités inégales

We will propose several sampling algorithms that match pre-determined first order inclusion probabilities, with one of them also matching joint inclusion probabilities. Some theoretical results will be presented as well.

Nous proposons plusieurs algorithmes d'échantillonnage qui respectent des probabilités d'inclusion du premier ordre prédéterminées, dont l'une concordant également aux probabilités d'inclusion conjointe. Des résultats théoriques sont également présentés.

XICHEN SHE & CHANGBAO WU (University of Waterloo)

[Monday May 26 / lundi 26 mai, 14:15]

Analysis of Complex Survey Data with Missing Ordinal Responses

Analyse de données d'enquêtes complexes avec réponses ordinales manquantes

Complex surveys are commonly used by social and health researchers. Many responses which measure knowledge, opinion and attitude, behaviour, psychological and health conditions are ordinal variables, with missing values frequently occurring, especially for variables with sensitive nature. In this talk, we first discuss statistical analysis of complex survey data with ordinal responses, with a focus on how to deal with missing values. We then explore methods on how to analyze survey data with missing ordinal responses when the survey designer also included "Don't Know" as part of the responses. Alternative approaches are compared and results from a limited simulation study are reported.

Les chercheurs en sciences sociales et de la santé ont souvent recours à des enquêtes complexes. Les réponses qui mesurent la connaissance, l'opinion et l'attitude, le comportement et les conditions psychologiques et de santé sont souvent des variables ordinales avec de nombreuses valeurs manquantes, surtout lorsque les variables sont de nature délicate. Dans cette présentation, nous discutons d'abord de l'analyse statistique des données d'enquêtes complexes avec réponses ordinales, en abordant le traitement des valeurs manquantes. Nous explorons ensuite des méthodes pour analyser les données d'enquête avec réponses ordinales manquantes dans le cas où le concepteur de l'enquête a également proposé « Ne sait pas » comme réponse. Nous comparons plusieurs approches et donnons les résultats d'une étude par simulation limitée.

AUDREY-ANNE VALLÉE & DAVID HAZIZA (Université de Montréal)

[Monday May 26 / lundi 26 mai, 14:30]

Approximation of the Variance in the Presence of Imputed Data

Approximation de la variance en présence de données imputées

Variance estimation in the case of item nonresponse treated by imputation is the main topic of this work. Treating the imputed values as if they

Les travaux porteront sur l'estimation de la variance dans le cas d'une non-réponse partielle traitée par une procédure d'imputation. Traiter les valeurs imputées comme si elles

1D-C5: Survey Methodology 1 Méthodologie d'enquête 1

were observed may lead to a substantial underestimation of the variance of point estimators. Classical variance estimators rely on the availability of the second-order inclusion probabilities, which are difficult to calculate. We propose to study the properties of variance estimators obtained by approximating the second-order inclusion probabilities. These approximations are usually valid for high entropy sampling designs. The results of a simulation study evaluating the properties of the proposed variance estimators in terms of bias and mean squared error will be presented.

avaient été observées peut mener à une sous-estimation substantielle de la variance des estimateurs ponctuels. Les estimateurs de variance usuels reposent sur la disponibilité des probabilités d'inclusion d'ordre deux, qui sont difficiles à calculer. Nous proposons d'examiner les propriétés d'estimateurs de variance obtenus au moyen d'approximations des probabilités d'inclusion d'ordre deux. Ces approximations sont généralement valides pour des plans à grande entropie. Les résultats d'une étude de simulation, évaluant les propriétés des estimateurs de variance proposés en termes de biais et d'erreur quadratique moyenne, seront présentés.

1E-I1: Delivering a Successful Introductory Stats Service Course Cours de service de base en statistique : meilleures pratiques

Chair/Président: Bruce Dunham (University of British Columbia)

Organizer/Responsable: Gordon Hines (University of Guelph)

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: Hart House - Debates Rm

Abstracts/Résumés

JEANETTE O'HARA HINES (University of Waterloo)

[Monday May 26 / lundi 26 mai, 15:30]

Statistics Service Courses and Evolving Undergraduate Science Education

Cours de service en statistique et évolution de l'enseignement scientifique du premier cycle

At the glacial pace of academic reform, a revolution is underway in undergraduate science education to active, collaborative and inquiry-based learning (IBL). The goal is students able to independently perceive the world, form hypotheses and infer weight of evidence. More than ever, service courses in statistics must prepare students effectively. Experience indicates that students' ability to apply statistical tools in IBL courses is greater when service courses do not reflect curricula familiar to instructors as the manner in which they learned statistics, but a flow more scientifically intuitive, i.e., from research hypotheses about observations, to experimental design (whether manipulative or observational), data collection and organization, estimation, statistical hypothesis testing and inference with respect to research hypotheses.

En comparaison avec la lenteur de la réforme universitaire, une révolution prend place actuellement dans l'enseignement scientifique universitaire et favorise l'approche active et collaborative de l'apprentissage par investigation. En voici l'objectif : les étudiants capables de percevoir indépendamment le monde formulent des hypothèses et infèrent le poids de la preuve. Plus que jamais, les cours de service en statistique doivent préparer efficacement les étudiants. L'expérience montre que l'habileté des étudiants à appliquer des outils statistiques dans des cours d'apprentissage par investigation est supérieure lorsque les cours de service diffèrent des programmes qui sont familiers aux formateurs et qui reflètent la manière dont ils ont étudié la statistique, et qu'ils font appel à un flux plus intuitif sur le plan scientifique, c'est-à-dire à partir d'hypothèses de recherche sur les observations, de plans expérimentaux (qu'ils soient conçus ou observés), de collecte et d'organisation des données, d'estimation, de tests d'hypothèse statistiques et d'inférence en regard des hypothèses de recherche.

GORDON HINES (University of Guelph), **R. JEANETTE O'HARA HINES** (Univeristy of Waterloo)

[Monday May 26 / lundi 26 mai, 16:00]

Useful Tactics For Teaching A Successful Stats Service Course

Stratégies utiles pour l'enseignement réussi d'un cours de service en statistique

The unconventional approach described in the previous talk inevitably provided new perspectives on various topics. This talk describes some useful tactics for discussing those topics, whether used in a

L'approche non traditionnelle décrite dans l'exposé précédent ouvre certainement la voie à des perspectives nouvelles sur divers sujets. Le présent exposé propose quelques stratégies utiles pour aborder ces sujets dans un cours, tra-

1E-I1: Delivering a Successful Introductory Stats Service Course Cours de service de base en statistique : meilleures pratiques

conventional course or in the suggested one.

Topics:

Template for a motivating first lecture; Surveys, case-control studies and experiments; Data types and grades, and selecting suitable methods; Introducing, using and comparing p-values; Relating statistical and practical significances of results; ANOVA diagrams that make ANOVA almost intuitive; Correlations for qualitative data; Extremely simple Principle Components.

ditionnel ou non.

Sujets :

Modèles pour un premier cours motivant ; Sondages, études cas-témoins et expériences ; Types et qualités de données, et sélection des méthodes appropriées ; Introduction, utilisation et comparaison des p-values ; Mise en relation de la signification statistique et pratique des résultats ; Diagrammes qui font de l'ANOVA une méthode d'analyse quasi intuitive ; Corrélations de données qualitatives ; Analyse en composantes principales très simple.

TOM NUDDS, GORDON HINES & BEREN ROBINSON (University of Guelph)

[Monday May 26 / lundi 26 mai, 16:30]

A Successful Structure for a Practical Introductory Statistics Service Course

Structure réussie pour un cours de service d'introduction à la statistique

By one measure, a service course in a subject succeeds if it prepares students to use the subject's relevant techniques effectively and appropriately. Focusing on that need results in a course of obvious relevance to students, even if it requires adjustments of material and presentation by instructors. We discuss an elementary, natural, integrated and (near-minimally) complete course flow that moves from research objectives to study design, data collection and summary, estimation, hypothesis testing and interpretations and discussions of practical conclusions.

The next talk discusses novel approaches to several topics developed for the service course created and delivered to meet student needs. The final talk discusses the (immediate and extended) success and relevance of that course.

Selon une définition possible, un cours de service est réussi s'il prépare les étudiants à utiliser efficacement et de façon appropriée des techniques pertinentes à un sujet. Lorsque l'accent est mis sur ce besoin, le cours présente un intérêt pour les étudiants, même s'il nécessite une adaptation de matériel et de présentation de la part des formateurs. Nous présentons un déroulement de cours quasi complet au niveau élémentaire, naturel et intégré qui comprend des objectifs de recherche, un protocole d'étude, la collecte de données, leur résumé, de l'estimation et des tests d'hypothèses, en plus d'interprétations et de discussions en matière de conclusions pratiques.

Le prochain exposé présente des approches novatrices de divers sujets élaborés pour un cours de service créé et présenté afin de répondre aux besoins de l'étudiant. L'exposé final discute du succès (immédiat et à plus long terme) et de la pertinence de ce cours.

1E-I2: Financial Mathematics Mathématique financière

Organizer and Chair / Responsable et président: Adam Kolkiewicz (University of Waterloo)

Room/Salle: Hart House - East Common

Abstracts/Résumés

MATT DAVISON & CHRISTIAN MAXWELL (Western University)

[Monday May 26 / lundi 26 mai, 15:30]

The impact of Regulatory Uncertainty on the Operation of Green Energy Projects
Impact de l'incertitude réglementaire sur le fonctionnement de projets d'énergie verte

The real options valuation framework is an exceptionally rich one for the analysis of energy projects, which can be modeled as strips of exchange options between inputs of one energy commodity and outputs of another. With subsidies that can change unexpectedly over time, green energy projects operate in a particularly complicated regulatory environment. Using a case study based on the analysis of a corn ethanol facility operating in the United States, we present an approach to modeling random changes in subsidy level and the approach taken by investors with various ambiguity aversions to operating a facility in face of this uncertainty and ambiguity. This is joint work with C. Maxwell (Applied Mathematics, Western University)

L'analyse par les options réelles constitue un contexte exceptionnellement riche pour l'analyse de projets énergétiques, qui peuvent être modélisés par des bandes d'options d'échange entre les intrants d'un bien énergétique et les extrants d'un autre. Compte tenu des subventions qui peuvent changer inopinément dans le temps, les projets d'énergie verte opèrent dans un contexte réglementaire particulièrement complexe. À partir d'une étude de cas fondée sur l'analyse d'une usine d'éthanol de maïs située aux États-Unis, nous présentons une méthode pour modéliser les changements aléatoires dans les niveaux de subventions et l'approche adoptée par les investisseurs à l'égard de diverses aversions de l'ambiguïté relativement à l'exploitation d'une installation confrontée à cette incertitude et ambiguïté. Ces travaux sont effectués conjointement avec C. Maxwell (Mathématiques appliquées, Western University).

ADAM METZLER (Wilfrid Laurier University), **R.M. REESOR** (Western University)

[Monday May 26 / lundi 26 mai, 16:00]

Pricing and Understanding Contingent Capital Bonds using Structural Models
Tarification et compréhension des obligations de capital contingent à l'aide de modèles structureux

We consider the valuation and analysis of zero-coupon contingent capital bonds (CCBs) in the structural framework. Using Doob's Optional Sampling Theorem (and making virtually no assumptions on asset value dynamics, the terms of conversion or the conversion trigger) we express the value of the CCB in terms of the effective loss imposed on CCB investors at conversion, quantify the impact that contingent capital has on traditional debt and equity, and describe how the conversion terms can be calibrated to ensure compliance with regulatory guidance. We provide numerical evidence

Nous considérons l'évaluation et l'analyse d'obligations à terme (zéro coupon) de capital contingent (OCC) dans un cadre structurel. À l'aide du théorème d'arrêt de Doob (en supposant aucune hypothèse sur la dynamique de la valeur des actifs, les conditions de conversion ou le déclencheur de la conversion), nous exprimons la valeur des OCC en termes de perte effective imposée aux investisseurs des OCC au moment de la conversion, nous quantifions l'impact causé par le capital contingent sur le financement classique par emprunt et par actions, et nous décrivons une façon de calibrer les conditions de conversion afin d'assurer la conformité avec les directives réglementaires. Nous présentons

1E-I2: Financial Mathematics Mathématique financière

indicating that the terms of conversion can fundamentally alter the nature of the CCB and suggests that a fixed conversion price may not be desirable when the conversion trigger involves regulatory discretion.

une preuve numérique indiquant que les conditions de conversion peuvent modifier fondamentalement la nature des OCC et qui suggèrent qu'un prix de conversion fixe ne serait pas souhaitable lorsque le déclencheur de la conversion fait appel à un pouvoir discrétionnaire en matière de réglementation.

ANATOLIY SWISHCHUK (University of Calgary), **GIOVANNI SALVI** (University of Rome 'La Sapienza')

[Monday May 26 / lundi 26 mai, 16:30]

Covariance and Correlation Swaps for Markov-modulated Volatilities

Swaps de covariance et de corrélation pour des volatilités

In this talk, we price covariance and correlation swaps for financial markets with Markov-modulated volatilities. As an example, we consider stochastic volatility driven by two-state continuous Markov chain. In this case, numerical example is presented for VIX and VXN volatility indices (S&P 500 and NASDAQ-100, respectively, since January 2004 to June 2012). We also use VIX (January 2004 to June 2012) to price variance and volatility swaps for the two-state Markov-modulated volatility and to present a numerical result in this case.

Dans cet exposé, nous évaluons la valeur de swaps de covariance et de corrélation pour les marchés financiers dont la volatilité est régi par une chaîne de Markov. Par exemple, nous considérons une volatilité stochastique déterminée par une chaîne de Markov continue à deux états. Dans ce cas, un exemple numérique est présenté pour les indices de volatilité VIX et VXN (respectivement S&P 500 et NASDAQ-100, de janvier 2004 à juin 2012). Nous utilisons aussi VIX (janvier 2004 à juin 2012) pour évaluer des swaps de variance et de volatilité pour la volatilité découlant de la chaîne de Markov continue à deux états et pour présenter un exemple numérique dans ce cas.

1E-13: New Developments in Nonlinear Time Series Modelling and Robust Methods in Time Series Analysis with Applications
Récents progrès en modélisation de séries chronologiques non linéaires; méthodes robustes d'analyse des séries chronologiques et applications

Organizer and Chair / Responsable et présidente: Ying Zhang (Acadia University)

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

HEROLD DEHLING (Ruhr-Universität Bochum, Germany), **ROLAND FRIED** (TU Dortmund), **AENEAS ROOCH** (Ruhr-University Bochum), **MURAD TAQQU** (Boston University), **MARTIN WENDLER** (Ruhr-University Bochum)
 [Monday May 26 / lundi 26 mai, 15:30]

Robust Change-Point Tests for Time Series

Tests de point de rupture robustes pour des séries chronologiques

We present recent developments on robust change-point tests for time series, both in the short-range as well as in the long-range dependent case. Given the data X_1, \dots, X_n , we test the hypothesis of a level shift at an unknown point in time. Our tests are based on common two-sample tests, such as the Wilcoxon test and the Hodges-Lehmann test. Specifically, we study the test statistics $\max_{1 \leq k \leq n-1} \sum_{i=1}^k \sum_{j=k+1}^n 1_{\{X_i \leq X_j\}}$, and $\max_{1 \leq k \leq n-1} \text{median}\{(X_j - X_i) : 1 \leq i \leq k < j \leq n\}$. We derive the asymptotic distribution of these test statistics, and more generally of two-sample U-processes and U-quantiles of dependent data, both under the hypothesis as well as under local alternatives.

Nous présentons des progrès récents en matière de tests de point de rupture robustes pour des séries chronologiques, dans les cas dépendants à courte et à longue portée. Étant donné les valeurs X_1, \dots, X_n , nous testons l'hypothèse d'un changement de niveau à un temps inconnu. Nos tests sont fondés sur des tests courants à deux échantillons, comme le test de Wilcoxon et celui de Hodges-Lehmann. En particulier, nous étudions les statistiques de $\max_{1 \leq k \leq n-1} \sum_{i=1}^k \sum_{j=k+1}^n 1_{\{X_i \leq X_j\}}$ et $\max_{1 \leq k \leq n-1} \text{median}\{(X_j - X_i) : 1 \leq i \leq k < j \leq n\}$. Nous obtenons la distribution asymptotique de ces statistiques de test et plus généralement de procédures en U et quantiles en U de données dépendantes sous l'hypothèse et selon des alternatives locales.

IAN MCLEOD & NAGHAM MOHAMMAD (Western University)

[Monday May 26 / lundi 26 mai, 16:00]

Censored Time Series Analysis

L'analyse des séries chronologiques censurées

Estimation and model checking for censored time series are discussed with applications.

Une discussion avec des applications au sujet de l'estimation et de la vérification des modèles des séries chronologiques censurées.

HAO YU, ZI ZHEN LIU & REG KULPERGER (Western University)

[Monday May 26 / lundi 26 mai, 16:30]

The PLAC-Weighted Adaptive LASSO for VAR(p) Models

LASSO adaptatif pondéré selon la fonction d'autocorrélation à décalage partiel pour les modèles VAR(p)

The LASSO (Tibshirani (1996)) and the adaptive LASSO (Zou (2006)) are popular in data analysis such as linear regression and Cox PH models for

Le LASSO (Tibshirani (1996)) et le LASSO adaptatif (Zou (2006)) sont des méthodes populaires en analyse de données, notamment les modèles de régression linéaire et de

1E-I3: New Developments in Nonlinear Time Series Modelling and Robust Methods in Time Series Analysis with Applications

Récents progrès en modélisation de séries chronologiques non linéaires; méthodes robustes d'analyse des séries chronologiques et applications

simultaneous variable selection and parameter estimation. Recently, the LASSO method has been applied to autoregressive (AR) and vector autoregressive (VAR) models. We propose the PLAC-weighted adaptive LASSO procedure for modelling VAR(p) processes. We formulate adaptive weights as functions of the norms of the sample partial lag autocorrelation (PLAC) matrix function (Heyse (1985)) and Yule-Walker estimates of a VAR model. The PLAC-weighted adaptive LASSO procedure achieves identification, selection and estimation all in one go. It shows promising results for VAR models.

Cox PH, pour la sélection de variables et l'estimation des paramètres effectuées de façon simultanée. Récemment, la méthode LASSO a été appliquée aux modèles autorégressif et vectoriel autorégressif (VAR). Nous proposons la procédure du LASSO adaptatif pondéré selon la fonction d'autocorrélation à décalage partiel (PLAC) pour la modélisation de processus VAR(p). Nous établissons les poids adaptatifs comme des fonctions normatives de la fonction matricielle PLAC de l'échantillon (Heyse (1985)) et des estimés Yule-Walker d'un modèle VAR. La méthode proposée permet d'obtenir l'identification, la sélection et l'estimation en même temps. Elle montre des résultats prometteurs pour les modèles VAR.

1E-15: Recent Developments in Small Area Estimation Récents progrès en estimation pour les petits domaines

Organizer and Chair / Responsable et président: Mahmoud Torabi (University of Manitoba)

Room/Salle: UC - A101

Abstracts/Résumés

JIMING JIANG (University of California, Davis), **SENKE CHEN** (Freddie Mac Foundation), **THUAN NGUYEN** (Oregon Health and Science University)

[Monday May 26 / lundi 26 mai, 15:30]

Observed Best Prediction for Small Area Counts

Meilleure prévision observée pour des décomptes de petits domaines

We extend the observed best prediction (OBP; Jiang, Nguyen & Rao 2011) method to small area estimation when the responses are counts at area-level. We show via a simulation study that the OBP outperforms the empirical best prediction method when the underlying model is misspecified. A bootstrap method is proposed for estimating the area-specific mean squared prediction error of the OBP conditioning on the small area mean counts. A real data example is considered.

Nous généralisons la méthode de meilleure prévision observée (OBP; Jiang, Nguyen et Rao, 2011) à une estimation pour les petits domaines lorsque les réponses sont des effectifs au niveau des domaines. Nous montrons à l'aide d'une étude par simulation que la méthode OBP surpasse la méthode de meilleure prévision empirique lorsque le modèle sous-jacent est mal spécifié. Nous proposons une méthode bootstrap pour estimer l'erreur quadratique moyenne de prévision spécifique aux petits domaines de l'OBP conditionnellement aux effectifs moyens des petits domaines. Nous examinons un exemple de données réelles.

GAURI DATTA (University of Georgia), **MAPLES JERRY** (US Census Bureau), **MAHMOUD TORABI** (University of Manitoba), **JIASHEN YOU** (US Census Bureau)

[Monday May 26 / lundi 26 mai, 16:00]

Best Predictive Small Area Estimation in Measurement Error Model

Meilleure estimation prédictive pour petits domaines dans un modèle d'erreur de mesure

Empirical best linear unbiased predictors (EBLUPs) are widely used in small area estimation. Improvement over direct estimators of EBLUPs, which trade bias for variance, stems from an assumed model. If the model is misspecified, the EBLUPs may be substantially biased resulting in a larger mean squared error. In a recent paper, Jiang et al. (JASA, 2011, pp. 732-745) proposed an alternative method of prediction of small area means, which they termed "overall best prediction" that enjoys superiority over EBLUPs when the model is misspecified, and remains asymptotically equivalent to EBLUPs, when the model holds. We generalize the method of Jiang et al. to handle measurement error in auxiliary variables when the working model is different from the true model.

Les meilleurs prédicteurs linéaires sans biais empirique (EBLUP) sont largement utilisés dans l'estimation pour petits domaines. Leur amélioration par rapport aux estimateurs directs des EBLUP, qui font un compromis entre le biais et la variance, résulte d'un modèle présumé. Si le modèle est mal spécifié, les EBLUP peuvent être biaisés de façon marquée et entraîner une erreur quadratique moyenne plus importante. Dans un article récent, Jiang et coll. (JASA, 2011, pp. 732-745) ont proposé une méthode de prévision différente pour les moyennes de petits domaines qu'ils ont appelée la «meilleure prévision globale», et qui est supérieure aux EBLUP lorsque le modèle est mal spécifié, tout en restant asymptotiquement équivalente aux EBLUP lorsque le modèle est vérifié. Nous généralisons la méthode de Jiang et coll. afin de traiter l'erreur de mesure des variables auxiliaires lorsque le modèle de travail est différent du vrai modèle.

1E-I5: Recent Developments in Small Area Estimation Récents progrès en estimation pour les petits domaines

THUAN NGUYEN (Oregon Health and Science University), **JIMING JIANG** (University of California, Davis), **J. SUNIL RAO** (University of Miami)

[Monday May 26 / lundi 26 mai, 16:30]

Observed Best Prediction via Nested-Error Regression with Potentially Misspecified Mean and Variance

Meilleure prévision observée par une régression à erreur emboîtée avec moyenne et variance mal spécifiées

We consider the observed best prediction (OBP; Jiang, Nguyen & Rao 2011) for small area estimation under the nested-error regression model, where both the mean and variance functions may be misspecified. We show via a simulation study that the OBP may significantly outperform the empirical best linear unbiased prediction method not just in the overall mean squared prediction error (MSPE) but also in the area-specific MSPE for every one of the small areas. A bootstrap method is proposed for estimating the design-based area-specific MSPE, which is simple and always produces positive MSPE estimates. The performance of the proposed MSPE estimator is evaluated through a simulation study. An application is considered.

Nous examinons la méthode de meilleure prédiction observée (OBP; Jiang, Nguyen et Rao, 2011) pour l'estimation de petits domaines sous un modèle de régression à erreur emboîtée, où les fonctions tant de moyenne que de variance peuvent être mal spécifiées. Dans une étude de simulation, nous montrons que la méthode OBP peut surpasser de façon significative celle des meilleurs prédicteurs linéaires sans biais empirique, non seulement dans l'erreur quadratique moyenne globale de prévision (MSPE), mais aussi dans les MSPE spécifiques au secteur pour chacun des secteurs restreints. Nous proposons une méthode bootstrap simple pour estimer le MSPE propre au domaine qui produit toujours des estimations de MSPE positives. Dans une étude de simulation, nous évaluons la performance de l'estimateur de MSPE proposé, et nous examinons une application.

1E-I6: Survival Methods in Genetics Méthodes de survie en génétique

Organizer and Chair / Responsable et président: M'Hamed Lajmi Lakhal Chaieb (Université Laval)

Room/Salle: UC - 161

Abstracts/Résumés

LAURENT BRIOLLAIS (Samuel Lunenfel Research Institute), **LAURENT BRIOLLAIS** (Lunenfeld-Tanenbaum Research Institute), **YUN-HEE CHOI** (Western University)

[Monday May 26 / lundi 26 mai, 15:30]

Statistical Challenges in Modeling Time-to-Event Data in Family Studies: Lessons Learned from the Analysis of Lynch Syndrome and Breast Cancer Families

Défis statistiques de la modélisation de temps avant l'événement dans les études familiales : Leçons apprises dans l'analyse de familles atteintes du syndrome de Lynch et du cancer du sein

Family studies have provided an unprecedented resource for identifying new genes underlying complex human diseases such as colorectal and breast cancer. Characterizing lifetime cancer risks and genetic risks in high-risk families is a major concern for genetic counselors but it raises many statistical challenges. The distribution of cancers in these families is complex with often, successive cancers observed in the same individual and various cancer types occurring among family members. The complex sampling of families, the presence of other risk factors besides the major gene, and missing genetic variables make risk estimation even more complicated. Based on our experience in analyzing large cancer families (Lynch, breast), we discuss various statistical issues and propose particular survival analysis approaches to address them.

Les études familiales ont constitué une ressource inégalée pour l'identification des nouveaux gènes sous-jacents aux maladies humaines complexes comme le cancer du sein ou le cancer colorectal. La caractérisation des risques à vie du cancer et des risques génétiques dans les familles à haut risque constitue une préoccupation majeure pour les conseillers en génétique, mais elle soulève bon nombre de défis statistiques. Dans les familles en cause, la distribution des cancers est complexe et souvent, des cancers successifs sont observés chez un même individu, et divers types de cancer surviennent parmi des membres d'une famille. L'échantillonnage complexe des familles, la présence de facteurs de risque autres que le gène principal, et l'absence de variables génétiques compliquent davantage l'estimation des risques. En fonction de notre expérience d'analyse des grandes familles de cancer (Lynch, sein), nous discutons divers problèmes statistiques et nous proposons des approches particulières à l'analyse des données de survie pour les aborder.

MARTIN LECLERC, JACQUES SIMARD & LAJMI LAKHAL-CHAIEB (Université Laval)

[Monday May 26 / lundi 26 mai, 16:00]

Ascertainment-Adjusted SNP Set Analysis for Clustered Age-at-Onset Outcomes

Analyse d'ensembles de SNPs ajustée pour le biais de détermination dans le cas de réponses d'âge d'apparition groupées

A high proportion of women carrying a mutation in the BRCA1 or BRCA2 gene develop breast cancer. So far, some SNPs modifying breast cancer risk among these women have been identified. How-

Une proportion élevée de femmes porteuses d'une mutation sur le gène BRCA1 ou BRCA2 développent un cancer du sein. Jusqu'à maintenant, quelques SNP modifiant le risque de cancer du sein chez ces femmes ont été identi-

1E-I6: Survival Methods in Genetics Méthodes de survie en génétique

ever, together they explain less than 10% of the polygenic variance observed. Few identified SNPs belonging to a same gene/region suggest the use of SNP-sets for breast cancer prediction, and within a survival framework as the risk modifiers are expected to influence occurrence of disease but also age at diagnosis. In this talk, we explore statistical approaches to testing for a gene/region effect in the context of BRCA1/2 data. Issues regarding intrafamily correlation, ascertainment bias, multiple outcomes (breast and ovarian cancer) and proportions of nonsusceptibles (cure fractions) are discussed.

fiés. Cependant, mis ensemble, ils expliquent moins de 10 % de la variance polygénique observée. Quelques SNP identifiés provenant d'un même gène/région donnent l'idée d'utiliser des ensembles de SNP pour prédire le risque de cancer du sein et ce, dans un contexte de durées de vie. Dans cette présentation, des méthodes statistiques pour tester l'effet d'un gène/région sont explorées dans le contexte des données BRCA1/2. Des questions concernant la corrélation intra-famille, le biais de détermination, les réponses multiples et la proportion de non-susceptibles sont discutées.

YILDIZ YILMAZ (Memorial University)

[Monday May 26 / lundi 26 mai, 16:30]

Analysis of Time-to-Event Phenotypes in Genetic Association Studies of Cancer

Analyse des phénotypes de temps avant l'événement dans les études d'association génétique sur le cancer

Time-to-event phenotypes typically arise in genetic association studies of complex traits in which individuals are followed through time and disease status is assessed longitudinally. For example, in the analysis of time to cancer recurrence of a type of cancer, it is of interest to identify clinical, pathological and molecular genetic markers that stratify newly diagnosed patients into subtypes differing in short- and long-term prognosis. Standard survival models can fail to describe adequately the relationship between subtype and cancer recurrence, particularly when there is a substantial proportion of long-term disease free survivors. I will discuss some appropriate modeling approaches and current issues in identifying novel genes involved in cancer recurrence following treatment.

Les phénotypes de temps avant l'événement proviennent surtout des études d'association génétique de traits complexes dans lesquelles des individus sont suivis au fil du temps, et leur état de maladie est évalué longitudinalement. À titre d'exemple, dans l'analyse du temps avant la récurrence d'un type de cancer, il est intéressant de reconnaître chez les patients nouvellement diagnostiqués les marqueurs génétiques cliniques, pathologiques et moléculaires qui les classifient en sous-types caractérisés par un pronostic à court ou à long terme. Les modèles de survie classiques n'arrivent pas toujours à décrire correctement les propriétés entre le sous-type et la récurrence d'un cancer, notamment en présence d'une proportion substantielle de survivants à long terme sans récurrence. Je discute des approches de modélisation appropriées ainsi que des problèmes courants d'identification des nouveaux gènes concernés lors d'une récurrence après traitement du cancer.

1E-C1: Biostatistics - Applications 2 Biostatistique - applications 2

Chair/Présidente: Katherine Davies (University of Manitoba)

Room/Salle: UC - 179

Abstracts/Résumés

OLAF BERKE (University of Guelph), **BIMAL CHHETRI** (Simon Fraser University), **ZVONIMIR POLJAK** (University of Guelph)

[Monday May 26 / lundi 26 mai, 15:30]

Diagnostic Misclassification Bias in Spatial Point Data Analysis -- A Simulation Study

Diagnostic des biais d'erreurs de classification dans une analyse de points de données spatiales - Une étude de simulation

Spatial statistical methods are widely used in disease surveillance systems, but their performance is not well understood when it comes to diagnostic misclassification.

The emergence of H3N2 influenza on Ontario swine farms is used as a scenario for a simulation study. The effect of diagnostic misclassification on the Cuzick-Edwards test for spatial clustering and spatial logistic regression modeling are investigated, assuming a true 10 percent prevalence among 550 simulated pork farms in southern Ontario.

The results show how varying values of herd-level sensitivity and specificity affect misclassification bias. The Cuzick-Edwards test might overlook spatial clustering due to misclassification bias; significance of binary and quantitative predictors in logistic regression models can vanish. In conclusion, spatial pattern get blurred by misclassification bias.

Les méthodes statistiques spatiales sont largement utilisées dans les systèmes de surveillance de maladies, mais leur performance n'est pas bien comprise quant au diagnostic des erreurs de classification.

L'émergence de l'influenza H3N2 dans les fermes porcines d'Ontario a été utilisée comme scénario pour une étude de simulation. L'effet du diagnostic des erreurs de classifications sur le test de regroupement spatial et de modélisation spatiale de régression logistique de Cuzick-Edwards a été étudié, en supposant une réelle prévalence de 10 pourcent parmi les 550 fermes porcines simulées dans le sud de l'Ontario.

Les résultats démontrent comment la variation des valeurs de sensibilité et de spécificité au niveau des troupeaux affecte le biais d'erreur de classification. Le test de Cuzick-Edwards peut ignorer un regroupement spatial en raison de biais due à une classification erronée ; la signification de prédicteurs binaires et quantitatifs dans des modèles de régression logistique peut disparaître. En conclusion, les structures spatiales peuvent être embrouillées par les biais de classifications erronées.

MOHMMAD KAVIUL KHAN, KATERINA MAXIMOVA & PAUL VEUGELERS (University of Alberta)

[Monday May 26 / lundi 26 mai, 15:45]

Are Overweight Children More Likely to Underestimate Their Weight Status: A Novel Approach to Quantify Weight Status Misperception

Les enfants atteints d'embonpoint sont-ils plus susceptibles de sous-estimer leur poids : une nouvelle approche pour quantifier les perceptions erronées de l'état pondéral

Despite high prevalence of childhood overweight and obesity, recent studies document that 70-80% of overweight children underestimate their weight

Malgré la prévalence élevée d'embonpoint et d'obésité chez les enfants, des études récentes révèlent que 70 % à 80 % des enfants atteints de surpoids sous-estiment leur état

1E-C1: Biostatistics - Applications 2

Biostatistique - applications 2

status. Weight status misperception is conceptualized as the discrepancy between actual (measured) and self-perceived weight status. However, quantification of weight status misperception remains challenging. To facilitate easier interpretation, we propose a novel approach to quantify weight status misperception using Stunkard body rating-scale and weight status categories (according to World Health Organization cut points) from measured height and weight among grade 5 students in Nova Scotia. Monte Carlo simulations were used to investigate whether overweight students are more likely to underestimate their weight status than normal weight students if students had selected their perceived weight status at random.

pondéral. Les perceptions erronées de l'état pondéral sont conceptualisées comme une différence entre le poids réel (mesuré) et le poids perçu. Toutefois, la quantification des perceptions erronées sur l'état pondéral continue de poser un défi. Pour en faciliter l'interprétation, nous proposons une nouvelle approche pour quantifier les perceptions erronées d'état pondéral en utilisant le barème de notation corporel de Stunkard et des catégories d'état pondéral (selon les seuils de l'Organisation mondiale de la santé) établies à partir des tailles et poids mesurés des élèves de cinquième année en Nouvelle-Écosse. Des simulations de Monte Carlo ont été utilisées pour déterminer si les élèves atteints de surpoids sont plus susceptibles de sous-estimer leur état pondéral que les élèves dont le poids est normal dans l'hypothèse que les élèves avaient sélectionné de façon aléatoire l'état pondéral perçu.

ZIHANG LU (The Hospital For Sick Children), **TETERINA ANASTASIA** (Toronto General Hospital)

[Monday May 26 / lundi 26 mai, 16:00]

Assessing the Performance of Different Indexes in Finding Cut-off Points for Continuous Biomarkers

Évaluation de la performance de divers indices afin de déterminer les seuils de biomarqueurs continus

The problem of choosing an appropriate cut-off point for a continuous biomarker to distinguish between diseased and healthy subjects is often encountered in clinical research. Currently, many indexes such as sensitivity, specificity, Youden Index, Gini index and total accuracy are used for this purpose. In many cases the researcher has to make decisions based on restricted sample due to low disease rates in the population. Using simulation, we investigate the performance of different indexes under different scenarios that include low and medium disease rates and different forms of predictors' distributions. For illustration, we are presenting a case study of predicting the presence of asthma in Canadian children.

La sélection d'un seuil approprié pour un biomarqueur continu afin de distinguer entre des sujets sains ou malades est un problème courant en recherche clinique. Actuellement, bon nombre d'indices sont utilisés à cette fin, et notamment les indices de sensibilité, de spécificité, de Youden, de Gini ou de précision totale. Dans plusieurs cas, le chercheur doit prendre des décisions basées sur un échantillon restreint en raison des faibles taux de maladie dans la population. À l'aide d'une simulation, nous investiguons la performance de différents indices sous divers scénarios comprenant des taux de maladie faibles et moyens, ainsi que des distributions de formes différentes pour le prédicteur. À des fins d'illustration, nous présentons une étude de cas pour prédire la présence d'asthme chez les enfants canadiens.

ABDOLLAH SAFARI (Simon Fraser University), **ROSA AGHDAM** (Shahid Beheshti University), **MEHDI SADEGHI & HAMID PEZESHK** (University of Tehran)

[Monday May 26 / lundi 26 mai, 16:15]

Generating a Random Gene Expression Dataset with a given Gene Regulatory Network

Comment générer un ensemble de données d'expression génique aléatoire avec un réseau de régulation génétique donné

1E-C1: Biostatistics - Applications 2

Biostatistique - applications 2

Inference of gene regulatory networks (GRN) from gene expression data is a major challenge in Biology. There are several methods to reconstruct a GRN from gene expression data sets. However, evaluating those methods is not easy since there is no true GRN for a given data set. We have provided an algorithm to generate a data set from a given GRN ("true GRN"). The generated data set was used to produce different GRNs which were then compared to the "true GRN". To produce those GRNs, correlation tests were applied to check relationships between genes, and networks from correlations were created by using different fixed thresholds.

L'inférence de réseaux de régulation génétique (RRG) à partir de données d'expression génique constitue un grand défi en biologie. Il existe plusieurs méthodes permettant de reconstruire un RRG à partir d'ensembles de données d'expression génique. Cependant, il n'est pas facile d'évaluer ces méthodes car il n'existe aucun RRG réel pour un ensemble donné. Nous proposons un algorithme capable de générer un ensemble de données à partir d'un RRG donné (RRG réel). Cet ensemble de données générées est utilisé afin de produire plusieurs RRG, qui sont comparés au RRG réel. Pour produire ces RRG, des tests de corrélation sont appliqués pour tester les relations entre gènes et des réseaux sont créés à partir de ces corrélations sur la base de plusieurs seuils fixes.

THUVA VANNIYASINGAM, CHARLES E. CUNNINGHAM, ALISON NICCOLS, GARY FOSTER, LEHANA THABANE & (CONNECTIONS) ON BEHALF OF CONNECTIONS (McMaster University)

[Monday May 26 / lundi 26 mai, 16:30]

A Discrete Choice Experiment of Evidence-based Practice Dissemination in Addiction Agencies for Women: An Empirical Comparison of Methods for Analyzing Clustered Data

Expérience à choix discret de la dissémination de pratiques fondées sur la preuve dans les organismes de toxicomanie pour les femmes : comparaison empirique de méthodes d'analyse de données regroupées

Efficient models of addiction services for the prevention and treatment of substance abuse among women are not effectively implemented by addiction agencies. A preference survey was conducted with 1379 service providers and administrators, from 333 addiction agencies, to explore the importance of 16 four-level attributes to effectively disseminate and implement evidence-based practices to addiction agencies. The purpose of this study was to empirically analyze discrete conjoint experimental data using various statistical methods to account for within-participant correlation --- specifically, multinomial logit and probit models.

Les organismes de toxicomanie ne recourent pas aux modèles de services de traitement efficaces qui existent pour prévenir et traiter la toxicomanie chez les femmes. Un sondage sur les préférences a été effectué auprès de 1379 fournisseurs de services et administrateurs de 333 organismes de toxicomanie, pour explorer l'importance de 16 attributs de quatre niveaux pour une dissémination et une implémentation efficaces des pratiques fondées sur la preuve dans les organismes de toxicomanie. L'objectif de cette étude est d'analyser empiriquement les données d'expérience discrètes conjointes à l'aide de diverses méthodes statistiques permettant de tenir compte de la corrélation en participants, notamment de modèles logit et probit multinomiaux.

YING WU & RICHARD COOK (University of Waterloo)

[Monday May 26 / lundi 26 mai, 16:45]

Variable Selection in Multi-State Analysis by Penalized Likelihood

Sélection de variables dans une analyse multi-états par vraisemblance pénalisée

We consider the problem of selecting among a high dimensional set of biomarkers, are prognostic for progression in chronic disease when such progression is characterized by a multi-state pro-

Nous étudions le problème de la sélection parmi un ensemble de biomarqueurs de grande dimension et le pronostic de progression d'une maladie chronique lorsque cette progression est caractérisée par un processus multi-états. Nous

1E-C1: Biostatistics - Applications 2

Biostatistique - applications 2

cess. We use the idea of penalized regression (e.g. LASSO, adaptive LASSO and SCAD) in the multi-state framework with multiplicative transition intensities. The baseline intensities are modeled using flexible piecewise constant functions. An expectation-maximization algorithm is developed which is shown to perform well empirically. An application to a motivating study is given in which the goal is to select important haplotype variables for predicting progression in joint damage in psoriatic arthritis.

utilisons l'idée de la régression pénalisée (p. ex. LASSO, LASSO adaptif et SCAD) dans le cadre multi-états avec des intensités de transition multiplicatives. Les intensités de base sont modélisées à l'aide de fonctions constantes par morceaux souples. Nous développons un algorithme EM dont nous prouvons la bonne performance empirique. Nous donnons une application à une étude dont l'objet est de sélectionner d'importantes variables d'haplotype pour prédire la progression des lésions articulaires dans l'arthrite psoriasique.

1E-C2: Biostatistics - Clinical Trials and Diagnostic Protocols Biostatistique - essais cliniques et protocoles diagnostiques

Chair/Présidente: Rhonda Rosychuk (University of Alberta)

Room/Salle: UC - 163

Abstracts/Résumés

CECILIA COTTON, OANA DANILA, STEFAN STEINER, DANIEL SEVERN & JOCK MACKAY (University of Waterloo)

[Monday May 26 / lundi 26 mai, 15:30]

Using Available Information in the Assessment of Diagnostic Protocols

Utiliser l'information disponible dans l'évaluation des protocoles de diagnostic

A new binary diagnostic test may be combined with an existing test to form either an add-on or triage protocol. We discuss methodological challenges in comparing the new combined protocol with the existing test using parallel or sequential assessment studies with complete verification via a gold standard. We also incorporate available information from a previous assessment study of the existing test. The comparison of the new protocol versus the existing test is made via likelihood ratios which capture the trade-off between sensitivity and specificity of a test or protocol. We compare estimation efficiency with and without baseline data and give planning advice for future studies. Methods are illustrated using mammography and ultrasound test results from a cohort of symptomatic women.

Un nouveau test de diagnostic binaire peut être combiné avec un test existant pour mettre sur pied, soit un protocole supplémentaire ou un protocole de triage. Nous discutons des défis méthodologiques dans la comparaison du nouveau protocole combiné avec le test existant en utilisant des études d'évaluation parallèles ou séquentielles avec une vérification complète à l'aide d'une méthode étalon. Nous incluons aussi de l'information provenant d'une étude précédente au test existant. La comparaison du nouveau protocole versus le test existant est effectuée à partir de ratios de vraisemblance qui capturent le compromis entre la sensibilité et la spécificité d'un test ou d'un protocole. Nous comparons l'efficacité des estimations avec et sans données de référence et nous donnons des conseils en matière de planification pour des études futures. Les méthodes sont illustrées à l'aide de résultats de mammographies et d'échographies d'un groupe de femmes symptomatiques.

SHUN FU LEE (Population Health Research Institute (PHRI)), **ALLAN DONNER & NEIL KLAR** (University of Western Ontario)

[Monday May 26 / lundi 26 mai, 15:45]

Power of the Adjusted Q Statistic for the Meta-Analysis of Cluster Randomization Trials

Puissance de la statistique Q ajustée pour la méta-analyse d'essais randomisés par grappes

Given that the adjusted Q statistic is frequently used for testing heterogeneity in the meta-analysis of cluster randomization trials, we derive an analytic expression for the power of this statistic focusing on binary outcomes. A simulation study is used to compare the performance of the adjusted Q statistic, in terms of Type I error and power, to that obtained from the proposed formula. It is found that the adjusted statistic has Type I error close to

Comme la statistique Q ajustée est souvent utilisée pour tester l'hétérogénéité d'une méta-analyse d'essais randomisés par grappe, nous dérivons une expression analytique pour la puissance de cette statistique qui met l'accent sur une variable réponse binaire. Une étude de simulation est utilisée pour comparer la performance de la statistique Q ajustée en ce qui a trait à l'erreur de type I et à la puissance, à celle obtenue avec la formule proposée. Il en résulte que l'erreur de type I de la statistique ajustée est près

1E-C2: Biostatistics - Clinical Trials and Diagnostic Protocols Biostatistique - essais cliniques et protocoles diagnostiques

the nominal level of 5%, as compared to the unadjusted Q statistic commonly used to test for heterogeneity in the meta-analysis of individually randomized trials. Data from a meta-analysis of four cluster randomization trials are used to illustrate the procedures.

du seuil nominal de 5 %, en comparaison avec la statistique Q non ajustée qui est couramment utilisée pour tester l'hétérogénéité d'une méta-analyse d'essais randomisés individuellement. Les données de la méta-analyse de quatre essais randomisés par grappes sont utilisées pour illustrer la procédure.

SELVAKKADUNKO SELVARATNAM, YANQING YI, ALWELL OYET & VEERESH GADAG (Memorial University of Newfoundland)

[Monday May 26 / lundi 26 mai, 16:00]

Estimation of Generalized Linear Mixed Model for Response Adaptive Designs of Multi-Center Clinical Trials

Estimation d'un modèle mixte linéaire généralisé pour des plans s'adaptant aux données dans le cas d'essais cliniques multicentres

Response adaptive designs are very important alternatives to Completely Randomized Design (CRD) in clinical trials because equal treatment allocation has been found to have ethical issues. Thus, in this paper we study the combination of response adaptive design and multi-center clinical trials. We develop a Generalized Linear Mixed Model (GLMM) for analyzing data from combining response adaptive design with multi-center clinical trials and discuss the estimation of the model parameters by Maximum Likelihood (ML). The asymptotic properties of the estimates are derived by applying influence function techniques. We examine the performance of our model, estimation method and asymptotics through a simulation study.

Les plans s'adaptant aux données sont de très importantes alternatives aux plans complètement aléatoires (PCA) dans les essais cliniques car l'affectation par traitement égal présente des problèmes éthiques. Dans cet article, nous étudions donc la combinaison de plans s'adaptant aux données et d'essais cliniques multicentres. Nous développons un modèle mixte linéaire généralisé afin d'analyser les données issues de ces plans s'adaptant aux données pour les essais cliniques multicentres et discutons de l'estimation des paramètres du modèle par maximum de vraisemblance. Nous développons les propriétés asymptotiques de ces estimations en appliquant des techniques de fonction d'influence. Nous examinons la performance de notre modèle, de cette méthode d'estimation et des asymptotiques via une étude par simulation.

LAN WEN, LEILEI ZENG & RICHARD COOK (University of Waterloo)

[Monday May 26 / lundi 26 mai, 16:15]

The Design and Analysis of Cluster-Randomized Longitudinal Trials Based on Transition Models

Conception et analyse d'essais longitudinaux randomisés en grappes fondés sur des modèles de transition

Cluster randomization is increasingly common in health, behavioural and clinical research where different interventions are allocated to clusters and the objective is to assess the cluster-level treatment effect. Often, cluster-randomized trials involve assessing each subject repeatedly over time after treatment assignments. Responses are then correlated longitudinally within each subject and cross-sectionally between subjects within a cluster. This has profound implications on the design and analysis as less efficient estimates can be obtained when these correlations are ignored. Focusing on the

La randomisation en grappes est de plus en plus utilisée en recherche clinique, sur la santé et sur le comportement où différentes interventions sont affectées à des grappes et l'objectif est d'évaluer l'effet de traitement au niveau de la grappe. Souvent, les essais randomisés en grappes exigent une évaluation répétée de chaque sujet dans le temps après l'avoir affecté à un traitement. Les réponses sont ensuite corrélées de manière longitudinale pour chaque sujet et de manière transversale entre sujets d'une même grappe. Cela a d'importantes implications pour la conception et l'analyse, car on obtient des estimations moins efficaces lorsque ces corrélations sont ignorées. En nous concentrant sur l'effet

1E-C2: Biostatistics - Clinical Trials and Diagnostic Protocols Biostatistique - essais cliniques et protocoles diagnostiques

treatment effect on the change in responses over time, we describe marginal transition models for the analyses of clustered longitudinal data and provide general formulas/software for sample size calculations. Empirical studies are also conducted to validate the proposed methods.

du traitement sur l'évolution des réponses dans le temps, nous décrivons des modèles de transition marginaux pour l'analyse des données longitudinales regroupées et offrons des formules/logiciels généraux pour le calcul de la taille d'échantillons. Nous effectuons des études empiriques pour valider les méthodes proposées.

YANQING YI (Memorial University of Newfoundland), **YANQING YI**

[Monday May 26 / lundi 26 mai, 16:30]

Dynamic Programming and Optimal Designs of Adaptive Clinical Trials

Programmation dynamique et plans optimaux pour les essais cliniques adaptatifs

Response adaptive designs have the ethical advantages over traditional randomization to benefit the patients in a clinical trial. However, the uncertainty introduced by adaptation of treatment allocation may lead to the loss of statistical power. We consider the costs of running a trial, type I error rate and statistical power simultaneously. The cost function of a trial could be the number of patients allocated to the inferior treatment. Our objective is to achieve the minimum cost while keeping a good statistical power and an adequate type I error rate. We establish a theorem to combine multiple objectives for response adaptive designs and prove the Bellman equation for the combined objective function. The properties of optimal strategies will be discussed.

Les plans s'adaptant aux données présentent des avantages éthiques par rapport à la randomisation traditionnelle, en ce qu'ils peuvent profiter aux patients participant à un essai clinique. Cependant, l'incertitude introduite par l'adaptation de l'affectation du traitement peut conduire à une perte de puissance statistique. Nous étudions simultanément les coûts d'exécution de l'essai, le taux d'erreur de type I et la puissance statistique. La fonction de coût de l'essai correspond au nombre de patients affectés au traitement inférieur. Notre objectif est de minimiser ce coût tout en maintenant une bonne puissance statistique et un taux d'erreur de type I adéquat. Nous établissons un théorème qui permet de combiner ces multiples objectifs pour les plans s'adaptant aux données et prouvons l'équation de Bellman pour la fonction d'objectifs combinés. Nous discutons des propriétés des stratégies optimales.

YUJIE ZHONG & RICHARD COOK (University of Waterloo)

[Monday May 26 / lundi 26 mai, 16:45]

Robust Design of Cluster-Randomized Trials With Censored Event Times

Plans robustes d'essais randomisés par grappes avec temps d'événements censurés

Cluster-randomized trials assign intact groups of individuals to one of several interventions, and the interest often lies in evaluating intervention effects. The intervention effects are often formulated by specifying marginal models, fitting them under a working independence assumption, and using robust variance estimates to address the within cluster association. We develop sample size criteria for cluster-randomized trials with right-censored time to event response. At the design stage, copula models are specified to facilitate the derivation of the asymptotic variance of parameter estimators. The required number of clusters are then obtained

Les essais randomisés par grappes affectent des groupes intacts d'individus à l'une de plusieurs interventions, l'objectif étant d'évaluer les effets de l'intervention. Ces effets d'intervention sont souvent formulés en spécifiant des modèles marginaux, en les ajustant dans le cadre d'une hypothèse d'indépendance de travail, puis en utilisant des estimations de variance robustes pour traiter l'association au sein de la grappe. Nous développons des critères de taille d'échantillon pour les essais randomisés par grappes avec réponse de temps avant événement censurée en aval. Au stade de la conception, nous spécifions des modèles de copule pour faciliter la dérivation de la variance asymptotique des estimateurs de paramètres. Nous obtenons alors assez de grappes

1E-C2: Biostatistics - Clinical Trials and Diagnostic Protocols **Biostatistique - essais cliniques et protocoles diagnostiques**

to ensure power objectives. Simulation studies are carried out to validate the accuracy of the sample size formula, and to explore the impact of misspecification of the copula functions.

pour garantir les objectifs de puissance. Nous effectuons des études par simulation pour valider la précision de la formule de taille d'échantillon et explorer l'impact de l'erreur de spécification des fonctions de copules.

1E-C3: Clustering, Classification and Dimension Reduction Partitionnement, classification et réduction de dimension

Chair/Président: Longhai Li (University of Saskatchewan)

Room/Salle: UC - 144

Abstracts/Résumés

NAOMI ALTMAN (Penn State University)

[Monday May 26 / lundi 26 mai, 15:30]

Extending PCA to Non-elliptical Data

Généralisation de l'ACP à des données non elliptiques

Principal component analysis plays an essential role in reducing the dimension of high dimensional problems. In this talk I argue that the ubiquitous role of PCA and its oft-noted robustness to non-normality of the data is due to its connection with the best L2 projections of the data matrix. Therefore useful extension of PCA should focus on generalizing L2 projections. Statistical models for extended PCA will also be introduced.

L'analyse en composantes principales (ACP) joue un rôle essentiel dans la réduction de dimension des problèmes de grande dimension. Dans cet exposé, je discute du rôle omniprésent de l'ACP et de sa robustesse souvent constatée en présence de non-normalité des données, qui résulte de son lien avec les meilleures projections L2 de la matrice des données. Par conséquent, une généralisation utile de l'ACP devrait mettre l'accent sur la généralisation des projections L2. Je présente des modèles statistiques pour généraliser l'ACP.

TORTORA CRISTINA, PAUL D. MCNICHOLAS & RYAN P. BROWNE (University of Guelph)

[Monday May 26 / lundi 26 mai, 15:45]

Mixture of Generalized Hyperbolic Factor Analyzers

Mélange d'analyseurs factoriels hyperboliques généralisés

The most widely used model-based clustering technique for high-dimensional data is the mixture of factor analyzers model. It is based on a latent Gaussian mixture model, and the aim of this talk is to extend it to the mixture of generalized hyperbolic factor analyzers (MGHFA) model. The mixture of generalized hyperbolic distributions is able to handle skewness and heavy tails, and it has the Gaussian, Student-t, normal inverse Gaussian and variance-gamma distributions as special cases. An alternating expectation-conditional maximization algorithm is used for parameter estimation and the Bayesian information criterion is used to select the number of factors and components. The performance of MGHFA is illustrated on real data sets, where it performs favorably when compared to its Gaussian analogue.

La technique de regroupement à l'aide de modèles la plus utilisée pour les données de grandes dimensions est le mélange du modèle d'analyseurs de facteurs. Cette technique est basée sur un modèle latent de mélange gaussien et cette présentation vise à l'étendre au modèle de mélange d'analyseurs factoriels hyperboliques généralisés (MAFHG). Le mélange d'analyseurs factoriels hyperboliques généralisés est capable de traiter l'asymétrie et les ailes lourdes. Celui-ci correspond aux distributions gaussiennes, t de Student, normalisée inverse gaussienne et variance-gamma comme cas spéciaux. Un algorithme alterné d'espérance conditionnelle de maximisation est utilisé pour l'estimation de paramètres et le critère d'information bayésien est utilisé pour choisir le nombre de facteurs et de composantes. L'efficacité du MAFHG est illustrée à l'aide des données réelles, où il se compare favorablement à son analogue gaussien.

1E-C3: Clustering, Classification and Dimension Reduction Partitionnement, classification et réduction de dimension

GEOFFROY MOURET, JEAN-JULES BRAULT & VAHID PARTOVI NIA (École Polytechnique de Montréal)

[Monday May 26 / lundi 26 mai, 16:00]

No Mean Clustering Algorithm

Algorithme de regroupement sans moyennes

The k -means clustering method is one of the frequently used unsupervised learning techniques. The main reasons for this popularity is its ease of understanding, computational efficiency, the wide applicability. We suggest a deeper view to unsupervised learning and derive the k -means algorithm as a variant of this broad view. This extension gives us the possibility of designing a k -mean algorithm with no step for computing the mean of the clusters.

L'algorithme des k -moyennes fait partie des méthodes d'apprentissage non-supervisé les plus utilisées. Les raisons de ce succès sont principalement sa simplicité, son efficacité computationnelle et son large champ d'application. Ce projet propose une approche bayésienne plus approfondie de l'apprentissage non-supervisé et présente une variante de l'algorithme des k -moyennes basée sur cette approche. Ce nouvel algorithme met en place une version des k -moyennes qui s'affranchit du calcul de la moyenne des grappes.

YANG TANG, RYAN BROWNE & PAUL MCNICHOLAS (University of Guelph)

[Monday May 26 / lundi 26 mai, 16:15]

Mixture of Latent Trait Models for High-Dimensional Categorical Data

Mélange de modèles de traits latents pour données catégoriques de grandes dimensions

We present a mixture of latent trait models with common slope parameters (MCLT) for high dimensional categorical data, a data type for which few established methods exist. Recent work on clustering of binary data, based on a d -dimensional Gaussian latent variable, is extended by implementing multinomial logistic response functions. The use of a logistic response function allows us to derive a variational EM algorithm which gives us a solvable EM algorithm of a latent variable density model. Fourteen covariance structures have been implemented to reduce further the number of parameters. Moreover, this approach allows a low-dimensional visual representation of the clusters. Our approach is illustrated using real and simulated data.

Nous présentons un mélange de modèles de traits latents avec des paramètres de pente communs pour des données catégoriques de grandes dimensions, type de données pour lequel il existe peu de méthodes établies. Les travaux récents sur le regroupement des données binaires, fondés sur une variable latente gaussienne de d dimensions, sont généralisés par l'emploi de fonction de réponse logistique multinomiales. L'utilisation d'une fonction de réponse logistique nous permet de dériver un algorithme EM variationnel qui nous donne un algorithme EM soluble d'un modèle de densité des variables latentes. Quatorze structures de covariance sont employées pour minimiser le nombre de paramètres. De plus, cette approche permet une représentation visuelle de faible dimension des grappes. Nous illustrons notre approche par des données réelles et simulées.

YUHONG WEI & PAUL MCNICHOLAS (University of Guelph)

[Monday May 26 / lundi 26 mai, 16:30]

Mixture Model Averaging for Clustering and Classification

Moyenne de modèles de mélange pour le groupage et la classification

In model-based clustering, it is common to fit a family of models and report clustering results from only the 'best' one. Rather than throw away all but the best model, we develop approaches to average

Dans le groupage par modèle, il est fréquent d'ajuster une famille de modèles puis de n'enregistrer que les résultats de groupage du « meilleur » modèle. Plutôt que de nous débarrasser de tous les modèles sauf le meilleur, nous

1E-C3: Clustering, Classification and Dimension Reduction Partitionnement, classification et réduction de dimension

multiple models that are in some sense close to the best one, thereby producing a weighted average of clustering results. Two averaging approaches are considered: 1) averaging the component membership probabilities, and 2) averaging models. In both cases, Occam's window is used to determine closeness to the best model and weights are computed within a Bayesian model averaging paradigm. In some cases, we need to merge components before averaging and we introduce a method for merging mixture components based on the adjusted Rand index.

mettons au point des approches permettant de faire la moyenne de plusieurs modèles qui sont proches du meilleur, produisant ainsi une moyenne pondérée des résultats du groupage. Nous étudions deux approches pour calculer cette moyenne : moyenne des probabilités des éléments membres et moyenne des modèles. Dans les deux cas, nous utilisons la fenêtre d'Occam pour déterminer la proximité au meilleur modèle et nous calculons la pondération dans un paradigme de moyenne de modèle bayésien. Dans certains cas, nous devons fusionner les éléments avant de calculer la moyenne : nous proposons une méthode pour ce faire fondée sur l'indice de Rand ajusté.

LU XIN & MU ZHU (University of Waterloo), **HUGH CHIPMAN** (Acadia University)

[Monday May 26 / lundi 26 mai, 16:45]

Continuous-time Stochastic Blockmodels for Transactional Networks

Modèles stochastiques par blocs en temps continu pour les réseaux transactionnels

Transactional network data describe communications among actors in a social network. Many have used the stochastic blockmodel to detect communities in such networks. However, they either do not take time into account or treat time in a discrete manner (as in discrete-time Markov chains). We introduce continuous-time stochastic blockmodels. Using ideas from mixture modeling and recurrent event analysis, our model is capable of not only detecting communities but also capturing how communication patterns evolve among communities. We demonstrate the advantages of our model with some NBA basketball data.

Les données de réseaux transactionnels décrivent les communications entre les acteurs d'un réseau social. Plusieurs chercheurs ont utilisé des modèles stochastiques par blocs pour détecter les communautés de ces réseaux. Par contre, soit ils ne tiennent pas compte du temps, soit ils le traitent de façon discrète (avec une chaîne de Markov en temps discret). Nous proposons un modèle stochastique par blocs en temps continu. En exploitant des idées puisées dans la modélisation par mélanges et l'analyse d'événements récurrents, notre modèle est non seulement capable de détecter des communautés, mais il permet aussi de décrire l'évolution des schémas de communications au sein de ces communautés. Nous illustrons les avantages de notre modèle à partir de données réelles sur le basketball de la ligue NBA.

1E-C4: Probability Theory and Estimation Théorie des probabilités et estimation

Chair/Président: Deli Li (Lakehead University)

Room/Salle: UC - Croft House

Abstracts/Résumés

ERNEST DANKWA (Memorial University)

[Monday May 26 / lundi 26 mai, 15:30]

Bias Study of the Naive Estimator in a Longitudinal Binary Mixed-effects Model with Measurement Error and Misclassification in Covariates

Étude des biais de l'estimateur naïf dans un modèle longitudinal binaire à effets mixtes avec les erreurs de mesure et de classification dans les covariables

When covariates in Longitudinal data are subject to error, the naive estimates of the model parameters are often biased. In this research, we exploit a Dynamic Binary Mixed-effects Model using a Generalized Quasi-likelihood approach. By simulations, we shall study the patterns in the bias of the naive estimates of the parameters that ignore the errors in the covariates.

Quand les covariables de données longitudinales sont sujettes à l'erreur, les estimations naïves des paramètres du modèle sont souvent biaisées. Dans cette recherche, nous explorons un modèle binaire dynamique à effets mixtes à l'aide d'une approche de quasi-vraisemblance généralisée. Nous étudierons, à l'aide de simulations, les schémas dans le biais des estimations naïves des paramètres qui négligent les erreurs des covariables.

SHAKHAWAT HOSSAIN (University of Winnipeg), **EJAZ AHMED** (Brock University), **GRACE YI** (University of Waterloo), **BAOJIANG CHEN** (University of Nebraska Medical Center)

[Monday May 26 / lundi 26 mai, 15:45]

Shrinkage and Pretest Estimators in Partially Linear Models for Longitudinal Data

Estimateurs de rétrécissement et d'essai préliminaire dans des modèles partiellement linéaires de données longitudinales

We consider the partially linear models for longitudinal data. We employ the pretest and shrinkage estimation procedures to estimate the mean response and the association parameters by accounting for within-subject correlation when some of the parameters may satisfy certain restrictions. We provide the expressions for the asymptotic biases and risks of pretest and shrinkage estimators. A Simulation study is conducted to evaluate the relative performance of the estimators with the semiparametric least squares estimator(USLSE). We show that if the dimension of parameters exceeds two, the risk of shrinkage estimators are strictly less than the USLSE and the risk of pretest estimator depends on the validity of restrictions. A real data example is used to illustrate the application of esti-

Nous analysons des modèles partiellement linéaires de données longitudinales. Nous utilisons les procédures d'estimation de rétrécissement et d'essai préliminaire pour estimer la réponse moyenne et les paramètres d'association en tenant compte de la corrélation pour un même sujet lorsque quelques-uns des paramètres peuvent satisfaire à certaines contraintes. Nous proposons des expressions pour le biais et le risque asymptotiques des estimateurs d'essai préliminaire et de rétrécissement. Nous menons une étude de simulation pour évaluer la performance relative des estimateurs avec la méthode d'estimation semiparamétrique aux moindres carrés (MEMCS). Nous montrons que lorsque la dimension des paramètres est supérieure à deux, le risque des estimateurs de rétrécissement est strictement inférieur à celui de la MEMCS, et le risque de l'estimateur de rétrécissement dépend de la validité des contraintes. Nous présentons un

1E-C4: Probability Theory and Estimation Théorie des probabilités et estimation

mation procedures.

exemple basé sur des données réelles pour illustrer l'application des procédures d'estimation.

MASOUD M NASARI & MIKLOS CSORGO (Carleton University)

[Monday May 26 / lundi 26 mai, 16:00]

On the Accuracy of the Bootstrap

À propos de la précision du bootstrap

In this talk we study the random error generated by approximating the sampling distribution of the t -statistic by the conditional, given the random sample X_1, \dots, X_n , $n \geq 1$, distribution of its bootstrap version. We present some theoretical results, supported by numerical illustrations, which indicate that the bootstrap approximation of the sampling distribution of the t -statistic is of the same magnitude of error as that of the classical CLT, i.e., $O(1/\sqrt{n})$.

Dans cet exposé, nous étudions l'erreur aléatoire générée par une approximation conditionnelle de la distribution d'échantillonnage de statistique t , compte tenu de la distribution de l'échantillon aléatoire X_1, \dots, X_n , $n \geq 1$, dans sa version bootstrap. Nous présentons des résultats théoriques, corroborés par des illustrations numériques, qui révèlent que l'approximation bootstrap de la distribution d'échantillonnage de la statistique t présente une marge d'erreur équivalente à celle du TCL classique, soit $O(1/\sqrt{n})$.

MATTHEW SCHMIRLER (University of Saskatchewan)

[Monday May 26 / lundi 26 mai, 16:15]

Strand Passage Probabilities in an Interacting Self-Avoiding Polygon Model

Probabilités de passage de brins dans un modèle de polygones auto-évitant interactif

We model here the action of the enzyme type-II topoisomerase on circular DNA using self-avoiding polygons (SAPs) in the simple and face-centred cubic lattices. Samples of essentially independent SAPs containing a specific fixed strand passage structure are generated via Composite Markov Chain Monte Carlo simulations. Strand passages are performed on these SAPs, the results of which are used to estimate probabilities of changing knot types and overall knot reduction after a strand passage.

Nous modélisons ici l'action de l'enzyme topoisomérase de type II sur l'ADN circulaire à l'aide de polygones auto-évitants (PAÉ) dans les treillis cubiques simples et à faces centrées. Des échantillons de PAÉ essentiellement indépendants contenant une structure de passage de brins fixe spécifique sont générés à l'aide de simulations de Monte Carlo par chaînes de Markov. Des passages de brins sont effectués sur ces PAÉ ; nous en utilisons les résultats pour estimer la probabilité de la modification du type de nœud et la réduction du nombre total de nœuds après un passage de brins.

TREVOR THOMSON, SHAKHAWAT HOSSAIN & MELODY GHAHRAMANI (University of Winnipeg)

[Monday May 26 / lundi 26 mai, 16:30]

Application of Shrinkage Estimation in Linear Regression Models with Autoregressive Errors

Application de l'estimation du rétrécissement dans les modèles de régression linéaire avec erreurs autorégressives

In this paper, we consider the shrinkage procedure in the linear regression model with autoregressive errors when it is conjectured that some of the regression parameters are inactive. We develop statistical properties of the shrinkage estima-

Dans cette présentation, nous étudions la procédure de rétrécissement dans le modèle de régression linéaire avec erreurs autorégressives dans l'hypothèse où certains des paramètres de régression sont inactifs. Nous développons les propriétés statistiques de la méthode d'estimation du

1E-C4: Probability Theory and Estimation Théorie des probabilités et estimation

tion method including asymptotic biases and risks. We show that the shrinkage estimator has a significantly higher relative efficiency than the classical estimator. A simulation study is conducted for different combinations of inactive predictors and the performance of each estimator is evaluated in terms of the simulated mean squared error. This study shows that the shrinkage estimator dominates the classical estimator in the entire parameter space. The shrinkage method is applied to a real data set to illustrate the usefulness of these procedures in practice.

rétrécissement, notamment les biais asymptotiques et les risques. Nous montrons que l'estimateur du rétrécissement présente une efficacité relative nettement supérieure à l'estimateur classique. Nous effectuons une étude par simulation pour différentes combinaisons de prédicteurs inactifs et évaluons la performance de chaque estimateur en ce qui concerne l'erreur quadratique moyenne simulée. L'étude montre que l'estimateur du rétrécissement domine sur l'estimateur classique dans tout l'espace de paramètres. La méthode du rétrécissement est appliquée à un ensemble de données réelles pour illustrer l'utilité de cette procédure dans la pratique.

MONICA H.T. WONG & PAUL D. MCNICHOLAS (University of Guelph)

[Monday May 26 / lundi 26 mai, 16:45]

Biclustering of Gene Expression Data with Applications in Nutrigenomics

Bipartitionnement parcimonieux des données d'expression de gènes avec des applications en nutriginomique

A family of Gaussian mixture models for the biclustering of high-dimensional gene expression data is introduced. Previously, a biclustering algorithm was introduced whereby the mixtures of factor analyzers model was extended to include a binary and row stochastic factor loadings matrix. This particular form of factor loadings matrix resulted in a block-diagonal covariance matrix. Our family is derived by imposing a factor analysis covariance structure and subsequently imposing constraints thereon. Parameter estimates are obtained through a variant of the expectation-maximization algorithm and model selection is performed via the Bayesian information criterion.

Une nouvelle famille de modèles de mélanges gaussiens pour le bipartitionnement des données d'expression de gènes à grande dimension est proposée. Auparavant, le modèle de mélange d'analyse factorielle a été perfectionné afin d'inclure une matrice de saturation binaire et stochastique pour se conformer à l'algorithme de bipartitionnement. Cette forme particulière de la matrice de saturation s'en suit de la matrice de covariance par blocs diagonales. Notre famille est obtenue en imposant une structure de covariance à l'analyse factorielle et ultérieurement, en imposant des contraintes. Les paramètres sont estimés par une variante de l'algorithme espérance-maximisation et le choix du modèle est accompli avec le critère de l'information bayésien.

1E-C5: Statistical Inference and Applications 5 Inférence statistique et applications 5

Chair/Président: Saman Muthukumarana (University of Manitoba)

Room/Salle: UC - 244

Abstracts/Résumés

AMY MORIN & ALISHA ALBERT-GREEN (University of Western Ontario), **DOUGLAS WOOLFORD** (Wilfrid Laurier University), **DAVID MARTELL** (University of Toronto)

[Monday May 26 / lundi 26 mai, 15:30]

Analyzing Forest Fire Lifetime Data Using Proportional Hazards Frailty Models

Analyse des cycles de vie des feux de forêt à l'aide des modèles de fragilité à risques proportionnels

The lifetime of a wildland fire depends on locally-observed covariates, such as fire weather variables, as well as fire management variables. Understanding how these affect the duration of a fire can lead to more effective suppression efforts as this information can be incorporated into decision support systems used by fire management agencies to help allocate resources and prioritize suppression efforts. We present a survival analysis of historical forest fire data from the Intensive Fire Management zone of Ontario and an investigation of spatial patterns across this region using proportional hazards Gaussian shared frailty models. The resulting posterior estimates suggest spatial patterns across this zone.

Le cycle de vie des feux de végétation est fonction des covariables observées à l'échelle locale, notamment des variables météorologiques propices aux incendies forestiers et des variables relatives à la gestion des incendies. Une compréhension claire de l'effet de ces variables sur la durée d'un incendie peut favoriser l'efficacité des moyens de suppression, puisque cette information peut être incorporée dans les systèmes de support décisionnel utilisés par les agences de gestion des incendies pour faciliter l'allocation des ressources et accorder la priorité aux efforts de suppression. Nous présentons une analyse de survie des données sur les feux de forêt historiques de la zone de gestion intensive des incendies de l'Ontario, ainsi qu'une investigation des structures spatiales de la région à l'aide des modèles à risques proportionnels gaussiens de fragilité partagée. Les estimations a posteriori qui en résultent proposent des structures spatiales pour la région.

WEI LIN, DON FRASER & NANCY REID (University of Toronto)

[Monday May 26 / lundi 26 mai, 15:45]

Adjustments to r : Insides from Taylor Expansions

Ajustements à r : le fonctionnement du développement en série de Taylor

The likelihood ratio test and the signed likelihood root (r) are two of the most commonly used statistics for inference in parametric models, with the r often viewed as the most reliable. A third-order adjusted signed likelihood root is available from likelihood theory, but the formulas and development methods are not always easily implemented and understood. Using a log-model Taylor expansion, we develop simple second order additive adjustments to the r . For a scalar linear interest param-

Le test du rapport de vraisemblance et la racine signée de vraisemblance (r) sont deux des statistiques d'inférence les plus courantes dans les modèles paramétriques, avec r souvent considérée comme la plus fiable. Une racine signée de vraisemblance ajustée en fonction du troisième ordre est disponible grâce à la théorie de la vraisemblance, mais les formules et les méthodes de développement ne sont pas toujours facilement appliquées et comprises. En utilisant un modèle logarithmique de développement en série de Taylor, nous avons formulé pour r des ajustements additifs simples

1E-C5: Statistical Inference and Applications 5

Inférence statistique et applications 5

ter, the adjustment to r are easy to calculate and easy to explain. In the curve interest parameter case, adjustment is available but need more calculations. The theory is developed and simulations are conducted to indicate the repetition accuracy.

du deuxième ordre. Pour un paramètre scalaire linéaire d'intérêt, l'ajustement de r est facile à calculer et à expliquer. Dans le cas d'un paramètre courbe, il est possible de faire l'ajustement, mais il nécessite d'autres calculs. La théorie est élaborée et les simulations ont été effectuées pour refléter la précision de répétition.

MATTHIEU WILHELM (Université de Neuchâtel), **LAURA M. SANGALLI** (Politecnico di Milano)

[Monday May 26 / lundi 26 mai, 16:00]

Generalized Models for Spatial Regression
Modèles généralisés pour régression spatiale

We propose a novel method for the analysis of spatially distributed data occurring over irregularly shaped domains. This allows to estimate functions defined over domains characterized by concavities, complex boundaries or interior holes. We consider a generalized additive framework allowing for responses from an exponential family distribution. Specifically, we maximize a penalized log-likelihood function where the roughness penalty term involves a suitable differential operator of the spatial field over the domain of interest. The proposed models exploit advanced scientific computing techniques and specifically make use of the Finite Element Method. They provide a basis for piecewise polynomial surfaces and allow to impose boundary conditions. Since this model can deal with responses from any exponential family distribution, it is broadly applicable.

Nous proposons une méthode novatrice pour l'analyse de données spatialement distribuées sur des domaines de forme irrégulière. Elle permet l'estimations de fonctions définies sur des domaines caractérisés par des concavités, des bords complexes ou des trous à l'intérieur. Le modèle proposé permet de traiter tout type de réponse dont la distribution appartient à une famille exponentielle. On maximise une fonction de log-vraisemblance pénalisée par un opérateur différentiel défini sur le domaine considéré. Le modèle proposé fait usage de méthodologies avancées issues de l'analyse numérique et utilise la méthode des éléments finis. Elle permet de construire une base pour des fonctions polynomiales par morceaux. Le modèle proposé peut être appliqué à de nombreux types de données spatiales.

SILE TAO & IVAN MIZERA (University of Alberta)

[Monday May 26 / lundi 26 mai, 16:15]

Prediction of Ice Hockey Performance: A Field Test of Empirical Bayes Procedure

Prédiction de la performance en hockey sur glace : test de terrain d'une procédure bayésienne empirique

We examine the records for each National Hockey League player in the regular season and the playoffs (2012-2013), with the objective of predicting the number of goals and assists per unit time for each player in the playoffs, and comparing the results obtained with those generated by other existing empirical Bayes methods. The prediction method is motivated by Koenker and Mizera (2013), and provides a new approach to computing the Kiefer-Wolfowitz nonparametric maximum likelihood estimator using duality results from con-

Nous examinons la performance de chaque joueur de la Ligue nationale du hockey en saison régulière et séries éliminatoires (2012-2013) dans le but de prédire le nombre de buts et de passes par unité de temps pour chaque joueur dans les séries éliminatoires et de comparer les résultats à ceux générés par d'autres méthodes bayésiennes empiriques existantes. La méthode de prédiction est motivée par Koenker et Mizera (2013) : une nouvelle approche pour calculer l'estimateur du maximum de vraisemblance non paramétrique de Kiefer-Wolfowitz en utilisant les résultats de dualité de l'optimisation convexe. Cette méthode réduit

1E-C5: Statistical Inference and Applications 5

Inférence statistique et applications 5

vex optimization. The method reduces the computational time by "several orders of magnitude", compared to the previous EM based methods.

le temps de calcul de plusieurs ordres de grandeur par rapport aux méthodes précédentes.

MARYAM SOHRABI & MAHMOUD ZAREPOUR (University of Ottawa)

[Monday May 26 / lundi 26 mai, 16:30]

Asymptotic Theory for M-Estimates in Unstable AR(P) Processes with Infinite Variance Innovations

Théorie asymptotique pour les M-estimations dans les processus AR(P) instables avec innovations de variance infinie

We present the asymptotic distribution of M-estimators for parameters in non-stationary AR(P) processes. The innovations are assumed to be in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. In particular, when the model involves with repeated unit roots or conjugate complex unit roots, M-estimators have higher asymptotic rate of convergence compared to the least square estimators. The asymptotic results can be written as Ito stochastic integrals.

Nous présentons la distribution asymptotique des M-estimateurs pour les paramètres dans les processus AR(P) non stationnaires. Les innovations sont supposées être dans le domaine d'attraction d'une loi stable avec un index $0 < \alpha \leq 2$. En particulier, lorsque le modèle inclut des racines unitaires répétées ou des racines unitaires complexes conjuguées, les M-estimateurs présentent un taux de convergence asymptotique supérieur aux estimateurs des moindres carrés. Les résultats asymptotiques peuvent s'écrire sous la forme d'intégrales stochastiques d'Ito.

AUDREY BÉLIVEAU, RICHARD LOCKHART & CARL SCHWARZ (Simon Fraser University), **MICHAEL SCHAUB** (Swiss Ornithological Institute), **ROGER PRADEL** (Centre d'écologie fonctionnelle et évolutive)

[Monday May 26 / lundi 26 mai, 16:45]

Explicit Integrated Population Modelling: Escaping the Conventional Assumption of Independence

Modélisation explicite intégrée de populations : échapper à l'hypothèse classique d'indépendance

Integrated population modelling is a relatively new development in statistical ecology that permits the joint analysis of different sources of data, e.g. capture-recapture and count data. Typically, the joint likelihood is obtained using an independence assumption so that it can be conveniently expressed as a product of the likelihoods of the respective datasets. In this talk, I will present a new Bayesian model that avoids the independence assumption by using latent variables that keep track of the marked and unmarked individuals separately while allowing unmarked individuals to become marked when so. I will present the results of a simulation study and apply the methods to data from a colony of Greater horseshoe bats (*Rhinolophus ferrumequinum*) in the Valais, Switzerland.

La modélisation intégrée de population est un développement relativement récent en écologie statistique qui permet l'analyse conjointe à partir de différentes sources de données, e.g. échantillonnage avec capture-recapture et données de dénombrement. Typiquement, la vraisemblance conjointe est obtenue à l'aide d'une hypothèse d'indépendance de manière à ce qu'elle puisse aisément être exprimée comme le produit de la vraisemblance de chaque jeu de données. Je présenterai un nouveau modèle bayésien qui évite l'hypothèse d'indépendance en utilisant des variables latentes qui tiennent compte séparément des individus marqués et non-marqués tout en permettant aux individus non-marqués de devenir marqués. Je présenterai les résultats d'une étude par simulation et appliquerai les méthodes à des données obtenues d'une colonie de grands rhinolophes fers à cheval (*Rhinolophus ferrumequinum*) dans le Valais, en Suisse.

2A-A: Gold Medal Address
Allocution du récipiendaire de la Médaille d'or

Chair/Président: Don McLeish (University of Waterloo)
Organizer/Responsable: Christian Léger (University of Montreal)

Room/Salle: MedSci Auditorium

Abstract/Résumé

JEFFREY ROSENTHAL (University of Toronto)

[Tuesday May 27 / mardi 27 mai, 8:45]

The Magic of Monte Carlo

La magie de Monte Carlo

Markov chain Monte Carlo (MCMC) algorithms have completely revolutionised statistical computations. One reason for their success is a strong theoretical foundation, which has allowed us to validate the basic algorithms, provide numerous extensions and generalisations of the algorithms, clarify different algorithm options and tunings, and evaluate the results. This talk will present simple graphical examples, dynamically displayed using Java applets, to illustrate the workings of MCMC. It will emphasise the impact and importance of various theoretical MCMC issues including ergodicity, qualitative and quantitative convergence rates, optimal scalings, and adaptive MCMC. It will not assume any previous MCMC background, and is designed to be accessible to everyone.

Les algorithmes de Monte Carlo par chaîne de Markov (MCMC) ont complètement révolutionné les calculs statistiques. Une des raisons qui expliquent leur succès réside dans leur fondement théorique rigoureux qui nous a permis de valider les algorithmes de base, d'en établir de nombreuses extensions et généralisations, de clarifier leurs diverses options et réglages et d'en évaluer les résultats. Cet exposé présente des exemples de graphiques simples, dont l'affichage dynamique réalisé avec des applets Java illustre bien le fonctionnement des MCMC. Il souligne l'impact et l'importance de divers problèmes théoriques de MCMC, y compris l'ergodicité, les taux de convergence qualitatifs et quantitatifs, un échantillonnage optimal et une MCMC adaptative. Cet exposé est conçu pour être facilement accessible à tous et il ne nécessite aucune formation préalable sur les méthodes MCMC.

2B-I1: Actuarial Finance Finance actuarielle

Organizer and Chair / Responsable et président: Jean-Francois Renaud (Université du Québec à Montréal)

Session sponsored by the Actuarial Science Section
Séance parrainée par le Groupe de science actuarielle

Room/Salle: Hart House - South Dining

Abstracts/Résumés

RUNHUAN FENG (University of Illinois at Urbana-Champaign), **HANS VOLKMER** (University of Wisconsin-Milwaukee)
[Tuesday May 27 / mardi 27 mai, 10:20]

Spectral Methods for the Valuation of Variable Annuity Guaranteed Benefits
Méthodes spectrales pour l'évaluation des prestations garanties de rente variable

Spectral methods have been extensively exploited for the pricing of exotic options, yet unknown to the risk management literature. In the first part, we demonstrate some novel applications of spectral methods for the computation of risk measures for variable annuity guaranteed minimum maturity benefit and guaranteed minimum death benefit, which is the key to determining risk-based capital according to regulatory requirements. Our examples show that spectral methods are highly efficient and numerically more stable than conventional known methods. In the second part, we explore an application of spectral methods in conjunction with probabilistic methods in the pricing of more complicated guaranteed minimum withdrawal benefit. Numerical examples are provided to demonstrate the great efficiency and accuracy of the resulting algorithm.

Les méthodes spectrales ont été largement utilisées pour l'établissement des prix d'options exotiques et, pourtant, elles ne figurent pas dans la littérature sur la gestion des risques. En première partie, nous démontrons des applications nouvelles des méthodes spectrales pour mesurer les risques des prestations minimales de rente variable garanties à l'échéance et ceux des prestations minimales garanties au décès, ce qui constitue un facteur clé dans la détermination des fonds propres fondée sur les risques, conformément aux exigences réglementaires. Nos exemples montrent que les méthodes spectrales sont très efficaces et plus stables sur le plan quantitatif que les méthodes traditionnelles reconnues. En deuxième partie, nous explorons une application des méthodes spectrales parallèlement à des méthodes probabilistes dans l'établissement des prestations minimales garanties de sortie qui sont plus complexes. Nous présentons des exemples quantitatifs pour démontrer l'efficacité et la précision très élevées de l'algorithme qui en résulte.

ANNE MACKAY & CAROLE BERNARD (University of Waterloo)
[Tuesday May 27 / mardi 27 mai, 10:50]

Reducing Surrender Incentives through Fee Structure in Variable Annuities
Structures de frais visant à réduire l'incitatif à la résiliation pour des produits d'assurance liés aux marchés

In this presentation, we discuss the effect the fee structure of a variable annuity on the embedded surrender option. We consider fees set as a fixed percentage of the variable annuity account, as well as regular fees set as a fixed, deterministic

Dans cette présentation, nous analysons l'effet de la structure des frais de gestion sur l'option de résiliation d'un produit d'assurance lié aux marchés. Nous considérons un frais exprimé en pourcentage du fonds, un montant pré-déterminé payé périodiquement ainsi que des pénalités à la résilia-

2B-I1: Actuarial Finance Finance actuarielle

amount. Surrender charges are also taken into account. For fairly general premium and fee schedules, we discuss theoretical conditions under which it is never optimal for the policyholder to surrender. Using finite difference methods and partial differential equations, we present numerical examples that highlight the effect of a combination of surrender charges and deterministic fees in reducing the value of the surrender option and raising the optimal surrender boundary.

tion. Pour différentes structures de primes et de frais, nous présentons des conditions théoriques pour éliminer la possibilité de résiliation optimale. En utilisant des différences finies pour résoudre numériquement des équations différentielles partielles, nous présentons des exemples numériques qui démontrent l'effet de la combinaison des pénalités de résiliation et des frais pré-déterminés sur l'option de résiliation et la barrière de résiliation optimale.

TOM SALISBURY & MOSHE MILEVSKY (York University)

[Tuesday May 27 / mardi 27 mai, 11:20]

Optimal Tontines

Tontines optimales

Tontines were once more popular than annuities, but have fallen out of favour. As risk charges for systemic longevity risk rise, tontines (and other alternatives to traditional annuity designs) are being reconsidered, that leave the purchaser insured against idiosyncratic but not systemic longevity. We will look at optimizing tontine design and use.

Les tontines ont déjà été plus en demande que les rentes, mais elles ont perdu en popularité. À mesure qu'augmentent les primes de risques pour couvrir le risque de longévité systémique, les tontines (ainsi que d'autres solutions de rechange aux modèles traditionnels de rentes) sont réexaminées afin d'assurer l'acquéreur contre la longévité idiosyncratique, mais non systémique. Nous examinons les façons d'optimiser le concept et l'utilisation des tontines.

2B-I2: New Developments in High Dimensional Data Analysis Récents progrès en analyse de données de grande dimension

Organizer and Chair / Responsable et président: Abbas Khalili (McGill University)

Room/Salle: Hart House - East Common

Abstracts/Résumés

JINCH LV, YINGYING FAN & ZEMIN ZHENG (University of Southern California)

[Tuesday May 27 / mardi 27 mai, 10:20]

High-Dimensional Thresholded Regression and Shrinkage Effect

Régression seuillée de grande dimension et effet de rétrécissement

The use of nonconvex penalty functions shows advantage in selecting important features in high dimensions, but the global optimality of such methods still demands more understanding. In this paper, we consider sparse regression with hard-thresholding penalty, which we show to give rise to thresholded regression. Under some mild conditions allowing exponentially growing dimensionality, we establish oracle inequalities of the regularized estimator, as the global minimizer, under various prediction and variable selection losses, as well as oracle risk inequalities of hard-thresholded estimator followed by a further L2-regularization. The risk properties exhibit interesting shrinkage effects under both estimation and prediction losses. We identify optimal choice of the ridge parameter, which is shown to have simultaneous advantages to both L2-loss and prediction loss.

L'utilisation de fonctions de pénalité non convexes présente un avantage en matière de sélection des caractéristiques importantes de haute dimensionnalité, mais l'optimalité globale de ce type de méthode exige une compréhension plus approfondie. Dans cet article, nous examinons une régression éparsée avec une pénalité à seuil fort, qui donne lieu à une régression seuillée. Sous des conditions faibles qui permettent une dimensionnalité croissant de façon exponentielle, nous établissons les inégalités d'oracle de l'estimateur régularisé, comme le réducteur global, selon diverses fonctions de pertes pour la prédiction et la sélection de variables, ainsi que les inégalités d'oracle en matière de risque de l'estimateur à seuil fort, suivi d'une régularisation L2. Les propriétés du risque présentent un effet de rétrécissement intéressant selon les fonctions de pertes pour l'estimation et la prédiction. Nous identifions le choix optimal du paramètre «ridge», qui présente des avantages simultanés à la fois pour les fonctions de pertes L2 et de prédiction.

ANAND VIDYASHANKAR (George Mason University)

[Tuesday May 27 / mardi 27 mai, 10:50]

Finite Sample Post Model Selection Inference

Inférence post-sélection de modèle sur des échantillons finis

Simultaneous model selection and inference is one of the intensively studied areas in theoretical and applied statistics. However, most of the asymptotic and finite sample results established so far in the literature do not take into account model selection uncertainty in inference. In this talk, we describe new methodologies that address this problem. Motivated by this question, we describe new central limit theorems in infinite dimensions and concentration inequalities which rigorously justify the pro-

La sélection et l'inférence simultanées de modèles représente l'un des domaines les plus étudiés en statistique théorique et appliquée. Cependant, la plupart des résultats asymptotiques sur des échantillons finis obtenus jusqu'à présent dans la littérature ne prennent pas compte de l'incertitude de la sélection du modèle dans l'inférence. Dans cet exposé, nous présentons de nouvelles méthodologies qui tiennent compte de ce problème. Motivés par cette question, nous décrivons de nouveaux théorèmes centraux limites de dimensions infinies et aux inégalités de concentration qui

2B-I2: New Developments in High Dimensional Data Analysis Récents progrès en analyse de données de grande dimension

posed methods. In the process, we solve some challenging open problems in the area of high-dimensional finite sample inference.

justifient rigoureusement les méthodes proposées. Nous résolvons au passage quelques problèmes complexes dans le domaine de l'inférence de grande dimension sur des échantillons finis.

HUI ZOU (University of Minnesota), **JIANQING FAN** (Princeton University), **LINGZHOU XUE** (Penn State University)
[Tuesday May 27 / mardi 27 mai, 11:20]

A General Theory on The Strong Oracle Property of Folded Penalized Estimation
Théorie générale sur la propriété oracle forte de l'estimation pénalisée repliée

Folded concave penalization methods have been shown to enjoy the strong oracle property for high-dimensional sparse estimation. However, the oracle property is established only for one of the unknown local solutions. A challenging fundamental issue still remains that it is not clear whether the local optimum computed by a given optimization algorithm possesses those nice theoretical properties. To close this important theoretical gap in over a decade, we provide a unified theory to show explicitly how to obtain the oracle solution via the local linear approximation algorithm. The general theory is demonstrated by using four classical sparse estimation problems, i.e., sparse linear regression, sparse logistic regression, sparse precision matrix estimation and sparse quantile regression.

Il a été démontré que les méthodes de pénalisation concaves repliées possèdent la propriété oracle forte pour l'estimation éparsée de grande dimension. Cependant, la propriété oracle n'est établie que pour l'une des solutions locales inconnues. Une question fondamentale demeure : il n'est pas évident de déterminer si l'optimum local calculé par un algorithme d'optimisation donné possède ces jolies propriétés théoriques. Afin de combler cette importante lacune théorique datant de plus d'une décennie, nous proposons une théorie unifiée qui montre explicitement la façon d'obtenir la solution oracle par le biais de l'algorithme d'approximation linéaire local. La théorie générale est démontrée au moyen de quatre problèmes classiques d'estimation éparsée : la régression linéaire, la régression logistique, l'estimation de matrice de précision et la régression quantiles.

2B-13: New Directions in Causal Inference Nouvelles directions en inférence causale

Chair/Président: Lawrence McCandless (Simon Fraser University)

Organizers/Responsables: Lawrence McCandless (Simon Fraser University) and/et Russell Steele (McGill University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: Hart House - Debates Rm

Abstracts/Résumés

PETER AUSTIN (University of Toronto)

[Tuesday May 27 / mardi 27 mai, 10:20]

The Performance of Different Propensity Score Methods for Estimating Marginal Hazard Ratios

La performance de différentes méthodes de score de propension pour estimer les rapports de risque marginaux

The propensity score is a popular statistical method to remove the effects of observed confounding when estimating the effect of treatments, interventions and exposures when using observational data. Most research on propensity score methods has been in the context of continuous or binary outcomes. We examine the performance of four different propensity score methods (matching, weighting, stratification, and covariate adjustment) to estimate hazard ratios for survival outcomes.

Le score de propension est une méthode statistique populaire pour éliminer les effets de confusion observée lors de l'estimation de l'effet de traitements, d'interventions et d'expositions lors de l'utilisation des données d'observation. La plupart des recherches sur les méthodes de score de propension a été faite dans le cadre de résultats continus ou binaires. Nous examinons la performance de quatre méthodes de score de propension différentes (correspondant, pondération, stratification et ajustement par covariable) pour estimer les rapports de risque pour les résultats de survie.

MIREILLE SCHNITZER (Université de Montréal)

[Tuesday May 27 / mardi 27 mai, 10:50]

Nonparametric Causal Inference through Targeted Maximum Likelihood Estimation: Options for Using Machine Learning in Causal Inference

Inférence causale non paramétrique par l'estimation du maximum de vraisemblance ciblé: options pour l'utilisation de l'apprentissage machine en inférence causale

Unbiased causal estimation inevitably relies on data-specific assumptions. In addition to these 'causal assumptions', the consistency of some estimators relies on the parametric specification of a full likelihood or of nuisance models. Researchers have sought to minimize model-dependence through the usage of nonparametric methods. Targeted Maximum Likelihood Estimation (TMLE) is a statistical framework that allows for nonparametric estimation of a causal effect using flexible modelling of

L'estimation causale non biaisée s'appuie inévitablement sur des hypothèses spécifiques aux données. En plus de ces «hypothèses causales», la convergence de certains estimateurs repose sur la spécification paramétrique de modèles de vraisemblance complète ou de nuisance. Les chercheurs tentent de minimiser la dépendance au modèle au moyen de méthodes non paramétriques. L'estimation du maximum de vraisemblance ciblé (EMVC) constitue un cadre statistique qui permet l'estimation non paramétrique d'un effet causal par la modélisation flexible de modèles de nuisance,

2B-13: New Directions in Causal Inference Nouvelles directions en inférence causale

the nuisance models, opening the door for machine learning and prediction methods. I will present the fundamentals of TMLE and demonstrate how inference can be consistently improved with machine learning. I will also describe some of the challenges in causal variable selection and how machine learning methods should and should not be used.

pavant la voie aux méthodes d'apprentissage machine et de prédiction. Je présente les fondements de l'EMVC et je démontre comment l'inférence peut être considérablement améliorée grâce à l'apprentissage machine. J'aborde certains défis de la sélection de variables causales et les façons dont les méthodes d'apprentissage machine devraient et ne devraient pas être utilisées.

DAVE STEPHENS (McGill University)

[Tuesday May 27 / mardi 27 mai, 11:20]

Discussion of New Directions in Causal Inference

Discussion à propos de nouvelles orientations en inférence causale

It is 40 years since the publication of Rubin's seminal paper 'Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies', which launched modern statisticians' interest in causal inference problems. In this talk I will discuss several recent developments, including those described in the previous talks, and also new methodology for longitudinal studies, and the potential for a return of causal inference to its explicitly model-based, Bayesian roots.

Quarante ans se sont écoulés depuis la publication de l'ouvrage précurseur de Rubin intitulé «Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies», qui a soulevé l'intérêt des statisticiens modernes à l'égard des problèmes d'inférence causale. Dans cet exposé, je discuterai divers progrès récents, y compris ceux décrits dans les exposés précédents, ainsi qu'une nouvelle méthodologie pour les études longitudinales et le potentiel pour un retour de l'inférence causale vers ses racines explicitement bayésiennes, fondées sur un modèle.

2B-I4: Semi and Unsupervised Learning Apprentissage semi- et non supervisé

Organizer and Chair / Responsable et président: Paul McNicholas (University of Guelph)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

SALVATORE INGRASSIA (University of Catania, Italy), **ANTONIO PUNZO** (University of Catania)

[Tuesday May 27 / mardi 27 mai, 10:20]

Diagnostic Tools for Data Classification via Mixture Models with Covariates

Outils de diagnostic pour la classification de données à l'aide de modèles de mélanges avec covariables

Mixture models with covariates - which include mixtures of regression, mixtures of regression with concomitant and cluster-weighted models - are flexible statistical methods for clustering heterogeneous populations based on within-group relationships between a response variable and a set of covariates. In this paper we introduce some diagnostic indices and graphical tools for model evaluation and selection. Case studies based on both simulated and real data are presented.

Les modèles de mélanges avec covariables, qui comprennent des mélanges de régression, des mélanges de régression avec variables concomitantes et des modèles pondérés par grappes, sont des méthodes statistiques flexibles pour le regroupement de populations hétérogènes basées sur les relations à l'intérieur des groupes entre une variable réponse et un ensemble de covariables. Dans cet article, nous présentons quelques indices diagnostiques et outils graphiques pour l'évaluation et la sélection d'un modèle. Nous présentons des études de cas fondées sur des données simulées et réelles.

GEOFF MCLACHLAN & SHARON LEE (University of Queensland, Australia)

[Tuesday May 27 / mardi 27 mai, 10:50]

On Finite Mixtures of Skew Distributions

À propos de mélanges finis de distributions asymétriques

Finite mixtures of multivariate skew symmetric distributions, in particular, the skew normal and skew t -mixture models, are emerging as promising extensions to traditional normal and t -mixture modelling. In this talk, we give a brief overview of various existing proposals for multivariate skew distributions. We compare the relative performance of so-called restricted and unrestricted skew mixture models in clustering and density estimation on some real datasets. We also compare their performance with mixtures having other non-normal component distributions.

Les mélanges finis de distributions asymétriques multivariées, en particulier les mélanges de lois asymétriques normales et de Student, se démarquent comme des généralisations prometteuses de la modélisation classique de lois normales et de Student. Dans cet exposé, nous présentons un aperçu sommaire de diverses propositions existantes en matière de distributions asymétriques multivariées. Nous comparons la performance relative des modèles de mélanges asymétriques avec et sans contraintes pour le regroupement et l'estimation de la densité à l'aide de quelques jeux de données réelles. Nous comparons aussi leur performance avec celle de mélanges présentant d'autres distributions de composantes non normales.

2B-I4: Semi and Unsupervised Learning Apprentissage semi- et non supervisé

IRENE VRBIK & PAUL MCNICHOLAS (University of Guelph)

[Tuesday May 27 / mardi 27 mai, 11:20]

Fractionally-Supervised Classification via Skew-Normal and Skew-t Mixture Models

Classification partiellement supervisée au moyen de modèles de mélanges de lois asymétriques normales et de Student

Traditionally, model-based classification can be supervised, semi-supervised, or unsupervised. Recently, a fractionally-supervised classification (FSC) approach was introduced to allow for any level of supervision--ranging from unsupervised to supervised--to be attained. This novel approach, which uses a weighted likelihood in a Gaussian framework, can be extended to include a more robust class of component densities. To increase the flexibility of these models, the FSC approach is explored using skew-normal and skew-t mixture models. The performance of these models are compared with the traditional classification techniques and symmetric counterparts.

Traditionnellement, la classification basée sur un modèle peut être supervisée, semi-supervisée ou non supervisée. Récemment, une méthode de classification partiellement supervisée (CPS) a été développée afin de permettre l'atteinte de tout niveau de supervision. Cette nouvelle approche, qui fait appel à une vraisemblance pondérée dans un contexte gaussien, peut être élargie pour inclure une classe plus robuste de densités des composantes. En vue d'augmenter la flexibilité de ces modèles, nous examinons la méthode de CPS en utilisant des modèles de mélanges de lois asymétriques normales et de Student. Nous comparons la performance de ces modèles à l'aide des techniques de classification traditionnelles et des équivalents symétriques.

2B-15: Shape-Constrained Maximum Likelihood: Methods and Applications Maximum de vraisemblance avec contraintes de forme : méthodes et applications

Organizer and Chair / Responsable et présidente: Hanna Jankowski (York University)

Room/Salle: UC - 140

Abstracts/Résumés

CHARLES DOSS (University of Minnesota), **FADOUA BALABDAOUI** (Université Paris Dauphine), **JON WELLNER** (University of Washington)

[Tuesday May 27 / mardi 27 mai, 10:20]

Semiparametric Modeling with Log-Concave Densities

Modélisation semiparamétrique avec des densités log-concaves

We discuss uses of log-concave density estimates. One reason log-concavity has become popular is that log-concave maximum likelihood estimates (MLEs) exist without requiring a bandwidth choice, unlike, e.g., kernel density estimates. The log-concave MLE also behaves well when the model is misspecified, i.e., when log-concavity does not actually hold. The family of log-concave densities serves as a generalization of many familiar parametric families, in particular, the Gaussian one. Thus, many parametric models can be generalized to semiparametric ones by replacing a parametric assumption by the assumption of log-concavity. We discuss some such uses of log-concavity, with specific focus on the use of symmetric log-concave densities in mixture modeling.

Nous discutons l'utilisation des estimations de densités log-concaves. L'une des raisons qui expliquent la popularité de la log-concavité est que les estimateurs au maximum de vraisemblance log-concaves (EMV) existent sans nécessiter un choix de longueur de bande, contrairement aux estimations à noyau de densité, par exemple. L'EMV log-concave se comporte bien également lorsque le modèle est mal spécifié, c'est-à-dire, lorsque la log-concavité n'est pas vérifiée. La famille des densités log-concaves sert à la généralisation de plusieurs familles paramétriques familières, notamment celle de type gaussien. Ainsi, bon nombre des modèles paramétriques peuvent être généralisés en modèles semiparamétriques en remplaçant une hypothèse paramétrique par une hypothèse de log-concavité. Nous discutons des exemples d'utilisation de log-concavité, en mettant un accent particulier sur l'utilisation de densités log-concaves symétriques dans une modélisation de mélanges.

BODHI SEN (Columbia University), **MARY MEYER** (Colorado State University)

[Tuesday May 27 / mardi 27 mai, 10:50]

Testing Against a Parametric Regression Function

Test pour l'existence d'une fonction paramétrique de régression

A formal likelihood ratio hypothesis test for the validity of a parametric regression function is proposed, using a large-dimensional, nonparametric double cone alternative. For example, the test against a constant function uses the alternative of increasing or decreasing regression functions, and the test against a linear function uses the convex or concave alternative. The proposed test is exact, unbiased and the critical value is easily computed. The power of the test increases to one as the sam-

Nous proposons un test d'hypothèse formel basé sur le rapport de vraisemblance pour la validité d'une fonction de régression paramétrique, en utilisant une contre-hypothèse à cônes doubles non paramétrique de grande dimension. Par exemple, la recherche d'une fonction constante fait appel à une contre-hypothèse d'augmentation ou de diminution des fonctions de régression, et la recherche d'une fonction linéaire utilise la contre-hypothèse convexe ou concave. Le test proposé est précis et non biaisé, et la valeur critique est facilement calculée. La puissance du test s'élève

2B-15: Shape-Constrained Maximum Likelihood: Methods and Applications

Maximum de vraisemblance avec contraintes de forme : méthodes et applications

ple size increases, under very mild assumptions -- even when the alternative is misspecified. Simulations show that our procedure behaves well consistently when compared with other methods. Although the alternative fit is non-parametric, no tuning parameters are involved.

à un, à mesure qu'augmente la taille de l'échantillon, sous des hypothèses non restrictives, et ce, même lorsque la contre-hypothèse est mal spécifiée. Des simulations montrent que la bonne performance de notre procédure est constante en comparaison avec d'autres méthodes. Même si l'ajustement de la contre-hypothèse est non paramétrique, aucun paramètre de réglage n'est nécessaire.

MING YUAN (University of Wisconsin-Madison)

[Tuesday May 27 / mardi 27 mai, 11:20]

Information-Theoretic Properties of Log-Concave Projections and Their Applications

Propriétés des projections log-concaves du point de vue de la théorie de l'information et de leurs applications

In this talk, I will describe a set of information-theoretic properties of log-concave projections of a probability measure. I shall also present a couple statistical applications of these results.

Dans le présent exposé, je décris un ensemble de propriétés des projections log-concaves d'une mesure de probabilité du point de vue de la théorie de l'information. Je présente aussi des applications statistiques de ces résultats.

2B-16: Stochastic Processes and Their Applications Processus stochastiques et leurs applications

Organizer and Chair / Responsable et présidente: Barbara Szyszkowicz (Carleton University)

Session sponsored by the Probability Section
Séance parrainée par le Groupe de probabilité

Room/Salle: UC - 161

Abstracts/Résumés

MIKLOS CSORGO & M. MASOUD NASARI (Carleton University)

[Tuesday May 27 / mardi 27 mai, 10:20]

Randomized Empirical Processes with Applications to Infinite Super-Populations and Big Data Sets

Processus empiriques randomisés et applications à des superpopulations infinies et à des jeux de données volumineux

A super-population outlook regards a finite population as a large imaginary random sample of N labeled units $\{X_1, \dots, X_N\}$ of real valued random variables from a hypothetical infinite super-population. One may also view a Big Data set of univariate observations as if it were a large imaginary random sample of N labeled units. In both of these scenarios, instead of trying to process the entire data set, we sample it via its index set $\{1, \dots, N\}$ and thus reduce the problem to dealing with significantly smaller sub-samples that, in turn, we study with the help of appropriately randomized empirical processes.

Dans le contexte d'une superpopulation, une population finie est considérée comme un échantillon aléatoire imaginaire de grande taille de N unités identifiées $\{X_1, \dots, X_N\}$ de variables aléatoires de données réelles tirées d'une hypothétique population infinie. Un ensemble volumineux de données univariées peut aussi être considéré comme un échantillon aléatoire imaginaire de grande taille de N unités identifiées. Dans le cadre de ces deux scénarios, plutôt que de traiter le jeu de données entier, nous l'échantillonnons au moyen de son ensemble d'indices $\{1, \dots, N\}$, ce qui réduit le problème à créer des sous-échantillons de taille significativement plus petite que nous étudions ensuite à l'aide de processus empiriques randomisés de façon appropriée.

DON DAWSON (Carleton University)

[Tuesday May 27 / mardi 27 mai, 10:50]

Long Time Behaviour of a Class of Multidimensional Diffusions

Comportement à long terme d'une classe de processus de diffusion multidimensionnels

We consider some classes of Markov diffusion processes with values in the positive quadrant and related systems of interacting diffusions on the hierarchical group and their hierarchical mean field and continuum limits. In particular we consider the behaviour of catalytic branching and mutually catalytic branching systems. A review of recent results and open problems will be given.

Nous étudions quelques classes de processus de diffusion markoviens présentant des valeurs dans le cadran positif et des systèmes connexes de diffusions interactives au sein du groupe hiérarchique, du domaine hiérarchique moyen et des limites du continuum. En particulier, nous examinons le comportement d'embranchements catalytiques et de systèmes d'embranchements mutuellement catalytiques. Nous présentons un aperçu des résultats récents et des problèmes en suspens.

2B-I6: Stochastic Processes and Their Applications Processus stochastiques et leurs applications

DON MCLEISH (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 11:20]

The Importance of Importance Sampling: Measure Change in Probability, Statistics and Rare Event Simulation
Importance de l'échantillonnage par importance: changement de mesure en probabilité, statistique et simulation d'événements rares

Arguably nothing has contributed more to the practice of statistics than measure change and likelihood ratios. These are fundamental objects throughout statistics, probability and finance and provide the underpinning of much of statistical estimation and testing, and much of modern finance theory. We provide a brief review of many uses of Radon-Nikodym derivatives or likelihood ratios, together with some practical cautionary notes on their use. We will show examples of the extraordinary power of importance sampling, especially for simulation of stochastic processes and for rare event simulation, including the use of the generalized extreme value distributions for rare event simulation.

Probablement qu'aucune contribution n'a apporté plus à la pratique de la statistique que le changement de mesure et les rapports de vraisemblance. Ces techniques sont fondamentales en statistique, en probabilité et en finance, et elles constituent la base de la plupart des estimations et des sondages, ainsi qu'une bonne partie de la théorie moderne en matière de finance. Nous présentons un bref aperçu d'une foule d'usages des instruments dérivés de Radon-Nikodym ou des rapports de vraisemblance, ainsi que quelques mises en garde pratiques sur leur utilisation. Nous proposons des exemples qui montrent l'extraordinaire puissance de l'échantillonnage par importance, en particulier en matière de simulation de processus stochastiques et de simulation d'événements rares, qui comprend l'utilisation de distributions généralisées de valeurs extrêmes.

2B-C1: Biostatistics - Measurement Errors and Prediction Errors Biostatistique - erreurs de mesure et erreurs de prédiction

Chair/Présidente: Elif Acar (University of Manitoba)

Room/Salle: UC - 179

Abstracts/Résumés

LONGHAI LI, SHI QIU, BEI ZHANG & CINDY X. FENG (University of Saskatchewan)

[Tuesday May 27 / mardi 27 mai, 10:20]

Approximating Cross-validators Predictive Evaluation in Bayesian Latent Variables Models with Integrated IS and WAIC
Approximation de l'évaluation prédictive par validation croisée dans des modèles bayésiens à variables latentes à l'aide des critères IS et WAIC intégrés

A natural method for approximating out-of-sample predictive evaluation is leave-one-out cross-validation (LOOCV). However, actual LOOCV is time-consuming. This paper introduces two methods, namely iIS and iWAIC, for approximating LOOCV with only Markov chain samples simulated from a posterior based on a *full* data set. In iIS and iWAIC, we first integrate the predictive density over the distribution of the latent variables associated with the held-out without reference to its observation, then apply IS and WAIC approximations to the integrated predictive density. We compare iIS and iWAIC with other approximation methods in three real data examples. Our empirical results show that iIS and iWAIC give substantially better approximates than non-integrated IS and WAIC and other methods.

Une méthode naturelle pour approximer une évaluation prédictive hors échantillon est la validation croisée avec omission d'une variable (LOOCV). Toutefois, la véritable méthode LOOCV prend un temps considérable. Cet exposé présente deux méthodes, notamment iIS et iWAIC, pour effectuer une approximation LOOCV sur des chaînes de Markov simulées à partir d'une loi a posteriori basé sur un jeu de données *complet*. Dans les méthodes iIS et iWAIC, nous intégrons d'abord la densité prédictive à la distribution des variables latentes associées à la variable omise sans référence à son observation, puis nous appliquons les méthodes d'approximation iIS et iWAIC à la densité prédictive intégrée. Nous comparons iIS et iWAIC à d'autres méthodes d'approximation dans trois exemples de données réelles. Nos résultats empiriques montrent que iIS et iWAIC produisent des approximations sensiblement meilleures que les autres méthodes.

SHOMOITA ALAM & YAN YUAN (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 10:35]

Average Precision – An Overall Measure for Evaluating Classification Prediction Performance
Précision moyenne : mesure globale pour l'évaluation de la performance prévisionnelle de classification

To evaluate medical tests for disease classification, the area under the Receiver-Operating Characteristics Curve (AUC) has been widely used. Similar to AUC, Average precision (AP) is an overall measure of the predictive performance of a classifier. Previous work has shown that AP is related to AUC and is a better measure than AUC if prediction of low prevalence class is of interest, e.g., the medical screening test. The asymptotic properties of AP have been developed for categorical test scores. In

Pour évaluer des tests médicaux en vue d'une classification de maladies, l'aire sous la courbe ROC (AUC) a été largement utilisée. De façon similaire, la précision moyenne (AP) est une mesure globale de la performance prévisionnelle d'un classificateur. Des travaux antérieurs ont démontré que l'AP est liée à l'AUC et qu'elle constitue une meilleure mesure que l'AUC lorsque la prévision d'une classe de faible prévalence présente un intérêt, par exemple pour un test de dépistage médical. Les propriétés asymptotiques de l'AP ont été établies pour des scores de test catégoriel. Dans cet ex-

2B-C1: Biostatistics - Measurement Errors and Prediction Errors Biostatistique - erreurs de mesure et erreurs de prédiction

this talk, we extend AP to deal with continuous test scores, and provide simulation results and demonstrate its use with a real data set.

posé, nous généralisons l'AP pour traiter des scores de test continu, nous présentons des résultats de simulation et nous en démontrons l'utilisation avec un ensemble de données réelles.

SHAHADUT HOSSAIN (UAE University)

[Tuesday May 27 / mardi 27 mai, 10:50]

Covariate Mismeasurement in Logistic Regression: Bayesian Adjustment of Main and Interaction Effects and the Sample Size Implications

Erreur de mesure dans les covariables en régression logistique : ajustement bayésien des effets principaux et des effets d'interaction, et implications pour la taille d'échantillon

Mismeasurement in both quantitative and categorical covariates are common in epidemiological studies. Regression analysis ignoring such mismeasurements seriously biases the estimated main and interaction effects of the covariates on the outcome of interest. In this research, we propose a Bayesian parametric framework for eliminating deleterious impacts of covariate mismeasurements in logistic regression. Furthermore, adjustment for covariate mismeasurements requires validation data, usually in the form of replicates of the mismeasured covariates on a subset of the study population. Adequacy of the proposed adjustment method depends on the sizes of the main and the validation samples. Thus, we also provide a general guideline about the required sizes of the main and the validation samples based on simulation studies.

Les erreurs de mesure dans les covariables quantitatives et catégorielles sont courantes dans les études épidémiologiques. L'analyse de régression qui ignore de telles erreurs de mesure biaise considérablement l'estimation des effets principaux et des effets d'interaction des covariables sur les résultats d'intérêt. Dans cette recherche, nous proposons un cadre paramétrique bayésien pour l'élimination de l'impact néfaste des erreurs de mesure des covariables dans la régression logistique. De plus, un ajustement pour des erreurs de mesure des covariables nécessite des données de validation, normalement sous forme de répliqués des covariables mesurées erronément dans un sous-ensemble de la population à l'étude. Une adéquation de la méthode d'ajustement proposée dépend des tailles de l'échantillon principal et de l'échantillon de validation. Nous fournissons également des directives générales à propos des tailles requises pour l'échantillon principal et l'échantillon de validation en fonction d'études de simulation.

SHOFIQL ISLAM, JEMILA HAMID, LEHANA THABANE & JOSEPH BEYENE (McMaster University)

[Tuesday May 27 / mardi 27 mai, 11:05]

A Copula-based Method to Classify Individuals into Different Disease Categories Using Correlated Biomarkers

Méthode basée sur les copules pour classifier les patients dans différentes catégories de maladies à l'aide de biomarqueurs corrélés

We developed a new method of classifying patients into disease categories. Our method is flexible and allows classification based on broad classes of bivariate distributions that takes correlation into account and leads to a unique classification. We consider Frank, Clayton and Gumbel's copula with Gamma marginal's to construct joint probability distribution for this purpose. Comparing the association parameter in copula with Kendall's Tau, we observe that a given value of the parameter represent

Nous avons élaboré une nouvelle méthode de classification des patients par catégories de maladies. Cette méthode flexible conduit à une classification unique tout en permettant une classification en fonction de deux grandes classes de distributions bivariées qui tiennent compte de la corrélation. Aux fins de l'étude, nous considérons les copules de Frank, Clayton et Gumbel avec des distributions marginales gamma pour produire une distribution de probabilité conjointe. En comparant le paramètre d'association de la copule avec le tau de Kendal, nous observons qu'une certaine

2B-C1: Biostatistics - Measurement Errors and Prediction Errors Biostatistique - erreurs de mesure et erreurs de prédiction

higher levels of association in Gumbel followed by Clayton and Frank. Optimal threshold for disease classification converges to a stationary distribution regardless of the choices of copula. Furthermore, the simulation result indicates that Frank's copula perform better than Clayton followed by Gumbel in terms of predictive ability.

valeur du paramètre représente des niveaux d'association supérieurs avec Gumbel, suivi par Clayton, puis par Frank. Le seuil optimal de classification par maladie converge vers une distribution stationnaire, quel que soit le choix de copule. De plus, les résultats de simulation montrent que, d'un point de vue prévisionnel, la performance de la copule de Frank est supérieure à celle de Clayton, qui est elle-même supérieure à celle de Gumbel.

DAVID SOAVE & LEI SUN (University of Toronto)

[Tuesday May 27 / mardi 27 mai, 11:20]

Improved Detection of Variants with Main or Interaction Effects using a Robust Location and Scale Testing Framework
Comment améliorer la détection de variantes avec effets principaux et d'interaction à l'aide d'un cadre de vérification de l'emplacement et de l'échelle robuste

The common approach to identifying genetic variants associated with a quantitative phenotype is to test for phenotypic differences in mean (location) across the three genotypic groups. This approach ignores potential variance (scale) differences due to (e.g.) interactions with other risk factors. We propose a joint location-scale testing framework that includes a full likelihood approach and robust methods that combine evidence from the individual location and scale test classes. Efficiency and robustness of the proposed approach are studied via extensive simulation studies and multiple data applications. Complexities of the framework are investigated including effects of phenotypic distribution and frequency of the minor allele on the accuracy of the tests. A generalization of the framework to incorporate genotype group uncertainty is discussed.

L'approche la plus commune pour identifier les variantes génétiques associées à un phénotype quantitatif consiste à tester les différences phénotypiques en moyenne (position) sur les trois groupes génotypiques. Mais cette approche ignore les éventuelles différences de variance (d'échelle) dues, entre autres, aux interactions avec d'autres facteurs de risque. Nous proposons un cadre de test position/échelle conjoint qui inclut une approche de vraisemblance complète et des méthodes robustes qui combinent des données des classes de test concernant les emplacements et échelles individuels. Nous étudions l'efficacité et la robustesse de l'approche proposée via de nombreuses études par simulation et de multiples applications concrètes. Nous explorons les complexités du cadre de travail, notamment les effets de la distribution phénotypique et de la fréquence de l'allèle mineur sur la précision des tests. Nous discutons d'une généralisation du cadre de travail pour y inclure l'incertitude concernant les groupes de génotypes.

YING YAN & GRACE YI (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 11:35]

Model Misspecification and Model Checking for Survival Data with Covariate Measurement Error
Erreur de spécification du modèle et vérification du modèle pour les données de survie avec erreur de mesure des covariables

Survival models with covariate measurement error have been studied extensively, and most methods that correct for measurement error effects are developed by assuming a correct error model for the measurement error process. It is not clear what the impact would be if the assumption of the survival model and/or error model are incorrect. In this

Les modèles de survie avec erreur de mesure des covariables ont fait l'objet de nombreuses études et la plupart des méthodes qui corrigent les effets de cette erreur de mesure partent de l'hypothèse d'un modèle d'erreur exact pour le processus d'erreur de mesure. Mais il est difficile de savoir quel serait l'impact si les hypothèses du modèle de survie et/ou du modèle d'erreur sont inexactes. Dans cette présen-

2B-C1: Biostatistics - Measurement Errors and Prediction Errors Biostatistique - erreurs de mesure et erreurs de prédiction

talk, we study the impact of model misspecification on parameter estimation, and propose an omnibus goodness of fit test to detect departure from the survival model or the error model assumptions. Simulation studies and a real data example are presented.

tation, nous étudions l'impact de l'erreur de spécification du modèle sur l'estimation des paramètres et nous proposons un test de qualité de l'ajustement global pour détecter un écart des hypothèses du modèle de survie et/ou du modèle d'erreur. Nous présentons des études par simulation et un exemple sur des données réelles.

2B-C2: Empirical and Composite Likelihood and Higher Order Asymptotics Vraisemblances empirique et composite et comportement asymptotique d'ordre supérieur

Chair/Présidente: Cindy Fu (York University)

Room/Salle: UC - A101

Abstracts/Résumés

SHIVA ASHTA & NANCY REID (University of Toronto)

[Tuesday May 27 / mardi 27 mai, 10:20]

Higher Order Asymptotics and Normal Coefficients of Variation

Propriétés asymptotiques d'ordre supérieur et coefficients de variation normaux

Krishnamoorthy and Lee (2013) compared three methods to study coefficients of variation of independent normal populations. A mean- and variance-corrected version of the signed likelihood root statistic and an approach using generalized pivotal quantities were compared, and shown to give very accurate inference. We propose the use of the r-star approximation (Barndorff-Nielsen 1991) constructed using the method of Fraser and Reid (1993) as an alternative method to compare normal coefficients of variation. Confidence intervals for the difference and ratios of two coefficients of variation are also developed. Finally, the relationship between the r-star approximation and generalized pivotal quantity approaches are studied numerically and analytically.

Krishnamoorthy et Lee (2013) ont comparé trois méthodes pour étudier les coefficients de variation de populations normales indépendantes. Une version corrigée pour les moyennes et les variances de la racine signée de vraisemblance, ainsi qu'une approche utilisant des quantités pivotaux généralisées ont été comparées, et elles ont produit une inférence très précise. Nous proposons l'utilisation d'une approximation R étoile (Barndorff-Nielsen, 1991) construite à l'aide de la méthode de Fraser et Reid (1993) comme une méthode de rechange pour comparer des coefficients de variation normaux. Nous établissons des intervalles de confiance pour la différence ainsi que des ratios de deux coefficients de variation. Enfin, nous examinons la relation entre l'approximation R étoile et l'approche des quantités pivotaux sur les plans quantitatif et analytique.

SONG CAI & JIAHUA CHEN (The University of British Columbia)

[Tuesday May 27 / mardi 27 mai, 10:35]

Empirical Likelihood Inference under Density Ratio Models Based on Type I Censored Multiple Samples

Inférence de la vraisemblance empirique sous les modèles de rapport de densité sur des échantillons multiples censurés de type I

We present an empirical likelihood (EL) inference framework for estimating quantiles of a given number of different population distributions, with Type I independent samples from each. This framework is based on the semi-parametric density ratio model (DRM), which gains efficiency by pooling the common information in multiple samples. A fast computational method for the maximum EL estimator of the DRM parameter is proposed, and the asymptotic property of that estimator is studied. We also propose an estimator for quantiles of dif-

Nous présentons un cadre d'inférence de la vraisemblance empirique (VE) pour estimer les quantiles d'un nombre donné de distributions de populations, avec des échantillons de type I pour chacune d'elles. Ce cadre est fondé sur le modèle de rapport de densité (MRD) semiparamétrique, qui gagne en efficacité par la combinaison d'informations communes à de nombreux échantillons. Nous proposons un modèle computationnel pour l'estimateur du maximum de VE du paramètre du MRD et étudions la propriété asymptotique de cet estimateur. Nous proposons également un estimateur pour les quantiles de différentes distributions fondé

2B-C2: Empirical and Composite Likelihood and Higher Order Asymptotics Vraisemblances empirique et composite et comportement asymptotique d'ordre supérieur

ferent distributions based on the DRM and show it to be asymptotically normal and more efficient than the sample quantiles, which do not borrow strength across samples. The proposed inference framework is also useful for hypothesis testing, density estimation, etc.

sur le MRD et montrons qu'il est asymptotiquement normal et plus efficace que les quantiles d'échantillon, qui ne partagent pas leur puissance entre échantillons. Le cadre d'inférence proposé est également utile pour la vérification d'hypothèses, l'estimation de la densité, etc.

MAHDIS AZADBAKHS, XIN GAO & HANNA JANKOWSKI (York University)

[Tuesday May 27 / mardi 27 mai, 10:50]

Multiple Comparisons using Composite Likelihood in Clustered Data

Comparaisons multiples par vraisemblance composite pour des données en grappes

Multiple comparison is a common problem in many applications. Existing multiple comparison procedures often use maximum likelihood estimation which is computationally intensive for multivariate and correlated data sets. To overcome this computational barrier, composite likelihood method is introduced to reduce the dimensionality and simplify the parameter estimation.

We propose to use composite likelihood in estimation and construct procedures for multiple comparisons on regression coefficients. We focus on multivariate clustered data with within cluster correlation from three different models: Multivariate Normal; Multivariate Probit and Quadratic Exponential model. Simulations are performed to assess the empirical performance of this procedure. A real data set is used to demonstrate the utility of the proposed multiple comparison procedure.

Les comparaisons multiples constituent un problème courant dans bon nombre d'applications. Les procédés de comparaisons multiples existants utilisent souvent un estimateur au maximum de vraisemblance, ce qui nécessite des calculs intensifs pour des ensembles de données multivariées et corrélées. Pour surmonter cette barrière informatique, une méthode de vraisemblance composite est présentée pour réduire la dimension et simplifier l'estimation de paramètres. Nous proposons l'utilisation de la vraisemblance composite dans les procédures d'estimation et de construction aux fins de comparaisons multiples des coefficients de régression. Nous mettons l'accent sur des données en grappes multivariées avec une corrélation intra-grappes issue de trois modèles différents : le modèle normal multivarié, le modèle probit multivarié et le modèle exponentiel quadratique. Des simulations sont effectuées pour évaluer la performance empirique de cette procédure. Un ensemble de données réelles est utilisé pour démontrer l'utilité de la procédure de comparaison multiple proposée.

MIN CHEN, CHANGBAO WU & MARY THOMPSON (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 11:05]

Empirical Likelihood Methods for Testing the Difference of Two Distributions in Pretest-Posttest Studies

Méthodes de vraisemblance empiriques afin de vérifier la différence entre deux distributions dans des études pré-test et post-test

In pretest-posttest studies, it is sometimes of interest to test the difference of distributions from the two treatment groups. Empirical likelihood (EL) has been found useful in pretest-posttest studies but the EL test on the difference of two distributions involves a U-statistic in the constraint. We first explore the use of the jackknife empirical likelihood (JEL), which is shown to be effective in dealing with two-sample U-statistics under independent data.

Lors d'études pré-test et post-test, il est parfois intéressant de tester la différence entre les deux distributions. La vraisemblance empirique (VE) s'est avérée utile dans les études pré-test et post-test mais le test de VE sur la différence de deux distributions implique un coefficient U dans la contrainte. Nous explorons d'abord l'utilisation de la méthode d'Eustache de vraisemblance empirique (MEVE), qui s'avère efficace lorsque l'on a des coefficients U pour deux échantillons de données indépendantes. Nous proposons

2B-C2: Empirical and Composite Likelihood and Higher Order Asymptotics Vraisemblances empirique et composite et comportement asymptotique d'ordre supérieur

We then propose an adjusted JEL method to handle the case of dependent data when the pretest information is incorporated. We also consider an alternative formulation of the Mann-Whitney test using the EL approach proposed by Huang, Qin and Follmann (2008). Performances of these tests are examined through simulation studies.

donc une méthode ajustée de MEVE pour les cas de données dépendantes lorsque les informations de pré-test sont incluses. Nous considérons aussi une autre formulation du test Mann-Whitney à l'aide de l'approche de VE proposée par Huang, Qin et Follmann (2008). L'efficacité de ces tests sera évaluée à l'aide d'études de simulation

THARSHANNA NADARAJAH, ASOKAN MULAYATH VARIYATH & J CONCEPCIÓN LOREDO-OSTI (Memorial University of Newfoundland)

[Tuesday May 27 / mardi 27 mai, 11:20]

Empirical Likelihood Based Longitudinal Studies
Études longitudinales basées sur la vraisemblance

In longitudinal data analysis, our interest is on the regression parameters for the marginal expectations of the longitudinal responses. The estimating equations approaches have received much attention in longitudinal regression setup. The confidence regions and test of hypothesis are constructed based on asymptotic normality.

These methods are at the risk of model misspecification. To overcome this problem, we propose an empirical likelihood modelling based on a set of estimating equations. Our simulations studies shown that proposed method has comparable performance when a parametric model is plausible where as it outperforms when the parametric model is miss-specified.

Dans l'analyse de données longitudinales, nous étudions les paramètres de régression pour l'espérance marginale des réponses longitudinales. Les approches d'estimation d'équations ont récemment suscité une attention soutenue dans le contexte d'une régression longitudinale. Les régions de confiance et les tests d'hypothèses sont structurés en fonction de la normalité asymptotique.

Ces méthodes peuvent entraîner de mauvaises spécifications de modèles. Pour surmonter ce problème, nous proposons une modélisation par la vraisemblance empirique basée sur un jeu d'équations d'estimation. Nos simulations ont révélé que la performance de la méthode proposée est comparable pour un modèle paramétrique plausible, tandis qu'elle est supérieure pour un modèle paramétrique qui est mal spécifié.

XIMING XU & HAL WHITEHEAD (Dalhousie University)

[Tuesday May 27 / mardi 27 mai, 11:35]

A Composite Likelihood Approach to Whale's Network Analysis
Approche de vraisemblance composée pour l'analyse des réseaux de baleines

Statistical models have been developed to model the social structures of animals. Due to the complex dependence among animals, the ordinal likelihood-based methods become computationally infeasible. Here we introduce a composite likelihood-based approach to fitting the data on the movement and association of identified whales. In particular, the variance estimation and model selection criteria are developed based on the proposed composite likelihood which enables us to compare different models.

Divers modèles statistiques ont été développés pour modéliser les structures sociales des animaux. En raison de la dépendance complexe entre animaux, les méthodes fondées sur la vraisemblance ordinaire sont impossibles à gérer d'un point de vue computationnel. Nous introduisons une approche fondée sur la vraisemblance composée pour ajuster les données au mouvement et à l'association des baleines identifiées. En particulier, l'estimation de la variance et les critères de sélection du modèle sont développés en fonction de la vraisemblance composée proposée, ce qui nous permet de comparer différents modèles.

2B-C3: Nonregular Statistical Test Problems Problèmes de tests statistiques non réguliers

Chair/Président: Farouk Nathoo (University of Victoria)

Room/Salle: UC - 244

Abstracts/Résumés

JOEL SANGO & DUCHESNE PIERRE (Université de Montréal)

[Tuesday May 27 / mardi 27 mai, 10:20]

Test for Non-Correlation Under Weak Dependence Between Two Stationary Processes

Test de non-corrélation sous faible dépendance entre deux processus stationnaires

Tests for non correlation have been well studied in the literature of time series analysis. However, most of the proposed tests are based on linear models and their asymptotic distributions are obtained under the hypothesis of independent and identically distributed errors, nullity of the fourth order cross cumulants or by assuming constant conditional moments for first through fourth order of the

innovations.

In this paper, we derive under weak dependence a consistent estimator of the variance-covariance matrix of the vector of cross correlation for finite number of lags; alongside with the asymptotic distribution of the portemanteau test statistic for non correlation. The performance of the proposed test is examined through a simulation study.

Tester la non-corrélation entre deux séries fait l'objet de nombreuses études dans la littérature économétrique et statistique. Cependant, les tests proposés jusqu'ici se basent sur des modèles linéaires et sont construits sous les hypothèses d'indépendance des erreurs, de nullité des cumulants croisés d'ordre quatre, ou en supposant que les moments conditionnels d'ordre un à quatre des erreurs sont constants.

Dans le présent papier, nous dérivons sous l'hypothèse de faible dépendance un estimateur convergent de la matrice

de variance-covariance du vecteur des auto-covariances croisées pour un nombre fini de retards; de même que la distribution asymptotique de la statistique portemanteau pour la non-corrélation. Des simulations sont effectuées pour étudier la performance du test.

LUAI AL LABADI & EMAD MASUADI (UAE University), **MAHMOUD ZAREPOUR** (University of Ottawa)

[Tuesday May 27 / mardi 27 mai, 10:35]

Two-Sample Bayesian Nonparametric Goodness of Fit Test

Test d'adéquation non paramétrique bayésien pour deux échantillons

Testing the difference between two data samples is of a particular interest in statistics. Precisely, given two samples $\mathbf{X} = X_1, \dots, X_{m_1} \stackrel{i.i.d.}{\sim} F$ and $\mathbf{Y} = Y_1, \dots, Y_{m_2} \stackrel{i.i.d.}{\sim} G$, with F and G unknown continuous cumulative distribution functions, we wish to test the null hypothesis $\mathcal{H}_0 : F = G$. In this talk, we propose an effective and a convenient Bayesian nonparametric approach to assess the equality of unknown distributions. The method is based on the Kolmogorov distance and approximate samples from the Dirichlet process. Our results show

Il est particulièrement intéressant en statistiques de tester la différence entre deux échantillons de données. Plus précisément, soit deux échantillons $\mathbf{X} = X_1, \dots, X_{m_1} \stackrel{i.i.d.}{\sim} F$ et $\mathbf{Y} = Y_1, \dots, Y_{m_2} \stackrel{i.i.d.}{\sim} G$, avec F et G des fonctions de répartition continues, et nous voulons tester l'hypothèse nulle $\mathcal{H}_0 : F = G$. Dans cet exposé, nous proposons une approche bayésienne non paramétrique pratique et efficace pour évaluer l'égalité de distributions inconnues. La méthode est fondée sur la distance de Kolmogorov et des échantillons approximatifs obtenus grâce au processus de Dirichlet. Nos résultats montrent que le test proposé est robuste par rap-

2B-C3: Nonregular Statistical Test Problems Problèmes de tests statistiques non réguliers

that the proposed test is robust with respect to any prior specification of the Dirichlet process. We provide simulated examples to illustrate the work of the method. Overall, the proposed method performs perfectly in many cases.

port à la loi a priori liée au processus de Dirichlet. Nous donnons des exemples simulés pour illustrer le fonctionnement de la méthode. Globalement, la performance de la méthode proposée est impeccable dans bon nombre de cas.

DEBARAJ SEN (Concordia University), **KRISHNA K. SAHA** (Central CT State University)

[Tuesday May 27 / mardi 27 mai, 10:50]

Robust Test for Testing Homogeneity of Dispersions for Several Groups of Counts Using Dissimilarity Measures

Test robuste de l'homogénéité des dispersions dans plusieurs groupes de dénombrements par mesures de dissimilitude

The robust test procedures for testing the equality of the dispersions of several groups in the analysis of extra-dispersed count data are developed. In this paper, we obtain the test statistics based on the dissimilarity measures for computing the within-group distances between individuals, and compute the critical values using the asymptotic distribution as well as the permutation procedure. These procedures are then compared with the existing method, in terms of their empirical levels and powers, through a simulation study. The results demonstrate that our proposed methods perform reasonably well as they hold their nominal levels and are robust to misspecification of the distribution of the responses. An example from biomedical practices is analyzed.

Nous développons des procédures de test robustes pour l'égalité des dispersions de plusieurs groupes dans l'analyse des données de dénombrement surdispersées. Dans cet article, nous obtenons des statistiques de test basées sur les mesures de dissimilitude pour calculer les distances entre individus au sein du groupe, puis nous calculons les valeurs critiques à l'aide de la distribution asymptotique ainsi que la procédure de permutation. Nous comparons ensuite le niveau et la puissance empirique de ces procédures à la méthode existante via une étude par simulation. Les résultats prouvent que les méthodes que nous proposons produisent des résultats relativement bons : leurs niveaux nominaux ne changent pas et ils sont robustes à l'erreur de spécification de la distribution des réponses. Nous analysons un exemple tiré de la pratique biomédicale.

JOSEPH FRANÇOIS TAGNE TATSINKOU, PIERRE DUSCHESNE & PIERRE LAFAYE DE MICHEAUX (Université de Montréal)

[Tuesday May 27 / mardi 27 mai, 11:05]

On Smooth Test of Goodness of Fit for Vector Time Series with Structured Parameterization

Un test lisse d'ajustement pour les séries chronologiques multivariées avec paramétrisation structurée

We propose in this work a smooth test of goodness of fit for vector autoregressive and moving average (VARMA) time series model with structured parameterization. As in Neyman (1937), the test consists in projecting the multivariate normal density in a space defined by a family of orthogonal functions and then to use the Rao score test for the parametric testing problem. The resulting test generalizes the popular Mardia test when using Hermite polynomials. Therefore, we provide a soft proof of affine invariance of that particular case avoiding some big calculus as in the Mardia proof of invariance. A simulation study and an application to real data are

Un test lisse d'ajustement est proposé pour les modèles vectoriels autorégressifs avec moyenne mobile (VARMA). Nous nous basons sur la méthodologie de Neyman qui consiste à projeter la densité nulle, qui est la loi normale multivariée, dans une famille d'alternatives. Le test obtenu généralise dans un cas particulier le populaire test de Mardia qui possède une propriété intéressante d'invariance. Nous avons montré, à partir d'une manipulation astucieuse des polynômes de Hermite, que la généralisation obtenue possède également cette propriété, évitant ainsi les calculs fastidieux comme dans le cas du test de Mardia. Une étude par simulation ainsi qu'une application aux données réelles sont présentées.

2B-C3: Nonregular Statistical Test Problems Problèmes de tests statistiques non réguliers

provided.

ZHENG SUN (Simon Fraser University)

[Tuesday May 27 / mardi 27 mai, 11:20]

Testing Goodness-of-Fit for Latent Variables

Comment tester la qualité d'ajustement de variables latentes

In this talk, a particular use of Bayesian methods allows consideration of the problem of testing the distribution of latent (unobserved) variables when these are connected by a known relationship to a set of observed variables. The technique is used to advance an interesting procedure introduced in geology by Krumbein and, for a modern example, to test the distribution of the frailty term (random effects) in a Cox Proportional Hazards (PH) model.

Dans cette présentation, une utilisation particulière de méthodes bayésiennes permet d'étudier le problème consistant à tester la distribution de variables latentes (non observées) lorsque celles-ci sont reliées par une relation connue à un ensemble de variables observées. La technique est utilisée pour généraliser une procédure intéressante introduite en géologie par Krumbein et, dans un exemple moderne, pour tester la distribution du terme de fragilité (effets aléatoires) dans un modèle des hasards proportionnels de Cox.

JOSHUA POHLKAMP-HARTT (Queen's University)

[Tuesday May 27 / mardi 27 mai, 11:35]

Bootstrapped Multitaper F-test for Signal Detection

Test F « multitaper » bootstrap pour la détection de signaux

Signal detection is an important part of applied sciences from wireless communications to healthcare diagnostics. A commonly used test for signal detection is the Multitaper F-test. While the F-test performs well on high powered signals, at moderately low power the detection rate diminishes. To improve upon the detection rates of moderately low powered signals we introduce a bootstrapping procedure for use in calculating the F-test. The procedure involves resampling the residuals within the F-test to produce a new F-statistic and then averaging over a set of these new estimates. We will discuss the convergence rate of the average of the resampled F-statistics and show that this new test out performs the traditional F-test for moderately powered signals on simulated data.

La détection de signaux constitue un élément important des sciences appliquées qui va des communications mobiles au diagnostic médical. Le test de Fisher (test F) de type « multitaper » est communément employé pour la détection de signaux. Bien que le test F donne de bons résultats pour les signaux de forte puissance, lorsque la puissance des signaux est plus faible son taux de détection diminue. Pour améliorer celui-ci, nous introduisons une procédure de bootstrap pour calculer le test F. Cette procédure implique un rééchantillonnage des résidus du test F pour produire une nouvelle statistique F ainsi que le calcul d'une moyenne sur l'ensemble de ces nouvelles estimations. Nous discuterons du taux de convergence de la moyenne des statistiques F rééchantillonnées et montrerons à l'aide de données simulées que ce nouveau test donne de meilleurs résultats que le test F traditionnel pour des signaux de puissance modérée.

2B-C4: Statistical Inference and Applications 3 Inférence statistique et applications 3

Chair/Présidente: Xin Gao (York University)

Room/Salle: UC - Croft House

Abstracts/Résumés

ALEXANDRE LEBLANC, STEPHANE DUROCHER, JASON MORRISON, MATTHEW SKALA & ROBERT FRASER

(University of Manitoba)

[Tuesday May 27 / mardi 27 mai, 10:20]

On the Probability of Being Interior to a Random Subset of a Sample

À propos de la probabilité d'être à l'intérieur d'un sous-ensemble aléatoire d'un échantillon

Let q denote a point of interest on the plane and S denote a sample of n bivariate observations in \mathbb{R}^2 . Now, consider taking a subset S' of S at random by including each element of S with probability $1/2$ independently of all other points. We derive the probability that the point q is interior to S' (i.e. inside the convex hull of S') and establish links between this problem and different concepts of data depth.

Soient q un point d'intérêt dans le plan et S un échantillon observé de vecteurs aléatoires dans \mathbb{R}^2 . Supposons qu'un sous-échantillon S' est construit en sélectionnant les points de S indépendamment les uns des autres avec probabilité $1/2$. Nous présentons le problème du calcul de la probabilité que le point q soit intérieur au sous-ensemble S' (i.e. à l'intérieur de l'enveloppe convexe de S') et établissons des liens entre ce problème et différents concepts de profondeur.

TOM LOUGHIN & WILLIAM RUTH (Simon Fraser University)

[Tuesday May 27 / mardi 27 mai, 10:35]

The Effect of Heteroscedasticity on Regression Trees

Effet de l'hétéroscédasticité sur les arbres de régression

Regression trees are becoming increasingly popular as omnibus predicting tools and as the basis of numerous modern statistical learning ensembles. Part of their popularity is their ability to create a regression prediction without ever specifying a structure for the mean model. However, the method implicitly assumes homogeneous variance across the entire explanatory-variable space. It is unknown how the algorithm behaves when faced with heteroscedastic data. In this study, we assess the performance of the most popular regression-tree algorithm in a single-variable setting under a very simple step-function model for heteroscedasticity. We use simulation to show that the locations of splits, and hence the ability to accurately predict means, are both influenced by the change in variance.

Les arbres de régression deviennent de plus en plus populaires comme outils de prédiction omnibus et comme fondement de nombreux ensembles d'apprentissage statistique moderne. Leur popularité tient en partie à leur capacité de réaliser une prédiction par régression, sans jamais spécifier une structure pour la moyenne dans le modèle. Toutefois, la méthode présuppose implicitement une variance homogène à travers l'espace des variables explicatives. Le mode de fonctionnement de l'algorithme avec des données hétéroscédastiques reste inconnu. Dans la présente étude, nous évaluons la performance de l'algorithme d'arbre de régression le plus populaire dans un contexte à variable simple sous un modèle en escalier très simple pour l'analyse de l'hétéroscédasticité. Nous utilisons la simulation pour montrer que l'emplacement des divisions et, par conséquent, la capacité de prédire précisément les moyennes, sont influencés par le changement de variance.

2B-C4: Statistical Inference and Applications 3 Inférence statistique et applications 3

MINA MIRSHAHI (École Polytechnique de Montréal), **MASOUD ASGHARIAN** (McGill University), **VAHID PARTOVI NIA** (École Polytechnique de Montréal)

[Tuesday May 27 / mardi 27 mai, 10:50]

Cubic Splines for Shapes in Two-Dimensional Space

Des splines cubiques pour des formes dans un espace à deux dimensions

The recent measurement systems provide three-dimensional image data for many micro-organisms including cells. We first model the projected cell surface using periodic splines. Other basis functions such as Fourier basis, or wavelet basis can be used as well. Having implemented R codes, the estimation process for fitting cubic splines, Fourier, and wavelets have been compared through some illustrative examples.

Les systèmes de mesure récents fournissent des données d'images en trois dimensions pour de nombreux micro-organismes, y compris les cellules. Nous commençons d'abord avec une spline unidimensionnelle en appliquant la même idée et en la généralisant aux fonctions en deux dimensions. Nous proposons une méthode de calcul relativement facile pour trouver une estimation lisse de fonctions périodiques et des courbes fermées dans l'espace à deux dimensions par le biais de splines de lissage. Après avoir mis en place des codes R, le processus d'estimation d'ajustement des splines cubiques, les fonctions de Fourier, et des ondelettes ont été comparées à travers quelques exemples illustratifs.

PAVEL PETROV & GISEON HEO (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 11:05]

Applications of Persistent Homology to DNA data

Applications de l'homologie persistante à des données sur l'ADN

Statistical applications of persistent homology are still being developed. This study attempts to identify whether these techniques can be used to find differences in DNA sequences. Patients with clostridium difficile are treated and their DNA sequences are sampled before and after treatment. Data is presented in the form of a distance matrix between DNA sequences in each patient. A summary of persistent homology, in particular, persistence landscapes are applied to search for differences where other statistical methods may fail. Once differences are found, the source of these differences has to be located

Les applications statistiques de l'homologie persistante sont encore en cours de développement. Cette étude vise à déterminer si ces techniques peuvent être utilisées pour repérer des différences de séquences d'ADN. Chez des patients traités pour une infection au Clostridium difficile, un échantillon de séquences d'ADN est pris avant et après leur traitement. Les données sont présentées sous forme de matrice de distances entre les séquences d'ADN de chaque patient. Un sommaire d'homologie persistante, et notamment des profils de persistance sont utilisés pour chercher des différences, là où d'autres méthodes statistiques échouent. Après avoir repéré des différences, il faut en déterminer la source.

WANHUA SU (MacEwan University), **PENG ZHANG** (Zhejiang University)

[Tuesday May 27 / mardi 27 mai, 11:20]

Quantile Methods for Estimating Bi-normal Parameters in ROC Analysis

Méthodes de quantiles pour estimer les paramètres bi-normaux dans une analyse ROC

Receiver operating characteristic (ROC) analysis is widely used in the evaluation of ranking systems and medical tests. When it comes to ROC curve

L'analyse de la fonction d'efficacité du récepteur (ou analyse ROC) est communément employée pour évaluer les systèmes de classification et les tests médicaux. S'agissant d'a-

2B-C4: Statistical Inference and Applications 3

Inférence statistique et applications 3

fitting and inference, little has been done outside the framework of Metz et al. (1998). We propose two new methods, based on the ranks of the two classes, for estimating the bi-normal parameters in ROC analysis: 1) the method of quantiles, and 2) the maximum likelihood method based on quantiles. Using both simulation and realdata examples, we study the performance of the two proposed quantile methods, and illustrate the use of the quantile methods in evaluating medical tests.

justement et d'inférence de la courbe ROC, peu d'études ont été effectuées en dehors du cadre de Metz et al. (1998). Nous proposons deux nouvelles méthodes fondées sur les cotes des deux classes pour estimer les paramètres binormaux dans une analyse ROC : 1) la méthode des quantiles, et 2) la méthode du maximum de vraisemblance fondée sur les quantiles. Avec des simulations et des exemples de données réelles, nous étudions la performance des deux méthodes proposées et en illustrons l'utilisation dans l'évaluation de tests médicaux.

KARIM RAHIM & DAVID THOMSON (Queen's University)

[Tuesday May 27 / mardi 27 mai, 11:35]

Change-point Detection using Level-of-change in Spectra

Détection des points de rupture à l'aide du degré de changement des spectres

Change-point detection approaches generally employ time domain techniques. We propose a frequency domain change-point detection method based on the estimated level-of-change in the spectra over time. Our method proposes using $\text{ave}(\log(S_1(f)) - \log(S_2(f)))$ where $S_1(f)$ and $S_2(f)$ are spectra at frequency f taken at different time sections. We demonstrate using simulations that when the spectra $S_1(f)$ and $S_2(f)$ are estimated using the lower bias, lower variance, higher computational cost multitaper spectral estimator, the proposed method is successful in change-point detection even when given a homoscedastic time series with no noticeable change in mean over time.

Les approches de détection des points de rupture emploient généralement des techniques temporelles. Nous proposons une méthode de détection des points de rupture fréquentielle fondée sur le degré de changement estimé des spectres dans le temps. Notre méthode propose d'utiliser $\text{ave}(\log(S_1(f)) - \log(S_2(f)))$ où $S_1(f)$ et $S_2(f)$ sont des spectres de fréquence f pris à différents temps. Nous montrons à l'aide de simulations que lorsque les spectres $S_1(f)$ et $S_2(f)$ sont estimés à l'aide de l'estimateur spectral à faible biais, faible variance et haut coût de calcul « multitaper », la méthode proposée donne de bons résultats même pour une série chronologique homoscedastique sans changement manifeste de moyenne dans le temps.

2C-P1: Actuarial Science Poster Session Séance d'affiches en science actuarielle

Room/Salle: Great Hall

Abstracts/Résumés

HUAN GAO, ROGEMAR MAMON & XIAOMING LIU (The University of Western Ontario)

[Tuesday May 27 / mardi 27 mai, 12:00]

Risk Measurement of a Guaranteed Annuity Option in a Stochastic Environment

Mesure du risque d'une option de rente garantie dans un environnement stochastique

We address the problem of setting capital reserves for a guaranteed annuity option (GAO). We consider a one-decrement actuarial model for the gross loss in which death is the only decrement, and the interest and mortality risk factors follow correlated affine structures. For the net loss assumption, a two-decrement actuarial model is proposed by incorporating a withdrawal probability with an affine functional form and correlated with both interest and mortality rates. Risk measures are determined using moment-based density method and benchmarked with Monte-Carlo simulation. We establish the relation between level of desired risk measure accuracy and required sample size under the constraints of computing time and memory. A sensitivity of parameter test demonstrates the feasibility of our proposed modelling frameworks.

Nous traitons le problème de la constitution de réserves de capital pour une option de rente garantie. Nous considérons un modèle actuariel à décrétement simple pour la perte brute où le décès est le seul décrétement et les facteurs de risque d'intérêt et de mortalité respectent des structures affines corrélées. Pour l'hypothèse de perte nette, nous proposons un modèle actuariel à décrétement double qui inclut une probabilité de retrait avec une forme fonctionnelle affine et corrélé aux taux d'intérêt et de mortalité. Nous déterminons les mesures de risque à l'aide de la méthode de densité basée sur les moments et par comparaison à une simulation de Monte-Carlo. Nous établissons la relation entre le niveau de précision de mesure du risque souhaité et la taille d'échantillon nécessaire sous la contrainte du temps de calcul et de la mémoire. Un test de la sensibilité du paramètre prouve la faisabilité de nos cadres de modélisation.

LUYAO LIN, LI CHEN, YI LU & GARY PARKER (Simon Fraser University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Insurance and Investment Risks of a Joint Life Insurance Portfolio

Risques d'assurance et d'investissement d'un portefeuille d'assurance-vie conjointe

A general portfolio of joint life insurance and last survivor insurance contracts is studied in a stochastic interest rate environment with dependent mortality models. The first two moments of the prospective losses are derived using two methods, one considering individual losses and the other studying random cash flows. The aggregated riskiness of the portfolio is decomposed into insurance risk and investment risk. For illustrative purposes, we use an autoregressive process to model the investment returns and a copula approach to capture mortality dependence. The effects of mortality dependence and model parameters on the riskiness of the portfolios are analyzed.

Nous étudions un portefeuille général de contrats d'assurance-vie conjointe et d'assurance du dernier survivant dans un environnement de taux d'intérêt stochastique avec modèles de mortalité dépendante. Nous dérivons les deux premiers moments des pertes potentielles par deux méthodes, l'une relative aux pertes individuelles et l'autre qui étudie les flux de trésorerie aléatoires. Nous décomposons le niveau de risque composé du portefeuille en risque d'assurance et risque d'investissement. À des fins d'illustration, nous utilisons un processus autorégressif pour modéliser le rendement de placement et une approche de copule pour saisir la dépendance de la mortalité. Nous analysons les effets de la dépendance de la mortalité et des paramètres du modèle sur le niveau de risque des portefeuilles.

2C-P1: Actuarial Science Poster Session Séance d'affiches en science actuarielle

YUZHOU ZHANG & BRUCE L. JONES (Western University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Modelling Disability Using Mixtures of Markov Chains

Modélisation de l'invalidité à l'aide de mélanges de chaînes de Markov

Multi-state models provide a convenient way of representing disability processes. In the simplest case, the model involves three states: active (not disabled), disabled and dead. While calculations are easy if one assumes that the model is a Markov chain, this is rarely appropriate. It is more reasonable to assume that the future of the process is influenced by its history. This occurs when one uses a mixture of Markov chains. We consider how to construct mixture models and investigate how the intensities of these processes depend on their histories.

Les modèles d'états multiples offrent un moyen pratique de représenter les processus d'invalidité. Dans le plus simple des cas, le modèle inclut trois états : actif (non invalide), invalide et mort. Les calculs sont simples si l'on assume que le modèle est une chaîne de Markov, mais c'est rarement le cas. Il est plus raisonnable d'assumer que le futur du processus est influencé par son histoire. C'est le cas lorsqu'on utilise un mélange de chaînes de Markov. Nous explorons comment construire des modèles de mélange et comment l'intensité de ces processus dépend de leur histoire.

2C-P2: Poster Session Séance d'affichage

Room/Salle: Great Hall

Abstracts/Résumés

JUSTIN ANGEVAARE, DANIEL GILLIS & GERARDA DARLINGTON (University of Guelph)

[Tuesday May 27 / mardi 27 mai, 12:00]

Efficient Bayesian Inference for Conditionally Autoregressive Models

Inférence bayésienne efficace pour modèles conditionnellement autorégressifs

We compare the performance of Metropolis-Hastings (MH) and Hamiltonian Monte Carlo (HMC) methods for Bayesian inference, with specific application to conditionally autoregressive (CAR) models. We perform a simulation study which investigates the efficiency of MH and HMC in estimation of the spatial correlation strength parameter of the CAR model. For this, data are simulated at various resolutions and spatial correlation strengths. A case study on the relative abundance of Lake Whitefish in Lake Huron through space is also presented. HMC has seen very little use with CAR models, and more complex CAR models, (e.g. multivariate or directional versions) may see benefit from its adaptation.

Nous comparons la performance des méthodes d'inférence bayésienne de Metropolis-Hastings (MH) et de Hamilton Monte Carlo (HMC), avec application aux modèles conditionnellement autorégressifs (CAR). Nous effectuons une étude par simulation qui étudie l'efficacité des méthodes MH et HMC pour estimer le paramètre de puissance de la corrélation spatiale du modèle CAR. Pour ce faire, nous simulons les données à plusieurs résolutions et puissances de corrélation spatiale. Nous présentons également une étude de cas sur l'abondance relative des corégones de lac dans le lac Huron dans l'espace. La méthode HMC est très peu employée avec les modèles CAR, mais les modèles CAR plus complexes (p. ex. version multivariée ou directionnelle) pourraient profiter de son adaptation.

CAROLYN AUGUSTA & ROB DEARDON (University of Guelph)

[Tuesday May 27 / mardi 27 mai, 12:00]

A Comparison of MCMC and MLE in Spatial Epidemic Models

Une comparaison entre MCMC et MLE pour des modèles spatiaux de maladie infectieuse

As an infectious disease epidemic emerges, implementation of control policies must be quick and efficient. To make informed decisions, it is key to identify important risk factors. This can most effectively be done using disease transmission models, typically fitted in a Bayesian Markov Chain Monte Carlo (MCMC) framework. Such a framework also allows the imputation of key, but often unobserved data, such as the time of infection. Maximum Likelihood Estimation (MLE) can be a faster alternative, especially if we have reliable data on infection times from good disease reporting systems. Here, we compare the performance of MCMC and MLE approaches via simulation studies for a spatial infectious disease model.

Au début d'une épidémie de maladie infectieuse, il faut que les méthodes de prévention et de contrôle des infections soient rapidement employées. Le développement des modèles de transmission des maladies infectieuses est la méthode la plus efficace d'identifier des facteurs de risques importants à la prise des décisions en connaissance de cause. Ces modèles suivent généralement le formalisme bayésien de Monte-Carlo par chaînes de Markov (MCMC). Ce formalisme permet l'imputation des données non-observées, comme le moment de l'infection. Si on peut accéder à un système fiable de déclaration des maladies, la méthode de vraisemblance maximale (MLE) peut être plus rapide que MCMC. L'efficacité des deux méthodes d'estimation est évaluée par simulation dans le contexte d'un modèle spatial de maladie infectieuse.

2C-P2: Poster Session Séance d'affichage

JOSE MARTIN CADENA-BARAJAS (UNIVERSIDAD VERACRUZANA), **SERGIO FRANCISCO JUAREZ-CERRILLO** (Universidad Veracruzana), **DAVID A. STEPHENS** (McGill University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Bayesian Modeling of Wind Speed and Direction

Modélisation bayésienne de la vitesse et de la direction du vent

We propose a Bayesian approach for analyzing wind speed and direction data recorded at "Laguna Verde" Nuclear Power Plant at Veracruz, Mexico. The model assumes speed follows a log-normal distribution and direction a multinomial distribution whose categories are given by a partition of the wind rose in 16 sectors. Selection of the conditional structure is performed by analysis of deviance for generalized linear models and correlation structure is modeled by regression functions. Inference about the unknown parameters is carried out with Markov chain Monte Carlo methods. The results showed a relationship between wind direction and speed, used for short-term forecast of both components.

Nous proposons une approche bayésienne pour analyser les données de vitesse et de direction du vent enregistrées à la centrale nucléaire de Laguna Verde à Veracruz, Mexique. Le modèle suppose que la vitesse obéit à une distribution log-normale et la direction à une distribution multinomiale dont les catégories sont données par une partition de la rose des vents en 16 secteurs. Nous sélectionnons la structure conditionnelle par analyse de la déviance pour les modèles linéaires généralisés et modélisons la structure de corrélation par fonctions de régression. Nous procédons à une inférence sur les paramètres inconnus par des méthodes Monte-Carlo par chaîne de Markov. Les résultats montrent une relation entre la direction et la vitesse du vent, relation qui servira pour la prévision à court terme des deux éléments.

YANLING CAI & JAMES ZIDEK (University of British Columbia)

[Tuesday May 27 / mardi 27 mai, 12:00]

"Quality Over Quantity": Damage Accumulation Models for Lumber Products

Qualité ou quantité : modèles d'accumulation des dommages pour les produits de bois d'œuvre

Lumber, as a natural building material, affects physical and mechanical properties of the structures built from it. Proof loading is a quality control technique in the lumber industry, which tests if a specimen survives breakage up to a pre-determined load level. The survivors are later distributed to the market. However, a large-scale experiment designed and conducted by statistics graduate students in a commercial testing lab indicates that the survivors are damaged. The quality control concerns the severity of the damage accumulated during proof loading. This poster explores potential statistical damage models driven by the data, and compares them to the damage accumulation models serving as a basis of the National Design Specification for Wood Construction by the American Wood Council.

Le bois d'œuvre, matériau de construction naturel, affecte les propriétés physiques et mécaniques des structures qu'il compose. L'épreuve de charge est une technique de contrôle de la qualité industrielle conçue pour tester si un spécimen résiste à la cassure jusqu'à un niveau de charge prédéterminé. Les survivants du test sont distribués sur le marché. Cependant, une expérience à grande échelle conçue et exécutée par des étudiants gradués en statistique dans un laboratoire de test commercial indique que ces survivants sont abîmés. Le contrôle de la qualité concerne la sévérité du dommage accumulé au cours de l'épreuve de charge. Notre affiche explore les modèles de dommages statistiques que les données évoquent et les compare aux modèles d'accumulation des dommages qui servent de base à la norme National Design Specification for Wood Construction du American Wood Council.

2C-P2: Poster Session Séance d'affichage

HAO CHEN & WILLIAM WELCH (University of British Columbia)

[Tuesday May 27 / mardi 27 mai, 12:00]

Sequential Experimental Design for Percentile Estimation -- A Floor System Example

Plan d'expérience séquentiel pour l'estimation des percentiles -- un exemple des systèmes de planchers

Computer models are often used to study physical and engineering processes. In this poster, we first introduce a computer (numerical) model implemented as part of the ongoing work of the Forest Products Stochastic Modelling Group at UBC. It is able to quantify the relationship between Modulus of Elasticity (MOE) and the corresponding deflection under static load in floor systems. We then build a statistical surrogate for the input-output relationship of the numerical model with a modest number of evaluations. We apply sequential methodology to the surrogate for estimating the extreme percentile, say 97.5 %, of the deflection distribution. We find that the accuracy of the surrogate is improved with a few more runs chosen by the sequential design.

Des modèles informatiques sont souvent utilisés pour étudier les processus physiques et d'ingénierie. Dans cette affiche, nous commençons par introduire un modèle informatique (numérique) mis en œuvre dans le cadre des travaux du Groupe de modélisation stochastique des produits forestiers de la UBC, modèle capable de quantifier la relation entre le module d'élasticité et la flexion sous charge statique correspondante dans les systèmes de planchers. Nous construisons ensuite un substitut statistique pour la relation entrées/sorties du modèle numérique avec un petit nombre d'évaluations. Nous appliquons une méthodologie séquentielle au substitut pour estimer le percentile extrême, soit 97,5 %, de la distribution de flexion. Nous concluons que la précision du substitut est améliorée par quelques itérations supplémentaires choisies par le plan séquentiel.

OSVALDO ESPIN-GARCIA (University of Toronto / Public Health Ontario), **MICHAEL LEBENBAUM** (Public Health Ontario), **LAURA ROSELLA & YE LI** (Public Health Ontario / University of Toronto)

[Tuesday May 27 / mardi 27 mai, 12:00]

Model Validation Via the 632+ Bootstrap Method in a Longitudinal Survey Sampling Context: An Application for the Obesity Population Risk Tool

Validation de modèles via la méthode du bootstrap 632+ dans un contexte d'échantillonnage de sondage longitudinal : application à l'Obesity Population Risk Tool

Built on the longitudinal component of the National Population Health Survey, the Obesity Population Risk Tool (OPoRT) is a longitudinal model aiming to estimate BMI trajectories over time in order to provide an accurate population-level estimate of the prevalence and cost of obesity, which is a serious public health problem. Since there is no external longitudinal population data to validate OPoRT, we use the 632+ bootstrap approach to assess calibration and discrimination. Considering that there is no standard when accounting for different sampling weights in longitudinal studies, three scenarios are proposed to compute the 632+ error estimator. In addition, a simulation study was conducted to assess these scenarios using well-known prediction measures such as Brier score and the Hosmer-Lemeshow statistic.

Basé sur le volet longitudinal de l'Enquête nationale sur la santé de la population, l'Obesity Population Risk Tool (OPoRT) est un modèle longitudinal conçu pour estimer les trajectoires d'IMC dans le temps afin de fournir une estimation précise à l'échelle de la population de la prévalence et du coût de l'obésité, qui constitue un sérieux problème de santé publique. Puisqu'il n'existe aucune donnée longitudinale démographique externe pour valider l'OPoRT, nous utilisons une approche de bootstrap 632+ pour évaluer la calibration et la discrimination. Étant donné qu'il n'existe aucune norme qui tienne compte de différents poids d'échantillonnage dans les études longitudinales, nous proposons trois scénarios pour calculer l'estimateur d'erreur 632+. De plus, nous effectuons une étude par simulation pour évaluer les scénarios à l'aide de mesures de prédiction comme l'indice de Brier et la statistique de Hosmer-Lemeshow.

2C-P2: Poster Session Séance d'affichage

ANDREW GARRO (Lakehead University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Sampling Techniques: A Detailed Analysis of Three Sector Funds

Techniques d'échantillonnage : analyse détaillée de trois fonds sectoriels

This case study will focus on an application involving data from three different sector mutual funds. The mutual funds under consideration include: the TD Health Sciences Fund, the TD Precious Metals Fund, as well as the TD Science & Technology Fund. The primary objective throughout the study will be to draw representative samples of daily closing prices by employing the simple random sampling, stratified sampling, and two-stage cluster sampling designs. In doing so, we will seek to determine which sampling technique best estimates the true distribution of the fund prices over the 2012-2013 fiscal year.

Cette étude de cas explore une application concernant des données en provenance de trois fonds communs de placement sectoriels : les fonds TD Sciences de la santé, TD Métaux précieux et TD Science et technologie. L'objectif principal de l'étude est de prélever des échantillons représentatifs des cours de clôture quotidiens en employant des plans d'échantillonnage simples aléatoires, stratifiés et en grappes à deux degrés. Ce faisant, nous cherchons à déterminer quelle technique d'échantillonnage estime au mieux la vraie distribution des cours des fonds sur l'exercice 2012-2013.

MOHAMMAD SAJJAD GHAEMI, BRUNO AGARD, MARTIN TRÉPANIÉ & VAHID PARTOVI-NIA (École Polytechnique de Montréal)

[Tuesday May 27 / mardi 27 mai, 12:00]

Computational Complexity in Hierarchical Clustering Algorithms

Complexité de calcul dans les algorithmes hiérarchiques

One of the most challenging tasks in data learning is grouping similar objects together. This problem, also called clustering, has been an important topic in different disciplines over the past decades. Hierarchical clustering is a breakthrough in this context, because of producing a visual guide as

a binary-tree to data grouping, known as dendrogram. Single-linkage, average-linkage, and complete-linkage are variants of hierarchical clustering. Though, the average-linkage provides a better statistical interpretation.

The trivial implementation of the average linkage is computationally expensive. Priority-queue data

structure can be used to decrease the computational cost. Furthermore, we suggest to benefit from smart updating of similarity measures.

Une des tâches les plus difficiles dans l'apprentissage de données est de regrouper ensemble des objets similaires. Ce problème, appelé segmentation, a été un thème important dans différentes disciplines au cours des dernières années.

Le regroupement hiérarchique est une percée dans ce contexte, parce qu'il produit un guide visuel sous la forme d'un arbre binaire de regroupement de données, appelé dendrogramme. Le lien-simple, le lien-moyen, et le lien-complet sont des variantes du regroupement hiérarchique, néanmoins le lien-moyen fournit une meilleure interprétation statistique.

La mise en œuvre triviale du lien-moyen, est coûteuse numériquement. Une structure de données de priorité peut être utilisée pour diminuer ce coût de calcul. En plus, nous proposons de bénéficier de la mise à jour intelligente des mesures de similarité.

2C-P2: Poster Session Séance d'affichage

CHRISTOPHER GRAVEL & PATRICK FARRELL (Carleton University), **DANIEL KREWSKI** (University of Ottawa)
[Tuesday May 27 / mardi 27 mai, 12:00]

Resampling for Outcome Misclassification: Internal Validation Sampling Concerns

Rééchantillonnage pour erreur de classement des résultats : problèmes d'échantillonnage de validation interne

Internal validation samples can be used to address issues of outcome misclassification. These samples are gathered by resampling the original data and validating a small proportion of it against a gold standard. The resulting information can be used to adjust estimates of risk or association for the presence of misclassification bias. Resampling procedures are often expensive and are subject to the probabilistic structure of the data, which can be difficult to acquire. Motivated by electronic health records data we propose a set of guidelines designed to assist in the selection of techniques for validation sampling and for the determination of validation sample size.

Les échantillons de validation interne permettent de résoudre les problèmes d'erreur de classement des résultats. Ces échantillons sont prélevés en rééchantillonnant les données originales puis en validant une petite partie à partir d'une norme de référence. Les informations obtenues permettent d'ajuster les estimations du risque ou d'association en fonction du biais d'erreur de classement. Les procédures de rééchantillonnage sont souvent coûteuses et souffrent de la structure probabiliste des données, qui peuvent être difficiles à acquérir. Motivés par les dossiers de santé électroniques, nous proposons un ensemble de directives conçues pour faciliter la sélection de techniques pour l'échantillonnage de validation et la détermination de la taille de l'échantillon de validation.

YUMI KONDO, YUMI KONDO, YINSHAN ZHAO & JOHN PETKAU (University of British Columbia)

[Tuesday May 27 / mardi 27 mai, 12:00]

Who Is a Responder? Identification of Treatment Responders Based on Longitudinally Collected Multiple Endpoints in Multiple Sclerosis Patients

Qui est un répondant? Identification des répondants au traitement en fonction de plusieurs critères collectés longitudinalement chez les patients atteints de sclérose en plaques

Currently, there is no general framework to determine whether a multiple sclerosis patient is responsive to a treatment based on longitudinally collected multiple MRI endpoints such as Burden of Disease (BOD) and lesion count. Existing criteria are often ad-hoc and based on only a single endpoint. For each treated patient with certain baseline characteristics, our novel procedure estimates the likelihood that in the absence of treatment the patient would have experienced a disease progression smaller than that observed. If this probability is small, then observing less disease progression in the absence of treatment is unlikely. Hence we can identify such treated patients as treatment responders. This poster demonstrates our procedure when disease progression is measured by BOD and lesion count.

Actuellement, il n'existe aucun cadre général permettant de déterminer si un patient atteint de sclérose en plaques réagit à un traitement sur la seule base des critères IRM collectés longitudinalement tels la charge de morbidité (CM) et le nombre de lésions. Les critères existants sont souvent arbitraires et fondés sur un seul paramètre. Pour chaque patient traité présentant certaines caractéristiques de base, notre procédure estime la probabilité qu'en l'absence de traitement il aurait vu sa maladie progresser moins que ce qui est observé. Si cette probabilité est faible, alors il est improbable que la maladie progresse moins en l'absence de traitement. Nous pouvons donc identifier ces patients traités comme étant des répondants au traitement. Notre affiche démontre notre procédure lorsque la progression de la maladie est mesurée en fonction de la CM et du nombre de lésions.

2C-P2: Poster Session Séance d'affichage

VIOLETA KOVACEV-NIKOLIC & GISEON HEO (University of Alberta), **PETER BUBENIK** (Cleveland State University), **DRAGAN NIKOLIC** (University of Alberta (at the time of initial research))

[Tuesday May 27 / mardi 27 mai, 12:00]

Topological Data Analysis of the Maltose Binding Protein

Analyse de données topologiques de la protéine de fixation du maltose

Fourteen structures of the maltose binding protein are investigated using topological data analysis. Starting from a dynamical physical model, each of the fourteen structures is approximated by a point-cloud dataset built from 370 amino-acid residues. The first part of analysis focuses on dissimilarity between closed and open structures of the protein; a test statistic based on the recently introduced notion of persistence landscapes yields compelling evidence on a statistically significant difference between the two protein groups for H_0 and H_1 homology. The second part of analysis shows that the majority of active sites and allosteric pathway residues cluster near the most persistent loop in the filtered simplicial complex, which is a qualitative result with potential in applications involving glucose biosensors.

Nous étudions quatorze structures de la protéine de fixation du maltose à l'aide d'une analyse de données topologiques. À partir d'un modèle physique dynamique, nous procédons à une approximation de chacune des quatorze structures à l'aide d'un ensemble de données par nuages de points construit à partir de 370 résidus d'acides aminés. Dans un premier temps, nous nous concentrons sur la dissimilitude entre les structures fermées et ouvertes de la protéine ; une statistique de test fondée sur la notion récente de paysages persistants (persistence landscapes) produit des preuves convaincantes d'une différence statistiquement significative entre les deux groupes de protéines pour l'homologie H_0 et H_1 . Ensuite, l'analyse montre que la majorité des sites actifs et des résidus de voie allostérique se retrouvent près de la boucle la plus persistante du complexe simplicial filtré, résultat qualitatif qui pourrait être utile pour les applications sur les biocapteurs du glucose.

JIA LI & TARANEH ABARIN (Memorial University of Newfoundland)

[Tuesday May 27 / mardi 27 mai, 12:00]

Bias Study of the Naive Estimator in a Longitudinal Linear Mixed-effects Model with Measurement Error and Misclassification in Covariates

Étude du biais de l'estimateur naïf dans un modèle à effets mixtes linéaire longitudinal avec erreurs de mesure et de classement des covariables

The research presents a generalized quasi-likelihood approach to estimate the parameters in a longitudinal linear mixed-effects model. In this model, we consider measurement error and misclassification in the covariates. Moreover, a classical measurement error for continuous covariates, and misclassification for discrete covariates up to three categories, is considered. Through simulation studies, we observe the impact of each parameter of the model on the bias of the naive estimation, when the other parameters stay unchanged.

La recherche présente une approche de quasi-vraisemblance généralisée pour l'estimation des paramètres dans un modèle à effets mixtes linéaire longitudinal. Dans ce modèle, nous étudions l'erreur de mesure et de classement des covariables. Nous explorons aussi une erreur de mesure classique pour les covariables continues, ainsi qu'une erreur de classement pour les covariables discrètes de trois catégories ou moins. À l'aide d'études par simulation, nous observons l'impact de chaque paramètre du modèle sur le biais de l'estimation naïve lorsque les autres paramètres demeurent inchangés.

2C-P2: Poster Session Séance d'affichage

ZHONG LI & KRISTINA SENDOVA (University of Western Ontario)

[Tuesday May 27 / mardi 27 mai, 12:00]

On a Dual Risk with Dependence Between Gains Sizes and Inter-Gain Times

Double risque avec dépendance entre la taille des gains et le temps entre gains

Dual risk models may be used to model the revenue process of a company with constant expense rate and occasional gains. In this paper, we consider a dual risk model with a dependence structure between the distributions of gain sizes and inter-gain times. Two different types of dependence structures are examined, where the pair-wise events of the elapsed time and gain amount are independent or not. Solutions for the Laplace transform of the time to ruin are obtained. Applications and numerical examples are also provided.

Les modèles à double risque permettent de modéliser le processus de création de revenus d'une entreprise dont le taux de frais est constant et les gains sont occasionnels. Dans cette présentation, nous étudions un modèle à double risque avec une structure de dépendance entre la distribution de la taille des gains et du temps entre gains. Nous examinons deux types de structures de dépendance, avec ou sans indépendance entre les paires d'événements « temps écoulé » et « montant du gain ». Nous obtenons des solutions pour la transformée de Laplace du temps avant la faillite. Nous proposons des applications et des exemples numériques.

ERIN LUNDY (Western University), **ELIZABETH JUAREZ-COLUNGA** (University of Colorado Denver), **CHARMAINE DEAN** (Western University)

[Tuesday May 27 / mardi 27 mai, 12:00]

A Joint Model for Detecting Patterns of Antisocial Behaviour Among Serious Juvenile Offenders

Modèle conjoint pour la détection de tendances de comportement antisocial chez les délinquants juvéniles graves

Longitudinal studies of criminal behaviour often collect multivariate count data which are characterized by excess zeros. Incorporating similar temporal trends of risk reflects underlying shared risk factors and provides improved precision for the estimation of underlying risks, by borrowing information across outcomes as well as from neighbouring time points. In the univariate case, zero-inflated Poisson regression models have been extended to the longitudinal setting by incorporating random effects to accommodate within-individual correlation. Joint modeling in which model components are linked across outcomes through individual-specific frailties provides a framework for extending the existing models for zero-inflated longitudinal count data to accommodate several outcomes. Analysis of data from a large-scale study on offending patterns of serious juvenile offenders is presented.

Les études longitudinales du comportement criminel recueillent souvent des données de dénombrement multivariées à surreprésentation de zéros. L'inclusion de tendances temporelles de risques similaires permet de faire ressortir des facteurs de risque sous-jacents partagés et d'améliorer la précision de l'estimation des risques sous-jacents, en partageant des informations en provenance d'autres résultats ou d'autres temps voisins. Dans le cas univarié, les modèles de régression de Poisson à surreprésentation de zéros ont été étendus au cadre longitudinal en y incluant des effets aléatoires afin de tenir compte de la corrélation chez l'individu. La modélisation conjointe, où les éléments de chaque modèle sont reliés pour plusieurs résultats via des fragilités spécifiques à l'individu, offre un cadre où les modèles existants de données de dénombrement à surreprésentation de zéros peuvent tenir compte de divers résultats. Nous présentons une analyse de données tirées d'une étude à grande échelle sur les tendances de délinquance des délinquants juvéniles graves.

2C-P2: Poster Session Séance d'affichage

MAHMUD MAMUN, H.S. SHIN, W. BURR, M. SMITH-DOIRON, B. JOVIC, D. STIEB & B. JESSIMAN (Health Canada)

[Tuesday May 27 / mardi 27 mai, 12:00]

Air Health Indicator (AHI)

Indicateur air-santé (IAS)

The Air Health Indicator (AHI) provides a measure of the impact of outdoor air pollution on Canadian public health, in particular cardiopulmonary (heart-lung) deaths. Heart-lung deaths do not result solely from air pollution exposure (e.g., ozone and fine particulate matter), but rather from the combination of a variety of causes. The AHI aims to identify the risk for heart-lung deaths related solely to exposure to air pollution. The AHI shows a slight increasing trend between 1990 and 2007 for heart-lung deaths that can be attributed to the exposure to ground-level ozone. The measure also indicates that 5

L'indicateur air santé (IAS) permet de mesurer l'impact de la pollution de l'air extérieur sur la santé publique Canadienne, en particulier les mortalités cardiopulmonaires (cœur-poumon). La mortalité cardiopulmonaire ne provient pas uniquement de l'exposition à la pollution de l'air (p. ex. l'ozone et les particules fines), mais plutôt de la combinaison d'un nombre de facteurs de risque. L'IAS a pour but d'identifier les risques de mortalité cardiopulmonaire qui sont strictement reliées à l'exposition à la pollution de l'air. L'IAS démontre une légère tendance à la hausse entre 1990 et 2007 pour les décès cardiopulmonaires qui peuvent être attribués à l'exposition à l'ozone au niveau du sol. Cette mesure indique aussi que 5

LUCIA MIREA (Maternal-Infant Care Research Centre, Mount Sinai Hospital and the University of Toronto), **JUNMIN YANG** (Maternal-Infant Care Research Centre, Mount Sinai Hospital), **SHOO K. LEE & PRAKESH S. SHAH** (Pediatrics, Mount Sinai Hospital and the University of Toronto)

[Tuesday May 27 / mardi 27 mai, 12:00]

Evaluation of Propensity Score Methods for Multiple Treatment Groups

Évaluation des méthodes de score de propension pour les groupes de traitement multiples

Propensity score (PS) methods can correct treatment selection bias in observational studies. For multiple treatments, standard PS-matching, PS-stratification and PS-adjustment can compare treatments pair-wise, or multinomial logistic regression can estimate a multi-component PS used in Multi-PS-weighted and Multi-PS-adjusted analyses to compare multiple treatments simultaneously. Bias, MSE and 5% type I error of PS methods for multiple treatments were evaluated using simulations. A binary confounder associated with a 3-level treatment (OR= 2, 5) and a binary outcome (OR=5), and several treatment-outcome association levels (OR=1, 2, 5) were generated for 1000 replicate datasets each with 5000 subjects. Among all methods, Multi-PS-adjustment had nominal 5% type I error, negligible bias (<0.004) and low MSE (<0.01), and is recommended for analyses of multiple treatments.

Les méthodes de score de propension (SP) permettent de corriger le biais de sélection du traitement dans les études par observation. Pour des traitements multiples, l'appariement des SP standard, la stratification des SP et l'ajustement des SP permet de comparer des paires de traitement ; la régression logistique multinomiale permet d'estimer un SP multi-éléments utilisé dans les analyse de SP multiples pondérés ou ajustés pour comparer plusieurs traitements en même temps. Nous évaluons le biais, l'erreur quadratique moyenne et l'erreur de type I à 5 % des méthodes de SP pour les traitements multiples à l'aide de simulations. Nous générons un facteur confondant binaire associé à un traitement de trois niveaux (OR= 2, 5) et un résultat binaire (OR=5), ainsi que plusieurs niveaux d'association traitement/résultat (OR=1, 2, 5) pour 1000 ensembles de données de réplication contenant chacun 5000 sujets. De toutes les méthodes, l'ajustement multi-SP présente une erreur de type I de 5 % nominale, un biais négligeable (<0,004) and une erreur quadratique moyenne faible (<0,01) ; elle est donc recommandée pour les analyses de traitements multiples.

2C-P2: Poster Session Séance d'affichage

KHURRAM NADEEM, YING ZHANG & HUGH CHIPMAN (Acadia University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Integrating Population Dynamics Models and Distance Sampling Data: A Hierarchical State-Space Approach
Intégration de modèles de dynamiques de la population et données d'échantillonnage de la distance : une approche espace d'états hiérarchique

A time series of animal abundance estimates, together with stochastic population dynamics (SPD) models, are key ingredients of a population viability analysis (PVA) routinely used to estimate extinction risk of the species. Recently, distance sampling methodology is becoming increasingly popular for wildlife abundance estimation owing to its cost-effectiveness and ease of implementation. In this work we use density dependent SPD models to model long-term dynamics of wildlife populations using distance sampling data. We employ hierarchical state-space models (HSSMs) to link the stochastic distance sampling model to the SPD model, while simultaneously allowing for valid inference on both model components. We also illustrate the proposed modeling approach in the context of white-tailed deer (*Odocoileus virginianus*) population reconstruction in Nova Scotia.

Une série chronologique d'estimations de l'abondance d'animaux et des modèles de dynamiques de la population stochastiques (DPS) sont des ingrédients clés d'une analyse de viabilité de la population (AVP) couramment utilisée pour estimer le risque d'extinction d'une espèce. Récemment, les méthodes d'échantillonnage de la distance sont devenues plus populaires pour estimer l'abondance de la faune, car elles offrent un bon rapport coût/efficacité et sont faciles à mettre en œuvre. Dans cette présentation, nous utilisons des modèles de DPS dépendants de la densité pour modéliser la dynamique à long terme de populations fauniques à l'aide de données d'échantillonnage de la distance. Nous employons des modèles espaces d'états hiérarchiques (MEEH) pour lier le modèle d'échantillonnage de la distance stochastique et le modèle de DPS, tout en tenant compte d'une inférence valide sur les deux éléments de modèles. Nous illustrons l'approche de modélisation proposée dans le contexte du repeuplement en cerfs de Virginie (*Odocoileus virginianus*) en Nouvelle-Écosse.

KUNASEKARAN NIRMALKANNA & BOXIN TANG (Simon Fraser University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Optimal Designs of Two-Level Factorials when $N \equiv 1$ and $2 \pmod{4}$ under a Baseline Parameterization
Plans factoriels optimaux à deux modalités où $N \equiv 1$ et $2 \pmod{4}$ avec un paramétrage comportant un niveau de base

This work considers two-level factorial designs under a baseline parameterization where the two levels are denoted by 0 and 1. When only main effects are of interest, such designs are equivalent to the biased spring balance weighing designs. Commonly we assume that the interactions are negligible, but if this is not the case then these non-negligible interactions will bias the main effect estimates. We define a design as optimal if it has the minimum bias among highly efficient designs. Optimal designs for $N \equiv 0 \pmod{4}$, where N is the run size, were discussed by Mukerjee & Tang (2012). We continue this line of study by investigating optimal designs for the cases $N \equiv 1$ and $2 \pmod{4}$.

Cette étude analyse des plans factoriels à deux modalités avec un paramétrage comportant un niveau de base, où les deux modalités sont désignées par 0 et 1. Lorsque seuls les effets principaux présentent un intérêt, de tels plans équivalents à des plans à pondération dynamométrique biaisée. En général, nous présumons que les interactions sont négligeables, mais si ce n'est pas le cas, ces interactions non négligeables biaiseront les estimations des effets principaux. Nous définissons un plan comme étant optimal s'il présente le biais le plus faible parmi des plans très efficaces. Des plans optimaux pour $N \equiv 0 \pmod{4}$, où N est la taille de l'essai, ont fait l'objet d'un article par Mukerjee et Tang (2012). Nous poursuivons cette étude en investiguant des plans optimaux pour des cas $N \equiv 1$ et $2 \pmod{4}$.

2C-P2: Poster Session Séance d'affichage

MOHAMMAD NOURMOHAMMADI, MOHAMMAD JAFARI JOZANI & BRAD JOHNSON (University of Manitoba)

[Tuesday May 27 / mardi 27 mai, 12:00]

Distribution-Free Tolerance Intervals with Randomized Nomination Samples

Intervalle de tolérance non paramétriques avec des échantillons nominatifs randomisés

Tolerance intervals based on randomized nomination sampling (RNS) is discussed in this poster. We study the performance of our proposed RNS tolerance intervals based on the corresponding coverage probabilities and the necessary sample size for their existence with those based on simple random sampling (SRS). The efficiency of the constructed RNS-based tolerance intervals compared to the SRS counterparts is discussed. We investigate the performance of RNS tolerance intervals for different values of the design parameters and various population shapes. We find the values of the design parameters which improves RNS over SRS. The RNS design in presence of ranking error is discussed and a new method for estimating ranking error probabilities is proposed.

Dans cet exposé, nous discutons des intervalles de tolérance basés sur l'échantillonnage nominatif randomisé (ÉNR). Nous étudions la performance des intervalles de tolérance ÉNR proposés en fonction de leur probabilité de couverture correspondantes et de la taille d'échantillon nécessaire pour leur existence, puis nous les comparons à ceux basés sur un échantillonnage aléatoire simple (ÉAS). Nous discutons de l'efficacité des intervalles de tolérance structurés en fonction d'un ÉNR en comparaison avec un ÉAS. Nous étudions la performance des intervalles de tolérance des ÉNR pour des valeurs de paramètres différentes et des formes de population variées. Nous signalons les valeurs des paramètres pour lesquels les performances de l'ÉNR sont supérieures à celles de l'ÉAS. Nous discutons aussi du plan d'ÉNR en présence d'erreurs de rangs, et nous proposons une nouvelle méthode d'estimation des probabilités d'erreurs de rangs.

ALBERTO PADILLA (BANCO DE MEXICO)

[Tuesday May 27 / mardi 27 mai, 12:00]

Two-Stage Cluster Sampling: Bounds for the Variance, Coefficient of Variation and Design Effect of Proportions

Échantillonnage en grappes à deux étapes : bornes pour la variance, le coefficient de variation et l'effet du plan sur les proportions

It is a well known result in the estimation of proportions using simple random sampling that the variance of an estimated proportion attains its maximum when the proportion equals 0.5. In the absence of information of the variable of interest, the maximum value of the variance can be used to compute the sample size. We build expressions for the minimum and maximum variances of an estimated proportion under two-stage cluster sampling with equal sizes, using simple random sampling in both stages. As a by-product it is immediate to obtain bounds for the coefficient of variation and the design effect. Some examples are given related to the computation of the bounds and to sample size determination.

Pour l'estimation d'une proportion à l'aide d'un échantillonnage aléatoire simple, il est bien connu que la variance de la proportion estimée atteint son maximum lorsque la proportion égale 0,5. En l'absence d'information sur la variable d'intérêt, la valeur maximale de variance peut être utilisée pour calculer la taille d'échantillon. Nous formulons des expressions de variance minimum et maximum d'une proportion estimée dans un échantillonnage en grappes à deux étapes et de taille égale, en utilisant dans les deux étapes un échantillonnage aléatoire simple. Accessoirement, des bornes pour le coefficient de variation et l'effet de plan sont immédiatement établies. Des exemples traitent des calculs des bornes et de la détermination de la taille d'échantillon.

2C-P2: Poster Session Séance d'affichage

LEILA SINAEI (Shiraz University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Estimation Of Mean In Double Sampling For Stratification Using Multivariate Auxiliary Information

Estimation de la moyenne avec un échantillonnage double stratifié à l'aide d'information auxiliaire multidimensionnelle

Double sampling (or two-phase sampling) is a sampling design which utilizes information on one or several auxiliary variables, correlated with the study variable, to increase the precision of estimators of population parameters of interest. In double sampling for stratification (DSS), the auxiliary information gathered in the first phase is used to stratify the sampled units. Then, in the second phase, subsamples are selected and the study variable is measured. For DSS sampling, two classes of estimators are proposed and asymptotically optimum estimators in these classes are identified, with approximate variance formulae. The proposed estimators are compared with the corresponding estimators based on un-stratified double sampling, and illustrated in an empirical study.

L'échantillonnage double (ou l'échantillonnage à deux degrés) est une méthode d'échantillonnage qui utilise l'information d'une ou de plusieurs variables auxiliaires, corrélées avec la variable à l'étude, afin d'augmenter la précision des estimateurs des paramètres de la population qui présentent un intérêt. Dans l'échantillonnage double de stratification (ÉDS), l'information auxiliaire recueillie dans la première phase est utilisée pour stratifier les unités échantillonnées. Dans la deuxième phase, des sous-échantillons sont sélectionnés et la variable à l'étude est mesurée. Pour un ÉDS, deux classes d'estimateurs sont proposés, parmi lesquelles des estimateurs asymptotiquement optimaux sont identifiés, avec des formules de variance approximative. Les estimateurs proposés sont comparés avec les estimateurs correspondants basés sur un échantillonnage double non stratifié, puis illustrés dans une étude empirique.

ANU STANLEY, ROB DEARDON & ZENY FENG (University of Guelph)

[Tuesday May 27 / mardi 27 mai, 12:00]

Using School Absenteeism in Disease Surveillance Models: Refining of Surveillance Threshold

Utilisation de l'absentéisme scolaire dans des modèles de surveillance épidémiologique : le raffinement du seuil de surveillance

We evaluated the effectiveness in using school absenteeism data for the early detection of influenza outbreaks in the community, by examining data collected by the Wellington-Dufferin-Guelph Public Health (WDGPH) for the period of January 2008 to May 2013. The model was developed linking real outbreaks, defined by aggregated hospital data in the WDG area, to the school absenteeism data. We used classical surveillance models such as the moving average(MA) and exponentially weighted moving average(EWMA) to compute a unique baseline for each school on a given day and compared its sensitivity/specificity to that of a steady baseline currently used by the WDGPH. Therefore we were able to decrease the false-positive and false-negative absenteeism alerts to improve the surveillance program at WDGPH.

Nous évaluons l'efficacité de l'utilisation des données d'absentéisme scolaire pour la détection précoce des épidémies de grippe dans la collectivité, en examinant des données recueillies par le service de santé Wellington-Dufferin-Guelph Public Health (WDGPH) durant la période de janvier 2008 à mai 2013. Le modèle est élaboré en liant des épidémies véritables, telles que définies par les données hospitalières cumulatives de l'hôpital pour la région de WDG, aux données d'absentéisme scolaire. Nous utilisons des modèles de surveillance classiques comme la moyenne mobile et la moyenne mobile à pondération exponentielle afin de calculer, pour une journée quelconque, une valeur normale unique pour chaque école, puis nous comparons sa sensibilité et sa spécificité à celle de la valeur normale couramment utilisée par le WDGPH. En conséquence, nous pouvons diminuer les fausses alertes positives ou négatives afin d'améliorer le programme de surveillance au WDGPH.

2C-P2: Poster Session Séance d'affichage

JIAYING (VICKY) WENG (Simon Fraser University)

[Tuesday May 27 / mardi 27 mai, 12:00]

Maximin Strong Orthogonal Arrays

Tableaux orthogonaux robustes maximin

Orthogonal arrays have been widely used in computer experiments, either directly or via OA-based Latin hypercubes. He and Tang (2013) introduced and constructed a new class of arrays, strong orthogonal arrays, for computer experiments. Strong orthogonal arrays of strength t enjoy better space-filling properties than comparable orthogonal arrays of strength t in all dimensions lower than t . Given a single orthogonal array, many strong orthogonal arrays can be generated using the method of He and Tang (2013). We examine the selection of better strong orthogonal arrays using the maximin distance; a criterion attempting to place points in a design region so that no pairs of points are too close. This poster presents the method and results from such a study.

Les tableaux orthogonaux ont été largement utilisés dans des expériences informatiques, soit directement ou au moyen d'hypercubes latins basés sur des tableaux orthogonaux. He et Tang (2013) ont structuré une nouvelle classe de tableaux orthogonaux robustes à des fins d'expériences informatiques. Les tableaux orthogonaux robustes de force t sont dotés de propriétés d'encombrement spatial supérieures à celles des tableaux orthogonaux de force t comparables dans toutes dimensions inférieures à t . Grâce à la méthode de He et Tang (2013), plusieurs tableaux orthogonaux robustes peuvent être générés avec un seul tableau orthogonal. Nous examinons la sélection des meilleurs tableaux orthogonaux robustes en utilisant la distance maximin, un critère conçu pour placer des points dans une zone, de sorte qu'aucune paire de points ne soit trop proche. Cet exposé présente la méthode et les résultats de cette étude.

JINKUN XIAO & A.IAN MCLEOD (University of Western Ontario)

[Tuesday May 27 / mardi 27 mai, 12:00]

Bootstrapping Improvement to Portmanteau Test in ARFIMA Model

Amélioration au test portemanteau par le bootstrap dans un modèle ARFIMA

The portmanteau test statistics Pena and Rodriguez (2003) developed serves as an effective goodness-of-fit test in time series analysis. Lin and McLeod(2006) points out its limit and propose the algorithm to employ this test statistics in Monte Carlo Test. However,when this test was applied to ARFIMA model,the p-value estimation is highly biased and unreliable. In order to improve the test power for the algorithm,wild dependent bootstrapping are applied with modification and significant improvement are achievement. Due to the computation intensity,parallel computing technique are employed to reduce the computational time.

Le test statistique du portemanteau élaboré par Pena et Rodriguez (2003) sert efficacement de test d'adéquation dans l'analyse des séries chronologiques. Lin et McLeod (2006) soulignent ses limites et proposent un algorithme pour utiliser ces statistiques dans le test de Monte Carlo. Par contre, lorsque ce test est appliqué au modèle ARFIMA, l'estimation de la valeur p est très biaisée et non fiable. Afin d'améliorer l'efficacité du test pour l'algorithme, un procédé d'auto-amorçage à signes aléatoires est utilisé avec des modifications, et il en résulte une amélioration significative. En raison de l'envergure des calculs, des techniques de calculs parallèles sont utilisées pour réduire le temps de calcul.

2C-P3: Probability Student Poster Session Séance d'affiches étudiants en probabilité

Room/Salle: Great Hall

Abstracts/Résumés

CHI DONG (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 12:00]

Nonlinear Filtering on Completely Regular Spaces

Filtrage non linéaire sur des espaces complètement réguliers

This project generalizes the classical nonlinear filtering problem from Polish or compact Hausdorff spaces to completely regular spaces. In the first theme we develop a new technique of constructing an alternative compact Polish state space as well as an indistinguishable replicate signal in it. We re-define and discuss the proposed filtering problem in this Polish space, from which the desired filtering equations can be generalized within the original space. In the second theme, the long-time typical behavior of the signal-filter pair is innovatively analyzed by a redefined filtering problem on an enlarged probability space. Therein we establish the weak convergence of finite-dimensional distributions of the signal-filter pair. As a byproduct, we identify the limit process as a stationary signal-filter pair.

Ce projet généralise le problème classique de filtrage non linéaire d'espaces polonais ou Hausdorff compact à des espaces complètement réguliers. En premier thème, on développe une nouvelle technique pour construire un espace d'état polonais compact alternatif comportant une copie de signal indiscernable. On redéfinit et discute du problème de filtrage proposé dans cet espace polonais, à partir duquel les équations de filtrage désirées peuvent être généralisées dans l'espace d'origine. Dans le second thème, le comportement typique à long terme de la paire signal-filtre est analysé de manière innovatrice par un problème de filtrage redéfini sur un espace de probabilité élargi. On y établit la convergence en loi des distributions de dimension finie de la paire signal-filtre. Comme sous-produit, on identifie le processus limite en tant que paire signal-filtre stationnaire.

HAOSUI DUANMU (University of Toronto)

[Tuesday May 27 / mardi 27 mai, 12:00]

Nonstandard Analysis and Its Application to Convergence of Markov Processes

Analyse non standard appliquée à la convergence des processus de Markov

This project uses nonstandard analysis to construct so-called *hyperfinite Markov processes*, and applies them to the convergence to stationarity of standard continuous-time Markov processes. A hyperfinite Markov process is a Markov process with time line $T = \{0, \delta t, \dots, 1, \dots, K\}$ for some (infinitely large) hyperinteger K and (infinitely small) infinitesimal δt . For every regular continuous-time Markov process $\{X_t\}_{t \in \mathbb{R}}$, we construct a corresponding hyperfinite Markov process $\{^*X_t\}_{t \in T}$. We show that these two process have the same stationary distribution and that $\{^*X_t\}_{t \in T}$ behaves like $\{X_t\}_{t \in \mathbb{R}}$ in a precise mathematical sense. The proof that $\{^*X_t\}_{t \in T}$ converges to its stationary distribution

Ce projet fait appel à une analyse non standard pour construire des procédés de Markov de type hyperréel *Processus de Markov hyperfinis* et les appliquer à la convergence des processus de Markov à temps continu standard vers la stationnarité. Un processus de Markov hyperréel comprend des données temporelles $T = \{0, \delta t, \dots, 1, \dots, K\}$ pour des nombres entiers hyperréels (infiniment grands) K et infinitésimaux (infiniment petits) δt . Pour tout processus de Markov à temps continu standard, $\{X_t\}_{t \in \mathbb{R}}$, nous construisons un processus de Markov hyperréel correspondant, $\{^*X_t\}_{t \in T}$. Nous démontrons que ces deux processus ont une même distribution stationnaire et que, dans un sens mathématique précis, $\{^*X_t\}_{t \in T}$ se comporte comme $\{X_t\}_{t \in \mathbb{R}}$. La preuve que $\{^*X_t\}_{t \in T}$ converge vers une distribution stationnaire nous

2C-P3: Probability Student Poster Session Séance d'affiches étudiants en probabilité

then leads to a proof of convergence of standard continuous-time Markov processes.

conduit ensuite à une preuve de convergence des processus de Markov à temps continu standards.

2D-I1: David Binder Memorial Session Séance commémorative David Binder

Organizer and Chair / Responsable et présidente: Mary Thompson (University of Waterloo)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: Hart House - South Dining

Abstracts/Résumés

ABDELLATIF DEMNATI (Statistics Canada), **RAO J.N.K.** (Carleton University)

[Tuesday May 27 / mardi 27 mai, 13:30]

Variance Estimation from Complex Survey Data Using Linearization Method: Impact of David Binder

Estimation de la variance par linéarisation à partir de données d'enquêtes complexes: influence de David Binder

David Binder's pioneering 1983 paper published in the *International Statistical Review* provided a unified linearization approach to variance estimation from complex survey data. This important paper stimulated much new research on linearization variance estimation in subsequent years. This talk will review some of that work, including the alternative linearization method of Demnati and Rao (2004, 2010). We apply our approach to the total variance of calibration estimators of the generalized linear models parameters when missing items have been imputed using either deterministic or random imputation. Simulation results on the performance of the Demnati-Rao variance estimator are also presented.

L'article novateur de David Binder publié en 1983 dans le *International Statistical Review* a présenté une approche unifiée de l'estimation de la variance par linéarisation à partir de données d'enquête complexes. Cet article important a fait souffler un vent nouveau sur la recherche sur l'estimation de la variance par linéarisation dans les années qui ont suivi. Nous passerons en revue certains de ces travaux, y compris la méthode de rechange par la linéarisation de Demnati et Rao (2004, 2010). Nous appliquons notre approche à la variance totale des estimateurs par calage pour les paramètres des modèles linéaires généralisées lorsque les réponses manquantes ont été imputées de façon déterministe ou aléatoire.

MILORAD KOVACEVIC (United Nations Development Programme)

[Tuesday May 27 / mardi 27 mai, 14:00]

Measuring Multidimensional Poverty and Inequality

Mesure multidimensionnelle de la pauvreté et de l'inégalité

Most studies of poverty are concerned with monetary poverty; however there is also a strong consensus that poverty has many facets that are not necessarily or easily expressed in dollars, such as poor health, lack of education, lack of access to basic services, etc. The multidimensional poverty index introduced in 2010 by the Human Development Report was designed to capture the overlapping deprivations that people face, reflecting both the incidence of multidimensional poverty, and its

La plupart des études sur la pauvreté ont trait à la pauvreté financière ; un fort consensus existe toutefois sur le fait que la pauvreté comporte bon nombre de facettes qui ne s'expriment pas nécessairement ou facilement en dollars, et notamment la mauvaise santé, le manque d'éducation ou d'accès à des services de base, etc. L'indice de pauvreté multidimensionnel élaboré en 2010 par le Rapport mondial sur le développement humain est conçu pour illustrer les privations qui se chevauchent, reflétant à la fois l'incidence et l'intensité de la pauvreté, notamment, dans le nombre de privations

2D-I1: David Binder Memorial Session Séance commémorative David Binder

intensity how many deprivations people experience at the same time. In a similar way, a distribution-sensitive multidimensional inequality index was introduced to assess inequality in several dimensions of well-being.

que les gens subissent simultanément. De la même façon, un indice d'inégalité multidimensionnel a été élaboré pour évaluer l'inégalité de plusieurs dimensions du bien-être en étant sensible à leur distribution.

NATALIE SHLOMO (University of Manchester), **RODOLPHE PRIAM** (University of Southampton)

[Tuesday May 27 / mardi 27 mai, 14:30]

Calibration of Small Area Estimates in Business Surveys

Calibrage des estimations pour les petits domaines dans les enquêtes auprès des entreprises

There is growing demand for disaggregated estimates for business statistics at small areas or domains. Small area estimation models however break down when applied to business statistics due to skewed distributions and outliers. Since many countries have business registers where covariate information is available, the nested-error unit-level small area model under the log-transformation will be applied. Generally, small area predictors do not add up to the design-based direct estimator of the total computed for a large (and planned) domain. Several methods for calibrating small area predictors have been proposed and we review and propose methods for calibration under the log-transformation. Empirical results based on simulated data from the unit-level model are presented. This is joint work with Rodolphe Priam.

La demande est croissante pour des estimations désagrégées pour des statistiques au près des entreprises au niveau de petits domaines. Toutefois, les modèles conçus à cette fin ne conviennent pas aux applications en gestion en raison de distributions asymétriques et de valeurs aberrantes. Comme plusieurs pays ont établi un registre d'entreprises comportant des covariables, un modèle pour les petits domaines avec erreur imbriquée au niveau de l'unité sera appliqué avec une transformation logarithmique. Règle générale, la somme des prédicteurs pour les petits domaines ne correspond pas à l'estimateur direct calculé pour un domaine large (et faisant parti du plan). Bon nombre de méthodes sont proposées pour calibrer des prédicteurs pour les petits domaines, et nous examinons et proposons des méthodes de calage utilisant une transformation logarithmique. Nous présentons des résultats empiriques basés sur des données simulées du modèle au niveau de l'unité. Ce travail a été réalisé conjointement avec Rodolphe Priam.

2D-I2: Gaussian Process Regression: Innovations and Applications Régression par processus gaussien : innovations et applications

Chair/Président: Mu Zhu (University of Waterloo)

Organizer/Responsable: Martin Lysy (University of Waterloo)

Room/Salle: Hart House - East Common

Abstracts/Résumés

DEREK BINGHAM (Simon Fraser University), **ROBERT GRAMACY** (Booth School of Business, The University of Chicago)

[Tuesday May 27 / mardi 27 mai, 13:30]

Fast Gaussian Process Calibration of Computer Models with Large Ensembles

Calage rapide de modèles de processus gaussiens présentant de grands ensembles

Statistical model calibration of computer models is commonly done in a wide variety of scientific endeavours. In the end, this exercise amounts to solving an inverse problem and a form of regression. Gaussian process model are very convenient in this setting as non-parametric regression estimators and provide sensible inference properties. However, when the data structures are large, fitting the model becomes difficult. In this work, new methodology for calibrating large computer experiments is presented. We proposed to perform the calibration exercise by modularizing a hierarchical statistical model with approximate emulation via local Gaussian processes. The approach is motivated by an application to radiative shock hydrodynamics.

Le calage de modèles informatiques est courante dans un large éventail d'initiatives scientifiques. Cet exercice permet en fait de résoudre un problème inverse et une forme de régression. Les modèles de processus gaussiens sont très utiles dans ce contexte à titre d'estimateurs de régression non paramétriques et fournissent des propriétés d'inférence raisonnables. Cependant, il devient difficile d'ajuster le modèle lorsque les structures de données sont grandes. Dans cet article, nous présentons une nouvelle méthodologie pour le calage d'expériences informatiques massives. Nous proposons d'effectuer le calage en modulant un modèle statistique hiérarchique par émulation approximative au moyen de processus gaussiens locaux. Notre approche est motivée par une application à l'hydrodynamique de chocs radiatifs.

JASON LOEPPKY (University of British Columbia), **ERIK ROSOLOWSKY** (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 14:00]

Comparing Simulated Emission from Molecular Clouds using Gaussian Process Regression

Comparaison d'émissions simulées de nuages moléculaires à l'aide de la régression de processus gaussiens

The process of star formation is the fundamental agent at determining how a galaxy evolves over the course of the Universe. Decades of observation of star-forming regions suggest that several different physical effects shape the star formation process: gravitation, magnetism, chemistry, and radiation. However, it is not understood exactly how these effects shape the star formation process. Large-scale, multi-physics computer simulations are routinely used to model the process of star formation.

Le processus de formation des étoiles constitue l'élément fondamental permettant d'évaluer l'évolution d'une galaxie dans l'univers. Des décennies d'observations de régions propices à la formation d'étoiles laissent croire qu'une foule d'effets physiques influencent le processus de formation des étoiles, y compris la gravitation, le magnétisme, la chimie et la radiation. Cependant, le rôle précis de ces effets sur le processus en question n'est pas connu avec précision. Des simulations multi-physiques à grande échelle sont utilisées couramment pour modéliser ce processus. Ces simulations

2D-I2: Gaussian Process Regression: Innovations and Applications Régression par processus gaussien : innovations et applications

The simulations are computationally expensive and produce a three dimensional model that can only be compared statistically to observations. In this talk we will explore the use of a GP regression as a tool for providing a sensitivity analysis of various metrics used to compare astronomical data sets.

nécessitent une grande puissance de calcul et produisent un modèle tridimensionnel qui peut être comparé aux observations uniquement d'un point de vue statistique. Dans cet exposé, nous étudions l'utilisation d'une régression de processus gaussiens comme un outil d'analyse de sensibilité de différentes mesures utilisées dans la comparaison de jeux de données astronomiques.

MARTIN LYSY & ETHAN FANG (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 14:30]

Parameter Inference for Diffusion Processes via Gaussian Process Regression

Inférence sur les paramètres de processus de diffusion à l'aide de la régression de processus gaussiens

Diffusion processes are used to model a wealth of stochastic phenomena in the natural sciences, engineering, and finance. For many such processes, the likelihood function is only available for continuous-time data. Since actual data recordings are discrete, parameter inference is typically achieved by integrating over the missing paths between observations. However, most Markov chain Monte Carlo algorithms used to this end impose a formidable computational burden. Here, we propose an Importance Sampling approach to the missing path imputation via Gaussian Process Regression. To account for the highly non-Gaussian nature of these paths, both model-free and model-dependent features are included in the variance function. The methodology is illustrated with several financial and biological examples.

Les processus de diffusion sont utilisés pour modéliser une panoplie de phénomènes stochastiques en sciences naturelles, en génie et en finance. Pour bon nombre de ces processus, la fonction de vraisemblance est disponible uniquement pour des données continues dans le temps. Comme les enregistrements de données réelles sont discrets, l'inférence sur les paramètres est généralement obtenue en procédant à l'intégration sur les trajectoires manquantes entre les observations. Cependant, la plupart des algorithmes de Monte Carlo par chaîne de Markov utilisés à cette fin imposent des calculs extrêmement lourds. Nous proposons une approche d'échantillonnage par importance pour l'imputation des trajectoires manquantes par le biais de la régression de processus gaussiens. Pour tenir compte de la nature hautement non gaussienne de ces trajectoires, les caractéristiques dépendantes et non dépendantes d'un modèle sont incluses dans la fonction de la variance. La méthodologie est illustrée par plusieurs exemples en finance et en biologie.

2D-I3: Graduate Student Presentations in Actuarial Science Présentations d'étudiants de cycle supérieur en science actuarielle

Organizer and Chair / Responsable et président: Jiandong Ren (Western University)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe de science actuarielle

Room/Salle: Hart House - Debates Rm

Abstracts/Résumés

ANAS ABDALLAH (Université Laval), **JEAN-PHILIPPE BOUCHER** (UQAM), **HÉLÈNE COSSETTE** (Université Laval)
[Tuesday May 27 / mardi 27 mai, 13:30]

Modeling Dependence Between Loss Triangles Using Hierarchical Archimedean Copulas

Modélisation de la dépendance entre triangles de développement à l'aide des copules archimédiennes hiérarchiques

One of the most critical problems in casualty property insurance is to determine an appropriate reserve for the incurred but unpaid losses. The provisions are the largest part of the liabilities of a non-life insurer. The global provisions are often determined under the independence assumption. Firstly, we suppose a dependence between all the observations that belong to the same calendar year for all lines of business using multivariate Archimedean copulas. Secondly, we suppose another dependence structure that links the calendar years of different lines of business by using hierarchical Archimedean copulas. When applied to data, our models provide a better fit than existing models, and offer a better and more realistic interpretation of the dependence between the lines of business.

L'une des tâches les plus critiques en assurance de dommages est de déterminer une réserve appropriée pour l'ensemble du portefeuille. La plupart des techniques se basent sur des segmentations en sous-portefeuilles homogènes. Les provisions constituent un élément majeur des états financiers d'une compagnie d'assurance et la volatilité reliée au montant des provisions totales de l'ensemble des engagements ne peut être ignorée. Pour modéliser cette dépendance, nous utilisons la classe des copules Archimédiennes hiérarchiques, qui généralise la famille des copules archimédiennes en introduisant une plus riche structure de corrélation que les modèles existants. Notre modèle nous permet de capter la dépendance de façon plus réaliste et flexible en imbriquant une notion de niveau et de hiérarchie entre les différentes lignes d'affaires.

FANGDA LIU, JUN CAI & CHRISTIANE LEMIEUX (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 13:45]

Optimal Reinsurance with Regulatory Initial Capitals and Default Risks

Réassurance optimale grâce à une réglementation en matière de capital initial et de risque de défaut de paiement

In a reinsurance contract, a reinsurer promises to pay the part of the loss faced by an insurer in exchange of receiving a reinsurance premium. However, when the promised indemnity exceeds the total of the reinsurance premium and the reinsurer's initial capital which is determined by the value-at-risk (VaR) of its promised indemnity, the reinsurer may fail to pay the promised amount or default may

Dans un contrat de réassurance, un réassureur promet de payer la partie de la perte d'un assureur et reçoit en échange une prime de réassurance. Par contre, lorsque l'indemnité promise excède le total de la prime de réassurance et du capital initial du réassureur qui est déterminé par la valeur à risque (VaR) de l'indemnité promise, il est possible que le réassureur ne paie pas l'indemnité promise et il peut y avoir défaut de paiement. Nous examinons, du point de vue de

2D-I3: Graduate Student Presentations in Actuarial Science Présentations d'étudiants de cycle supérieur en science actuarielle

occur. In the presence of regulatory initial capitals and counterparty default risks, we investigate optimal reinsurance designs from an insurer's point of view and derive optimal reinsurance strategies that maximize the expected utility of an insurer's terminal wealth or minimize the VaR of an insurer's total retained risk.

l'assureur et en présence d'une réglementation en matière de capital initial et de risque de la contrepartie, les plans optimaux de réassurance sur la base desquels nous établissons les stratégies optimales de réassurance qui maximisent l'espérance d'utilité de la richesse résiduelle de l'assureur, ou minimisent la VaR du risque conservé par l'assureur.

HASSAN OMIDI (Université de Montréal), **ZIED BEN-SALAH** (Concordia University), **HELENE GUERIN** (University of Rennes), **MANUEL MORALES** (University of Montreal)

[Tuesday May 27 / mardi 27 mai, 14:00]

On the Depletion Problem for an Insurance Risk Process: New Non-ruin Quantities in Collective Risk Theory

À propos du problème d'épuisement pour un processus de risque d'assurance : de nouvelles données quantitatives de non ruine en matière de théorie du risque collectif

The field of risk theory has traditionally focused on ruin-related quantities. Although interesting in their own right, ruin related quantities do not seem to capture path-dependent properties of the reserve. In this presentation we aim at presenting the probabilistic properties of drawdowns and the speed at which an insurance reserve depletes as a consequence of the risk exposure of the company. This type of quantities has never been proposed before as measures of riskiness in insurance. We derive expressions for the distribution of drawdowns and the Laplace transform for the speed of depletion. These expressions are given for some examples of Levy insurance risk processes for which they can be calculated, in particular for the classical Cramer-Lundberg model.

En général, le domaine de la théorie du risque met l'accent sur les données quantitatives relatives à la ruine. Même si ces dernières présentent un intérêt, elles ne capturent pas la dépendance aux trajectoires de la réserve. Dans cet exposé, nous présentons les propriétés probabilistes de prélèvement et le rythme auquel la réserve d'assurance s'épuise en conséquence de l'exposition au risque de l'entreprise. Ce type de données quantitatives n'a jamais été proposé auparavant en tant que mesure de risque d'assurance. Nous avons dérivé des expressions pour la distribution des prélèvements ainsi que la transformée de Laplace du rythme d'épuisement. Ces expressions sont fournies pour quelques exemples de processus de Lévy en matière de risque d'assurance pour lesquels elles peuvent être calculées, notamment pour le modèle classique de Cramer-Lundberg.

OSCAR QUIJANO & JOSÉ GARRIDO (Concordia University)

[Tuesday May 27 / mardi 27 mai, 14:15]

Linear Credibility for GLM Losses

Modèles linéaires généralisés avec crédibilité linéaire pour les pertes

We consider actuarial credibility theory for expected losses estimated with Generalized Linear Models (GLMs). We use a Bayesian approach and appropriate priors for the regression coefficients that lead to Linear Credibility. These conjugate priors are then used to find confidence intervals for the regression coefficients. Illustrative examples will also be discussed.

Nous analysons la théorie de la crédibilité actuarielle pour l'espérance des pertes estimée avec des modèles linéaires généralisés. Nous utilisons une approche bayésienne et des lois a priori appropriées pour les coefficients de régression qui mènent à une crédibilité linéaire. Ces lois a priori conjuguées sont ensuite utilisées pour trouver des intervalles de confiance pour les coefficients de régression. Nous présentons enfin des exemples pour illustrer.

2D-I3: Graduate Student Presentations in Actuarial Science Présentations d'étudiants de cycle supérieur en science actuarielle

DAMENG TANG, X. SHELDON LIN, ANDREI L. BADESCU & LAN GONG (University of Toronto)

[Tuesday May 27 / mardi 27 mai, 14:30]

Fitting Erlang-Based Mixture Models to Operational Loss Data

Ajustement de modèles de mélanges d'Erlang à des données de perte opérationnelle

Modeling and quantifying the operational risk are now required by the Basel Committee on Banking Supervision and become a critical part of the risk management of a financial institution. A general quantitative approach is the so-called Loss Distribution Approach (LDA). In this paper, we propose an Erlang-based mixture model approach in which the frequencies among the units are dependent and modeled via a multivariate mixed Poisson process, while the severity in each unit is modeled by a mixed Erlang model. Under our approach there is no need to use a copula and the use of Monte Carlo simulation is minimal. Using real operational loss data from the Operational Riskdata eXchange (ORX), we demonstrate the efficiency and accuracy of the approach.

Le Comité de Bâle sur le contrôle bancaire exige désormais la modélisation et la quantification du risque opérationnel, qui deviennent des éléments essentiels de la gestion du risque d'une institution financière. L'approche par distribution des pertes (LDA) est la méthode quantitative classique. Dans cet article, nous proposons un modèle de mélanges fondé sur la formule d'Erlang dans lequel les fréquences au sein des unités sont dépendantes et modélisées selon un processus de Poisson multivarié composé, alors que la sévérité dans chaque unité est modélisée au moyen d'un modèle d'Erlang composé. Dans le cadre de notre approche, aucune copule n'est nécessaire et l'utilisation de simulation de Monte Carlo est minimale. Nous démontrons l'efficacité et la précision de notre méthode à l'aide de données de perte réelles provenant de l'Operational Riskdata eXchange (ORX) Association.

CHEN YANG & KRISTINA SENDOVA (Western University)

[Tuesday May 27 / mardi 27 mai, 14:45]

On the Threshold Strategy for Dividend Payments Under the Dual Model Perturbed by Diffusion

À propos de la stratégie de seuil pour les versements de dividendes selon le modèle dual perturbé par la diffusion

In this paper, we revisit the threshold dividend payment strategy of the dual risk model with a surplus process perturbed by a Brownian motion. We study the relevant Gerber-Shiu functions and obtain explicit expressions of the expected discounted dividends until ruin in terms of the δ -scale functions. Finally, we illustrate our results by deriving a more specific explicit expression for the expected discounted dividends when the jump size distribution is from the K_n family.

Dans cet article, nous réexaminons la stratégie de versements par dividendes dite de seuil du modèle de risque dual dont le processus de surplus est perturbé par un mouvement brownien. Nous examinons les fonctions Gerber-Shiu pertinentes et nous obtenons des expressions explicites des dividendes actualisés prévus jusqu'à la ruine dans le contexte de fonctions δ . Nous illustrons aussi nos résultats par l'obtention d'une expression explicite particulière des dividendes actualisés prévus lorsque la distribution de la taille des chocs provient de la famille K_n .

2D-I4: New Methods for Analyzing Brain Imaging data Nouvelles méthodes d'analyse des données d'imagerie du cerveau

Chair/Présidente: Michelle Zhou (Simon Fraser University)
Organizer/Responsable: Ivor Cribben (University of Alberta)

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

IVOR CRIBBEN (University of Alberta), **CHRISTIAN HABECK** (Columbia University Medical Center)
[Tuesday May 27 / mardi 27 mai, 13:30]

A New Method for Jointly Estimating Dynamic Functional Connectivity

Nouvelle méthode pour l'estimation conjointe de la connectivité fonctionnelle dynamique

Previous investigations of fMRI data focused on static descriptions of functional connectivity (FC) between a set of brain regions. However, recently, the emphasis has shifted to estimating dynamic FC (or dynamic networks) between brain regions in both task-related and resting-state experiments. In an fMRI experiment, a major challenge is estimating a group-level dynamic FC structure because of the variability between subjects. In this work, we introduce a new technique for estimating dynamic FC. The technique firstly estimates FC change points on the group-level by borrowing strength across subjects and then estimates a graphical model for each subject, that share certain characteristics, between each pair of change points. The new method is applied to simulated and resting-state fMRI data sets.

Les études antérieures sur des données d'IRMf portaient sur des descriptions statiques de la connectivité fonctionnelle (CF) entre un ensemble de régions cérébrales. Cependant, l'intérêt s'est tourné vers l'estimation de la CF dynamique (ou des réseaux dynamiques) entre des régions cérébrales dans des expériences sur un état associé à une tâche et au repos. Dans une expérience d'IRMf, l'estimation de la structure de la CF dynamique au niveau du groupe constitue un grand défi en raison de la variabilité entre les sujets. Dans ce travail, nous présentons une nouvelle technique pour estimer la CF dynamique. La technique estime d'abord les points de rupture de la CF au niveau du groupe par l'emprunt d'information à tous les sujets, puis estime un modèle graphique pour chaque sujet partageant certaines caractéristiques, entre chaque paire de points de rupture. La méthode proposée est appliquée à des jeux de données simulées et d'IRMf au repos.

LINGLONG KONG (University of Alberta), **HONGTU ZHU** & **HONGYU AN** (University of North Carolina at Chapel Hill),
L. ANDRIA FORD (Washington University at St. Louis)

[Tuesday May 27 / mardi 27 mai, 14:00]

Simultaneously Sparse Partial Functional Linear Regression in Hyper-Acute Ischemic Stroke Study

Régression linéaire fonctionnelle partielle simultanément éparsée dans une étude sur l'accident ischémique cérébral hyper-aiguë

In partial functional linear regression, it is common to approximate the functional coefficients by function basis with various sparsely estimating methods. However, it may happen that several functional coefficients share the same sparse pattern. We propose to use group adaptive LASSO to simultaneously estimate the sparse functional coefficients. We use a perturbation bootstrap method to

En régression linéaire fonctionnelle partielle, il est courant d'approximer les coefficients fonctionnels par des bases de fonctions avec diverses méthodes d'estimation éparsée. Cependant, il est possible que plusieurs coefficients fonctionnels présentent le même motif d'éparpillement. Nous proposons d'utiliser le LASSO adaptatif en groupe pour estimer simultanément les coefficients fonctionnels éparsés. Nous utilisons une méthode de bootstrap par perturbation

2D-I4: New Methods for Analyzing Brain Imaging data Nouvelles méthodes d'analyse des données d'imagerie du cerveau

obtain the p values for each function basis. Simulation studies show that our proposed method outperforms others. We apply it to a real data analysis on Hyper-acute ischemic stroke study. The results provide some new insights in addition to confirming some previous findings.

pour obtenir la p -value de chaque base de fonctions. Des études de simulation montrent que la méthode proposée surpasse d'autres approches. Nous l'appliquons à une analyse de données réelles provenant d'une étude sur l'accident ischémique cérébral hyper-aiguë. Les résultats apportent de nouvelles connaissances en plus de confirmer des conclusions antérieures.

FAROUK NATHOO & ARIF BABUL (University of Victoria), **ALEXANDER MOISEEV** (Simon Fraser University), **NAZNIN VIRJI-BABUL** (University of British Columbia), **MIRSA FAISAL BEG** (Simon Fraser University)

[Tuesday May 27 / mardi 27 mai, 14:30]

Variational Bayes High-Dimensional Discrete-Time Dynamical Systems for Combined MEG and EEG Data
Systèmes bayésiens dynamiques variationnels de grande dimension à temps discret pour des données combinées de MEG et d'EEG

We present an approach for electromagnetic brain mapping using combined MEG and EEG data. We consider the inverse problem of reconstructing neural activity at a large number of locations within the brain, from time series measuring the electromagnetic field at a small number of sensors on the scalp. A system of underdetermined dynamic models is used for relating the data to the underlying non-Gaussian dynamics, which exhibit spatial dependence as well as bilateral dependence on opposite brain hemispheres. With a large number of brain locations and time points, the resulting model is parameterized over a high-dimensional space, making Bayesian inference based on MCMC infeasible. As an alternative, we propose a variational Bayes approximation and study the quality of this approximation.

Nous présentons une approche de cartographie cérébrale électromagnétique qui utilise des données combinées de MEG et d'EEG. Nous examinons le problème inverse de la reconstruction de l'activité neurale d'un grand nombre d'emplacements dans le cerveau à partir de séries chronologiques mesurant le champ électromagnétique d'un petit nombre de capteurs sur le cuir chevelu. Nous utilisons un système de modèles dynamiques sous-déterminés pour lier les données aux processus dynamiques non gaussiens, qui présentent une dépendance spatiale et une dépendance bilatérale dans les hémisphères cérébraux opposés. Avec un grand nombre d'emplacements cérébraux et de points dans le temps, le modèle obtenu est paramétré sur un espace de grande dimension, rendant impossible l'inférence bayésienne fondée sur les MCMC. À titre de solution de rechange, nous proposons une approximation bayésienne variationnelle et nous évaluons la qualité de cette approximation.

2D-I5: Statistical Challenges in High-Throughput Genomic Data Problèmes statistiques des données génomiques à débit élevé

Organizer and Chair / Responsable et président: Kun Liang (University of Waterloo)

Room/Salle: UC - 140

Abstracts/Résumés

HONGKAI JI (Johns Hopkins University), **BING HE & FANG DU** (Johns Hopkins Bloomberg School of Public Health)
[Tuesday May 27 / mardi 27 mai, 13:30]

A Model-based Approach for Characterizing Correlation in Big Genomic Data

Approche pour la caractérisation des corrélations données génomiques volumineuses basée sur un modèle

Projects such as ENCODE have created high-throughput genomic data for many different cell types and experimental conditions. Many fundamental data analysis tasks such as clustering or prediction require one to characterize correlation or similarity between two genomic features based on data collected from different cell types and conditions. In big data settings such as ENCODE, data are intrinsically heterogeneous as they are generated by different laboratories and possibly using different protocols. This heterogeneity can have tremendous negative impact on the analysis of correlation. Here we present a method that couples the correlation analysis with a model that removes the unwanted technical variation in the big genomic data. We demonstrate that this approach improves big data clustering and prediction.

Des projets du type d'ENCODE ont produit des données génomiques à très haut débit pour différents types de cellules et de conditions expérimentales. Pour bon nombre de tâches d'analyse de données fondamentales, notamment l'analyse en grappe ou la prévision, il faut caractériser la corrélation ou la similarité entre deux caractéristiques génomiques en fonction des données recueillies sur les différents types de cellules et de conditions. Dans un vaste contexte comme le projet ENCODE, l'hétérogénéité des données est intrinsèque, étant donné qu'elles sont générées par des laboratoires différents et possiblement en fonction de protocoles distincts, ce qui peut avoir un impact négatif substantiel sur l'analyse des corrélations. Nous présentons une méthode qui combine l'analyse des corrélations avec un modèle qui élimine les variations techniques indésirables dans les données génomiques massives. Nous démontrons que cette approche améliore l'analyse de regroupement et la prévision.

LEI SUN & ANDRIY DERKACH (University of Toronto), **JERRY LAWLESS** (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 14:00]

Methods for Association Analysis of Rare Variants: A Review and Some New Developments

Méthodes d'analyse d'association de variantes rares : bilan et nouveaux développements

The recent focus on rare variants has produced many testing strategies to assess association between a group of rare variants and a trait, with competing claims about the performance of various tests. We review frequently used tests and show that they fall into either linear or quadratic classes, and neither class consistently outperforms the other across genetic models. This understanding leads to development of robust tests that combine the two complementary classes. However, re-

L'attention portée récemment aux variantes rares a produit plusieurs stratégies de test pour évaluer l'association entre un groupe de variantes rares et un trait, avec des allégations concurrentes sur la performance de divers tests. Nous passons en revue des tests courants et nous montrons qu'ils appartiennent à des classes linéaires ou quadratiques, et qu'aucune de ces classes n'a surpassé l'autre de façon constante dans les modèles génétiques. Cette constatation nous oriente vers le développement de tests robustes qui combinent les deux classes complémentaires.

2D-15: Statistical Challenges in High-Throughput Genomic Data Problèmes statistiques des données génomiques à débit élevé

Regardless of the specific test used, power is generally low in current realistic settings due to various factors including rareness of the key variants. To increase power, we investigate various cost-effective response-dependent sampling strategies. This is joint work with graduate student Andriy Derkach and Professor Jerry Lawless.

Toutefois, quels que soient les tests spécifiques utilisés, la puissance est assez faible dans les contextes statistiques courants en raison de divers facteurs, notamment la rareté des variantes clés. Pour améliorer la puissance, nous investiguons des stratégies économiquement efficaces d'échantillonnage de données dépendantes de la réponse. Ce travail a été réalisé conjointement avec l'étudiant diplômé Andriy Derkach et le professeur Jerry Lawless.

WEI SUN (University of North Carolina at Chapel Hill)

[Tuesday May 27 / mardi 27 mai, 14:30]

A Study of RNA-seq Data in Cancer Patients

Étude des données de séquençage d'ARN des patients cancéreux

I will present a systematic study of RNA-seq data and its genetic architecture in 550 breast cancer patients from The Cancer Genome Atlas project. eQTL mapping of gene expression (measured by RNA-seq in tumor tissue) vs. germline genotype and tumor copy number aberrations show that both types of genetic variants have substantial influence on gene expression. I will further assess such associations after deconvoluting gene expression from tumor cells and normal cells (e.g. stromal cells within the tumor tissue), and discuss possible scenarios to use such eQTL results to obtain further biological insights.

Je présente une étude systématique des données de séquençage d'ARN et de son architecture génétique chez 550 patientes atteintes du cancer du sein dans le cadre du projet de répertoire génomique du cancer (The Cancer Genome Atlas). Une cartographie LCQ (eQTL) des gènes (mesurés par séquençage d'ARN dans les tissus tumoraux) comparée au génotype germline et aux variations aberrantes du nombre de copies a révélé que les deux types de variantes génétiques ont une influence substantielle sur l'expression génétique. Après une déconvolution de l'expression génétique des cellules tumorales et normales, (i.e. des cellules stromatolithiques à l'intérieur du tissu tumoral), j'évalue de telles associations et discute des scénarios possibles d'utilisation des résultats d'eQTL afin d'approfondir les connaissances biologiques.

2D-I6: Weakly Interacting Processes in Finance and Economics Processus à faible interaction en finance et économie

Organizer and Chair / Responsable et président: Bruno Remillard (HEC Montreal)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Room/Salle: UC - 179

Abstracts/Résumés

CHRISTOPH FREI (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 13:30]

Financial Markets with Interacting Investors and Their Relation to Backward SDEs

Marchés financiers en présence d'investisseurs en interaction et leur relation à l'égard d'équations différentielles stochastiques (EDS) rétrogrades

We consider a model of a financial market where investors take not only their own absolute performance, but also the relative performance compared to their peers into account. We are interested in equilibria where every investor can find an individually optimal strategy. We relate this problem to the study of backward stochastic differential equations (SDEs). Introducing a new notion of local solution to backward SDEs, we show that the backward SDE related to our financial market is locally but not globally solvable. From this, we deduce that there exist local but no global equilibria in our model of a financial market. By considering the relative performance, investors may ruin each other so that equilibria exist only over a short time.

Nous examinons un modèle de marché financier où les investisseurs prennent en considération à la fois leur performance absolue et leur performance relative par rapport à leurs pairs. Nous nous intéressons aux divers équilibres qui permettent à chaque investisseur d'obtenir une stratégie optimale d'un point de vue individuel. Nous établissons un parallèle entre ce problème et l'étude d'équations différentielles stochastiques (EDS) rétrogrades. Nous présentons une nouvelle notion de solution locale aux EDS rétrogrades et nous montrons que l'EDS rétrograde liée à notre marché financier est soluble localement sans l'être globalement. De ce fait, nous déduisons qu'il existe des équilibres locaux mais non globaux dans notre modèle de marché financier. En considérant la performance relative, les investisseurs peuvent se ruiner mutuellement, ce qui fait en sorte que les équilibres existent seulement à court terme.

MINYI HUANG (Carleton University)

[Tuesday May 27 / mardi 27 mai, 14:00]

Mean Field Capital Accumulation Optimization and Decentralized Strategies

Stratégies décentralisées et optimisation d'accumulation de domaine de capital moyen

This work presents neoclassical stochastic growth optimization with (i) a large number of agents and (ii) a mean field congestion effect in the production dynamics. The stochastic control problem generalizes the classical one sector optimization framework which dates back to the work of Brock and Mirman (1972) and Merton (1975). We formulate a mean field game and decentralized strategies by consistent mean field approximations.

Ce travail présente une optimisation néoclassique de croissance stochastique avec (i) un grand nombre d'agents et (ii) un effet moyen de congestion du domaine dans la dynamique de production. Le problème de contrôle stochastique généralise le cadre classique d'optimisation à un secteur, qui remonte aux travaux de Brock et Mirman (1972) et Merton (1975). Nous formulons un jeu de domaine moyen et des stratégies décentralisées par des approximations moyennes et cohérentes du domaine.

2D-I6: Weakly Interacting Processes in Finance and Economics Processus à faible interaction en finance et économie

MICHAEL KOURITZIN (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 14:30]

Resampled Particle Filtering by McKean-Vlasov Thinking

Tri de particules ré-échantillonnées par l'approche McKean-Vlasov

Particle filters (also known as sequential Monte Carlo methods) have been around since the 1960s. However, their popularity surged in the 1990s when resampling procedures were introduced to prune useless particles and thereby reduce computations dramatically. Unfortunately, the convergence analysis became far more difficult. We will demonstrate the process of designing an effective particle filter, starting with the classical weighted particle filter, moving through an unrealizable McKean-Vlasov filter and arriving at a computer-ready, weakly-interacting particle filter. Both algorithms and basic convergence results will be given. We will also motivate the filtering problem so expertise in filtering is not required. Proof sketches will be given to the degree that time allows.

Les tris de particules (aussi connu sous le nom de méthodes séquentielles de Monte Carlo) sont connues depuis les années 60. Par contre, leur popularité a bondi dans les années 90 lorsque les procédures de rééchantillonnage ont été introduites afin de retirer les particules inutiles et ainsi grandement réduire les calculs. Malheureusement, les analyses de convergence sont devenues beaucoup plus difficiles. Nous allons décrire le processus pour concevoir un filtre de particules efficace, en commençant avec le filtre de particules pondéré classique, puis allant vers un filtre McKean-Vlasov irréalisable, afin d'arriver à un filtre de particule avec interactions faibles, prêt à être encodé sur l'ordinateur. Les deux algorithmes et les résultats de convergence de base seront fournis. Nous allons aussi motiver le problème de tri de manière à ce qu'une expertise en tri ne soit pas requise. Si le temps le permet, des ébauches de preuves seront présentées.

2D-C1: Biostatistics - Applications 3 Biostatistique - applications 3

Chair/Présidente: Yeying Zhu (University of Waterloo)

Room/Salle: UC - 52

Abstracts/Résumés

CAITLIN DALY & JEMILA HAMID (McMaster University)

[Tuesday May 27 / mardi 27 mai, 13:30]

Establishing Paediatric Reference Intervals: Adjusting for Complex Data

Établir des intervalles de référence pédiatriques : ajuster pour les données complexes

Reference intervals (RIs) are one of the most crucial tools in laboratory and diagnostic medicine. The importance of establishing paediatric RIs has gained attention over recent years. However, the dynamic nature of paediatric data complicates statistical analysis. No standard has been set for researchers to follow worldwide, thus use of different statistical methodologies may produce varying and unreliable RIs. A simulation study was conducted to determine the impact various statistical methods have on resulting RIs in a variety of circumstances. Clear recommendations are given, accompanied with direction to underutilized approaches, to assist researchers in constructing RIs through a systematic approach.

Les intervalles de références (IR) sont un des outils les plus essentiels de la médecine diagnostique et de laboratoire. L'importance d'établir des IR pédiatriques a retenu l'attention ces dernières années. Par contre, la nature dynamique des données pédiatriques compliquent l'analyse statistique. Aucun standard international à respecter n'a encore été établi pour les chercheurs, ainsi l'utilisation de différentes méthodologies statistiques peut produire différents IR peu fiables. Une étude de simulation a été réalisée pour déterminer l'impact que différentes méthodes statistiques ont sur les IR résultants dans diverses circonstances. Des recommandations claires sont fournies, accompagnées d'instructions pour les approches sous-utilisées, pour aider les chercheurs dans la construction d'IR selon une démarche systématique.

SHAHEDUL KHAN (University of Saskatchewan)

[Tuesday May 27 / mardi 27 mai, 13:45]

A Multistate Modeling Approach to Study the Effects of Air Pollutants on Recurrent Hospital Admissions Due to Respiratory Diseases

Approche de modélisation multi-états pour l'étude des effets des polluants atmosphériques sur les admissions répétitives à l'hôpital pour cause de maladie respiratoire

Adverse effects in the respiratory system due to air pollutants are commonly observed across the world. Traditional approaches to analyze such data do not take into account the possibility of the increased risk of the recurrence of a disease due to (1) the history of the disease itself, and (2) the occurrences of other diseases that might have positive effects for the recurrence of the disease. We propose a multistate modelling framework for multitype recurrent hospital admissions to characterize possible paths for event occurrences; the pri-

Partout dans le monde, des effets néfastes sur le système respiratoire résultant des polluants atmosphériques sont observés. Des approches traditionnelles d'analyse de telles données ne tiennent pas compte de la possibilité d'un risque accru de récurrence de la maladie en raison (1) des antécédents de maladie, et (2) des occurrences d'autres maladies pouvant avoir des effets positifs sur la récurrence de la maladie. Nous proposons un cadre de modélisation multi-états pour des admissions répétitives à l'hôpital de types multiples afin de caractériser des trajectoires possibles d'occurrence d'événements ; l'objectif principal consiste à estimer les ef-

2D-C1: Biostatistics - Applications 3

Biostatistique - applications 3

mary objective is to estimate the effects of air pollutants on respiratory diseases. We illustrate our approach with an application to recurrent hospital admissions due to chronic obstructive pulmonary disease, asthma and pneumonia in Saskatchewan.

fets des polluants atmosphériques sur les maladies respiratoires. Nous illustrons notre approche avec une application portant sur les admissions répétitives à l'hôpital en raison de maladies pulmonaires obstructives chroniques, d'asthme et de pneumonie en Saskatchewan.

FARHOOD MOHAMMADI & KEUMHEE CARRIERE CHOUGH (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 14:00]

Comparison of Missing data Methods for Repeated Measures Binary Data

Comparaison des méthodes pour données manquantes appliquées aux données binaires de mesures répétées

Missing data is common in experiments involving human subjects. Using only the complete cases for analysis may cause biased inferences. Assuming that the data are missing at random (MAR), various methods have been developed to handle missing data. Among them, Generalized Estimating Equations(GEE), Non-linear mixed effects(NLME) and multiple-imputation based GEE(MI-GEE) are considered the most efficient methods. However, these guidelines are too limited to apply generally. We evaluated their performances with repeated measurement binary data using a simulation study under various missing data mechanisms. Based on empirical size control and power level, we conclude that NLME method performs the best.

Dans les expériences qui font appel à des sujets humains, les données manquantes sont courantes. Une restriction de l'analyse aux seuls cas complets pourrait entraîner une inférence biaisée. En présumant que les données sont manquantes complètement au hasard (MCAH), diverses méthodes ont été conçues pour les traiter. Parmi ces dernières, les équations d'estimation généralisées (EEG), les modèles non linéaires à effets mixtes (NLEM) et les EEG à imputations multiples (EEG-IM) comptent parmi les méthodes les plus efficaces. Par contre, ces directives sont trop restreintes pour une application générale. Nous avons évalué leur niveau de performance avec des données binaires de mesures répétées en utilisant une analyse de simulation en fonction de divers mécanismes de données manquantes. En fonction du niveau de contrôle et de puissance empirique, nos conclusions révèlent que la méthode NLEM obtient les meilleurs résultats.

THIERRY CHEKOUO TEKOUANG & FRANCESCO STINGO (UT MD Anderson Cancer Center), **JAMES DOECKE** (CSIRO Computational Informatics, Australia), **KIM-ANH DO** (UT MD Anderson Cancer Center)

[Tuesday May 27 / mardi 27 mai, 14:15]

A Bayesian Approach to Biomarker Selection Through miRNA Regulatory Network with Application to Kidney Cancer

Une approche bayésienne de la sélection des biomarqueurs par un réseau de régulation microARN avec application au cancer du rein

The availability of cross-platform, large-scale genomic data has enabled the investigation of complex biological relationships for many cancers. MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression; however, the direct relationship between a microRNA and its target gene is difficult to measure. We propose a novel Bayesian model to identify microRNAs and their target genes, which are associated with survival time, by incorporating the microRNA regulatory network through prior distributions. We as-

La disponibilité de données génomiques multiplateforme à grande échelle a permis l'étude de relations biologiques complexes pour de nombreux cancers. Les microARN sont de petits ARN non codants qui régulent l'expression génique ; toutefois, la relation directe entre un microARN et son gène cible est difficile à mesurer. Nous proposons un nouveau modèle bayésien pour identifier les microARN et leurs gènes cibles qui sont associés à un temps de survie en incluant le réseau de régulation microARN via ses distributions antérieures. Nous partons du principe que les biomarqueurs impliqués dans les réseaux de régulation sont prob-

2D-C1: Biostatistics - Applications 3

Biostatistique - applications 3

sume that biomarkers involved in regulatory networks are likely associated with survival time. Using simulation studies, we assess the performance of our method, and apply it to experimental data of kidney renal clear cell carcinoma (KIRC) obtained from The Cancer Genome Atlas.

ablement associés au temps de survie. À l'aide d'études par simulation, nous évaluons la performance de notre méthode et l'appliquons à des données expérimentales sur le cancer du rein et l'hypermérome obtenues du Cancer Genome Atlas.

SHABNAM VATANPOUR, FARZANA YASMIN & XIAOMING WANG (University of Alberta), **SAUMYADIPTA PYNE** (CR Rao Advanced Institute of Mathematics, Statistics and Computer Science/Public Health Foundation of India), **IRINA DINU** (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 14:30]

Linear Combination Test and Reduction in Microarray Studies

Test des combinaisons linéaires et de réduction dans les études de microréseau

Gene set analysis identifies sets of genes associated with outcomes in microarray gene expression studies. Many methods have been proposed for a binary outcome, but only a few for continuous outcomes. Two challenges consist of: 1) having a large number of genes in a set and a small sample size, referred as the small n , large p problem, and 2) accommodating correlations between genes across a set. We propose a new method for reducing gene sets associated with a continuous outcome to their core members. This method is based on the linear combination test for gene sets, which incorporates the gene-expression covariance matrix into the test statistic, via a shrinkage estimation approach. The method is tested in a real microarray study of prostate cancer patients.

L'analyse des ensembles de gènes permet d'identifier des ensembles de gènes associés à divers résultats dans les études d'expression génique sur les microréseaux. Plusieurs méthodes ont été proposées pour les résultats binaires, mais seules quelques-unes pour les résultats continus. Le défi : un grand nombre de gènes dans chaque ensemble et un échantillon de petite taille, soit un problème de petit n , grand p ; et comment tenir compte des corrélations entre gènes dans un ensemble. Nous proposons une nouvelle méthode pour réduire les ensembles de gènes associés à un résultat continu pour leurs membres centraux. La méthode est fondée sur le test de combinaison linéaire pour les ensembles de gènes, qui inclut la matrice de covariance d'expression génique dans la statistique de test, via une méthode d'estimation du rétrécissement. Nous vérifions la méthode dans une étude de microréseau réelle de patients souffrant du cancer de la prostate.

GUOSHENG YIN (The University of Hong Kong)

[Tuesday May 27 / mardi 27 mai, 14:45]

Two-stage Single-to-double Arm Design for Phase II Clinical Trials

Plan à branches simples et doubles à deux étapes pour les essais cliniques de phase II

Phase II clinical trials use either a single-arm or two-arm comparison scheme to examine the experimental drugs therapeutic effects. Both single-arm and multi-arm evaluations have their own merits. To bridge the single- and two-arm schemes, we propose a two-stage design, in which the first stage takes a single-arm comparison of the experimental drug with the standard response rate and the second stage imposes a multi-arm comparison of the experimental drug with an active control arm. The

Les essais cliniques de phase II utilisent un schéma de comparaison à branches simples ou doubles pour examiner les effets thérapeutiques des médicaments expérimentaux. Les évaluations à branches simples et doubles ont chacune leurs mérites. Pour combler l'écart entre les deux approches, nous proposons un plan en deux étapes : dans un premier temps, une comparaison à branche simple du médicament expérimental et du taux de réponse standard ; puis dans un second temps, une comparaison à branches doubles du médicament expérimental et d'une branche de contrôle actif. Nous

2D-C1: Biostatistics - Applications 3

Biostatistique - applications 3

design is calibrated using a new concept, the detectable treatment difference, to balance the trade-offs between adaptation, power, and sample size. We conduct extensive simulation studies to examine the operational characteristics of our proposed method, and give an illustrative example of our design.

calibrons le plan à l'aide d'un nouveau concept, la différence de traitement détectable, afin de maintenir l'équilibre entre l'adaptation, la puissance et la taille d'échantillon. Nous effectuons de nombreuses études par simulation pour examiner les caractéristiques opérationnelles de la méthode proposée, puis donnons un exemple pour illustrer notre plan.

2D-C2: Biostatistics - Association Studies and Correlation Structure Biostatistique - études d'association et structure de corrélation

Chair/Présidente: Yildiz Yilmaz (Memorial University)

Room/Salle: UC - 244

Abstracts/Résumés

LAJMI LAKHAL-CHAIEB (Université Laval), **KARIM OUALKACHA** (Université du Québec à Montréal), **CELIA GREENWOOD** (McGill University)

[Tuesday May 27 / mardi 27 mai, 13:30]

A Family-based Test for Rare Variants Association in the Presence of Non-Normal Outcomes

Un test d'association pour les variantes rares avec des données familiales et des observations non-normales

Recent progress in sequencing technologies makes it possible to identify rare and unique variants that may be associated with complex traits. However, existing rare variant association approaches assume either independence of observations or normality of outcomes. In this talk, we present a family-based test for rare-variants association in the presence of non-normal outcomes. We consider a marginal generalized linear model for the response variable and model the familial association with a Gaussian copula model. We derive a score test for the association. We approximate the distribution of the score test under the null hypothesis with permutations. The proposed method is illustrated with an application to a real data set and evaluated by simulations.

Les progrès récents dans les technologies de séquençage ont permis d'identifier les variantes rares et uniques potentiellement associées à des traits complexes. Cependant, les méthodes existantes pour étudier l'association de variantes rares supposent l'indépendance ou la normalité des observations. Dans cet exposé, nous présentons un test d'association de variantes rares en présence de données familiales et observations non-normales. Nous considérons un modèle linéaire généralisé marginal pour la variable réponse et modélisons l'association familiale avec une copule gaussienne. Nous développons un test de type score pour l'association. Nous approximons la distribution de la statistique du test sous l'hypothèse nulle par des permutations. La méthode proposée est illustrée par une application à un ensemble de données réelles et évaluée par des simulations.

MUHAMMAD ABU SHADEQUE MULLAH & ANDREA BENEDETTI (McGill University)

[Tuesday May 27 / mardi 27 mai, 13:45]

Smoothing in a Semi-Parametric Generalized Linear Mixed Model and its Effect on the Estimation of Covariance Structures for Correlated Data

Lissage dans un modèle linéaire généralisé mixte semi-paramétrique et son effet sur l'estimation des structures de covariance des données corrélées

Besides its main use for analysing correlated data, generalized linear mixed models can also be used for smoothing. The resulting model is called semi-parametric generalized mixed model (SPGLMM). We compare smoothing in a SPGLMM to other methods such as fractional polynomials, generalized additive models, and parametric nonlinear function for correlated normal, Poisson and binary data. Through simulations, the effect of

En plus de leur principale fonction d'analyse des données corrélées, les modèles généralisés linéaires mixtes peuvent aussi être utilisés pour le lissage. Il en résulte un modèle généralisé mixte semi-paramétrique (MGLMSP). Nous comparons le lissage d'un MGLMSP à d'autres méthodes comme les polynômes fractionnels, les modèles additifs généralisés et les fonctions non linéaires paramétriques pour les données normales corrélées, les données de Poisson et les données binaires. Au moyen de simulations, l'ef-

2D-C2: Biostatistics - Association Studies and Correlation Structure Biostatistique - études d'association et structure de corrélation

smoothing on correlation and variance parameters has been evaluated by calculating bias and mean squared error of the estimates. Results suggest that SPGLMM yields reasonable estimates of the correlation and variance parameters. However, for binary outcome, SPGLMM produces biased estimates of variance parameters for highly correlated data. These methods then have been applied to a real life data set.

fet du lissage sur les paramètres de corrélation et de variance a été évalué en calculant le biais et l'erreur quadratique moyenne des estimations. Les résultats montrent que les MGLMSP produisent des estimations raisonnables des paramètres de corrélation et de variance. Pour les résultats binaires, par contre, les MGLMSP produisent des estimations biaisées des paramètres de variance pour des données fortement corrélées. Ces méthodes ont ensuite été appliquées à un jeu de données réelles.

CHUL GYU PARK (Carleton University)

[Tuesday May 27 / mardi 27 mai, 14:00]

A Representation for Binary Random Vector with Fixed Mean and Correlation Matrix

Représentation pour un vecteur aléatoire binaire à matrice de corrélation et à moyenne fixes

In the assessment of marginal models for dependent binary outcomes, it has been one of important issues to simulate correlated binary variables under given conditions. Most of such studies have been conducted in pursuit of better computational efficiency and flexibility in correlation structure. To this end, we introduce a representation for a multivariate binary variable that requires only independent uniform variables. This approach can be applied for efficient and easy generation of binary variables when their means and correlation matrix are fixed.

L'un des problèmes importants que présente l'évaluation des modèles marginaux pour des variables binaires dépendantes consiste à simuler des variables binaires corrélées dans des conditions diverses. La plupart des études de ce genre ont été réalisées avec un souci d'amélioration de l'efficacité des calculs et de la flexibilité de la structure de corrélation. À cette fin, nous présentons une représentation pur une variable binaire multidimensionnelle qui nécessite seulement des variables uniformes indépendantes. Cette approche peut être utilisée pour générer facilement et efficacement des variables binaires lorsque leur moyenne et leur matrice de corrélation sont fixes.

HELA ROMDHANI, AURELIE LABBE & HEUNGSUN HWANG (McGill University)

[Tuesday May 27 / mardi 27 mai, 14:15]

Genome-Wide Association Study Using Generalized Structured Component Analysis

Étude d'association pangénomique avec l'analyse en composantes structurées généralisées

We are interested in developing a statistical framework for the joint analysis of multiple correlated traits and multiple genotype measures from candidate regions in genetic studies. We propose to use the structural equation modeling with latent variables for the association structure between the observed variables and some components mediating the relationships between genotypes and phenotypes. The model is constructed on the basis of prior biological knowledge on both clinical and genetic pathways. We use the Generalized Structured Component Analysis (GSCA) to estimate the model's parameters. Tests procedures for different kinds of directed effects measured by GSCA

On s'intéresse à l'analyse simultanée de plusieurs traits corrélés et d'un ensemble de génotypes mesurés sur une ou plusieurs régions candidates du génomes. On propose d'utiliser les modèles à équations structurelles avec variables latentes pour la structure d'association entre les variables observées et certaines composantes médiatrices des relations entre les génotypes et les phénotypes. Le modèle est construit en se basant sur des connaissances biologiques préalables sur les cheminements cliniques (ou moléculaires) et génétiques. On utilise l'analyse en composantes structurées généralisées (GSCA) pour estimer les paramètres du modèle. Des procédures de test des différents effets mesurés par GSCA ont été développés et leur puissance étudiée par simulation. Enfin, cette nouvelle approche est

2D-C2: Biostatistics - Association Studies and Correlation Structure Biostatistique - études d'association et structure de corrélation

have been developed and their power has been assessed by simulations. Finally, an analysis of the QCAHS survey data is performed using this new approach.

utilisée pour analyser les données de l'enquête ESSEAQ.

JI-HYUNG SHIN & SHELLEY B. BULL (University of Toronto and Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital)

[Tuesday May 27 / mardi 27 mai, 14:30]

Can Sparse Data Methods Rescue Low Frequency Variants in Genome Wide Association Analysis?

Les méthodes de données éparses peuvent-elles sauver les variantes à faible fréquence dans l'analyse d'associations pangénomiques?

In genome-wide scans of binary traits, genetic variants that occur with low frequency (e.g., MAF < 1%) may be excluded even when the sample size is moderate (e.g., 1000), because standard logistic regression tests are subject to inflated type 1 error and/or low power. We investigate the performance of test statistics based on Firth's penalized logistic regression likelihood using analytic methods and finite-sample simulations across various scenarios. Overall, type 1 error is close to nominal for the penalized likelihood ratio test, the usual score test and the small-sample-adjusted SKAT. Neither the usual nor the penalized Wald tests maintain nominal level. We also identify some circumstances where alternative tests can recover information from low-frequency variants that otherwise would have been lost.

Dans les études pangénomiques de caractères binaires, les variantes génétiques peu fréquentes (p. ex., < 1 %) risquent d'être exclues même si la taille de l'échantillon est modérée (p. ex., 1000), parce que les tests de régression logistique standard sont sujets à une erreur de type 1 dilatée et/ou une faible puissance. Nous étudions la performance de statistiques de test fondées sur la vraisemblance de régression logistique pénalisée de Firth en utilisant des méthodes analytiques et des simulations par échantillons finis sur divers scénarios. Globalement, l'erreur de type 1 est proche de l'erreur nominale pour le test du rapport des vraisemblances pénalisé, le test de score habituel et le SKAT ajusté pour petits échantillons. Les tests de Wald habituels et pénalisés ne maintiennent pas leur niveau nominal. Nous identifions aussi certaines circonstances pour lesquelles des tests alternatifs permettent de récupérer des informations sur des variantes à faible fréquence qui seraient perdues sinon.

JIANPING SUN (McGill University), **KARIM OUALKACHA** (Université du Québec à Montréal), **CELIA GREENWOOD** (McGill University, Lady Davis Institute for Medical Research)

[Tuesday May 27 / mardi 27 mai, 14:45]

A Method for Analyzing Multiple Continuous Phenotypes in Rare Variant Association Studies

Méthode d'analyse de phénotypes continus multiples dans les études d'associations de variantes rares

Advances in high-throughput sequencing technologies have made it possible for researchers to conduct large-scale sequencing studies to study the association of rare genetic variants with complex traits or diseases. However, the power for detecting rare-variant associations is usually limited because very few individuals carry the (rare) variant alleles. Since often a study will collect several measures of related traits, i.e. a multivariate phenotype, it could be of interest to develop a test statistic that uses correlated traits in order to increase

Les progrès en technologie de séquençage à haut débit ont permis aux chercheurs d'effectuer des études de séquençage à grande échelle pour étudier l'association de variantes génétiques rares avec des caractères complexes ou des maladies. Cependant, la capacité à détecter des associations de variantes rares est généralement limitée car très peu d'individus sont porteurs des allèles de variantes (rares). Puisque les études recueillent souvent plusieurs mesures de caractères associés (ou phénotype multivarié), il pourrait s'avérer intéressant de mettre au point un test statistique qui utilise des caractères corrélés pour améliorer sa puissance.

2D-C2: Biostatistics - Association Studies and Correlation Structure Biostatistique - études d'association et structure de corrélation

test power. In this talk, I will introduce a newly developed statistical method for jointly testing the association between rare genetic variants and multiple correlated continuous traits jointly, and show its performance via simulations and real data analysis.

Dans cette présentation, j'introduis une nouvelle méthode statistique qui permet de tester l'association entre des variants génétiques rares et des caractères continus corrélés multiples en même temps ; j'en prouve la performance via des simulations et l'analyse de données réelles.

2D-C3: Business and Industrial Statistics 1 Statistique industrielle et de gestion 1

Chair/Président: Hugh Chipman (Acadia University)

Room/Salle: UC - A101

Abstracts/Résumés

KANIKA ANAND (Acadia University)

[Tuesday May 27 / mardi 27 mai, 13:30]

Sequential Design for Noisy Computer Experiments with a Bayesian Ensemble Model

Plan séquentiel pour des expériences informatiques perturbées par des bruits avec un modèle d'ensemble bayésien

Computer simulators are large computer codes written to represent numerical models of real phenomena that cannot be solved analytically. We consider problems with outputs that include noise, and the objective is to identify values of inputs that maximize the output. The computer simulator itself is often quite time-consuming to run, and thus we consider approximating it with a surrogate statistical model. Simulation results on test functions are presented using a quantile based approach for sequential design. A new surrogate model, a Bayesian ensemble of trees is considered to learn more about the simulator and optimize it.

Un simulateur informatique est un vaste code informatique écrit pour représenter des modèles numériques de phénomènes réels qui ne peuvent être résolus sur le plan analytique. Nous étudions des problèmes dont les extrants sont perturbés par des bruits et l'objectif consiste à identifier des valeurs d'intrants qui maximisent les extrants. Comme le simulateur informatique lui-même nécessite souvent beaucoup de temps de calcul, nous examinons la possibilité de l'approximer à l'aide d'un modèle statistique auxiliaire. Nous présentons des résultats de simulation sur des fonctions de test utilisant une approche à base de quantiles pour le plan séquentiel. Pour en apprendre davantage sur le simulateur et son optimisation, nous examinons un nouveau modèle auxiliaire et un ensemble d'arbres bayésien.

NEIL SPENCER (University of British Columbia), **PRITAM RANJAN & FRANKLIN MENDIVIL** (Acadia University)

[Tuesday May 27 / mardi 27 mai, 13:45]

Obtaining Two-Level Factorial Designs with Pre-specified Randomization Restrictions

Comment obtenir des plans factoriels à deux niveaux avec des restrictions de randomisation préspecifiées

Full factorial and fractional factorial designs with randomization restrictions (e.g., block designs, split-plot designs, and split-lot designs) are often used for industrial experiments when complete randomization of the trials is impractical. In this talk, we focus on the randomization defining contrast subspace (RDCSS) formulation of randomization-restricted experiments, namely spread-based (Ranjan, Bingham, and Dean (2009)) and star-based (Ranjan, Bingham, and Mukerjee (2010)) designs. We propose a new efficient relabelling-based method for obtaining such RDCSS-based designs with pre-specified randomization restrictions. The novelty in our approach lies in the use of the underlying ge-

Les plans factoriels complets et fractionnaires avec restrictions de randomisation (p. ex., plans en blocs, plans en parcelles divisées et en lots divisés) sont souvent utilisés dans les expériences industrielles pour lesquelles une randomisation complète de l'essai est impossible à réaliser. Dans cette présentation, nous nous intéressons à la formulation de sous-espace RDCSS (randomization defining contrast subspace) des expériences avec restrictions de randomisation, à savoir les plans de répartition (Ranjan, Bingham et Dean (2009)) et en étoile (Ranjan, Bingham et Mukerjee (2010)). Nous proposons une nouvelle méthode efficace fondée sur le réétiquetage pour obtenir de tels plans de type RDCSS avec des restrictions de randomisation préspecifiées. La nouveauté de notre approche tient à l'utilisation de

2D-C3: Business and Industrial Statistics 1

Statistique industrielle et de gestion 1

ometry of spreads and stars to reduce the search space of possible relabellings.

la géométrie sous-jacente des répartitions et des étoiles pour réduire l'espace de recherche de réétiquetages potentiels.

NATHANIEL STEVENS & STEFAN STEINER (University of Waterloo), **RYAN BROWNE** (University of Guelph), **JOCK MACKAY** (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 14:00]

Incorporating Baseline Information When Assessing Measurement Systems

Comment inclure les données de base dans l'évaluation de systèmes de mesure

The standard design of measurement system assessment studies is for each of r operators to measure k parts n times, for a total of $N = krn$ measurements. These studies are usually planned and conducted in isolation, and ignore available baseline data generated by the measurement system during regular use. The benefit of incorporating baseline data into the analysis is significant, and most of the gains in precision can be obtained with small baseline sample sizes. In this talk, we will first quantify the substantial benefits of incorporating baseline data into the analysis of measurement study data. Secondly, we will recommend optimal study designs that take into account available baseline data when the total number of measurements is fixed.

La plupart des plans d'évaluation de systèmes de mesure prévoit que chacun des r opérateurs doit mesurer k parties n fois pour un total de $N = krn$ mesures. Ces études sont généralement planifiées et effectuées isolément, sans tenir compte des données de base disponibles générées par le système de mesure en cours d'utilisation normale. Or il s'avère très utile de pouvoir inclure ces données de base dans l'analyse et la plupart des gains en précision s'obtiennent avec des échantillons de base de petite taille. Dans cette présentation, nous commencerons par quantifier les avantages d'une inclusion de ces données de base dans l'analyse des données d'étude de mesure. Ensuite, nous recommanderons des plans d'étude optimaux qui tiennent compte des données de base lorsque le nombre total des mesures est fixé.

SI CHEN (University of Waterloo), **ZILIN WANG** (Wilfrid Laurier University)

[Tuesday May 27 / mardi 27 mai, 14:15]

A Bias Correction for Capital Buffer Estimation in Retail Credit Risk Management

Une correction du biais dans l'estimation de la marge de capital dans la gestion du risque du crédit au détail

In credit risk management, capital buffer is prepared to cover potential loss, particularly in an economic downturn period. An adequate capital buffer is crucial to mitigate the risk. One of the determining factors of the capital buffer is an asset correlation between obligors. For retail portfolios, the single factor Vasicek model is adopted for underlying obligor's asset with homogeneous correlation. A commonly used approach under-estimates the correlation, which in turn would cause a smaller capital buffer than necessary, and thus would expose financial institutions in a risky situation. In this paper, we employ a bootstrapping method to correct the bias of estimation so as to aid risk management of regulators and practitioners.

Dans la gestion du risque du crédit, une marge de capital est préparée pour couvrir les pertes potentielles, particulièrement dans une période de ralentissement économique. Une marge de capital suffisante est cruciale pour atténuer le risque. Un des facteurs déterminants de la marge de capital est une corrélation des actifs entre les débiteurs. Pour les portefeuilles de détail, le modèle à facteur unique de Vasicek est adopté pour les actifs sous-jacents du débiteur avec une corrélation homogène. Une approche couramment utilisée sous-estime la corrélation, ce qui pourrait entraîner une marge de capital plus petite que nécessaire et par conséquent exposerait les institutions financières à une situation plus risquée. Dans cet article, nous utilisons une méthode d'auto-amorçage pour corriger le biais dû à l'estimation et ainsi aider les responsables de la réglementation et les praticiens dans la gestion du risque.

2D-C3: Business and Industrial Statistics 1

Statistique industrielle et de gestion 1

MATTHEW VAN BOMMEL, HUGH CHIPMAN & PRITAM RANJAN (Acadia University)

[Tuesday May 27 / mardi 27 mai, 14:30]

Stage-Wise Surrogate Modeling of Tidal Power Functions

Modélisation de substituts par étapes des fonctions d'énergie marémotrice

The Minas Passage in the Bay of Fundy has the potential to generate significant power from in-stream tidal energy turbines. Determining the optimal locations for these turbines can be seen as finding the (2D) locations which optimize a sequence of (potential) power functions. Approximating these functions with Gaussian Process (GP) models significantly decreases the required number of computer simulator evaluations. This talk will discuss a "stage-wise surrogate modeling" approach designed to approximate the power functions using GP models and further decrease the total simulator evaluations required to fit an accurate model at each step.

Minas Passage, dans la Baie de Fundy, possède le potentiel nécessaire pour générer beaucoup d'électricité via des turbines marémotrices. Pour déterminer l'emplacement optimal de ces turbines, il s'agit de trouver les emplacements (en 2D) qui optimisent une séquence de fonctions de puissance (potentielle). Une approximation de ces fonctions par modèles avec processus gaussien (PG) permet de nettement réduire le nombre d'évaluations simulées nécessaires. Nous discutons d'une approche de modélisation « de substituts par étapes » conçue pour approximer les fonctions d'énergie à l'aide de modèles PG et réduire le nombre d'évaluations simulées nécessaires pour ajuster un modèle précis à chaque étape.

FATEHA KHANAM BAPPEE & XU (SUNNY) WANG (St. Francis Xavier University)

[Tuesday May 27 / mardi 27 mai, 14:45]

Adaptively Chosen k in Weighted k -Nearest Neighbour Method for Drug Discovery Data

Sélection adaptative du k dans une méthode des k plus proches voisins pondérée pour des données de découverte de médicaments

Identifying active compounds against some biological target is of great value for drug discovery. Within millions of screened compounds there are very unbalanced class distributions. This paper proposes some innovative modifications to classify unbalanced data using k -Nearest Neighbour (KNN) method. The new method is named Adaptively Chosen k in Weighted KNN (Ad-WKNN). The main feature of Ad-WKNN is choosing an adaptive k for each test point. This method also employs a different weighting scheme for neighbours from the active and inactive classes and handles a mixed type of descriptor variables. AIDS antiviral screening data from the National Cancer Institute are used for the analysis. The experimental results indicate a significant classification performance improvement with Ad-WKNN on this type of data.

L'identification de composés actifs envers certaines cibles biologiques est d'une grande importance pour la découverte de médicaments. Parmi des millions de composés évalués, il y a des distributions de classes très déséquilibrées. Cet article propose des modifications novatrices pour la classification de données déséquilibrées à l'aide de la méthode des k plus proches voisins (KNN). Cette nouvelle méthode est nommée sélection adaptative dans un KNN pondéré (Ad-WKNN). La caractéristique principale du Ad-WKNN est de choisir un k adaptatif pour chaque point d'essai. Cette méthode utilise aussi une procédure de pondération différente pour les voisins des classes actives et inactives et permet de gérer un ensemble de variables descriptives mixtes. Des données de dépistage antiviral du SIDA de l'Institut National du Cancer ont été utilisées dans cette analyse. Les résultats expérimentaux indiquent une amélioration significative de la performance de la classification de ce type de données à l'aide du Ad-WKNN.

2D-C4: Quantile, Spatial and Logistic Regression Régression quantile, spatiale et logistique

Chair/Présidente: Cecilia Cotton (University of Waterloo)

Room/Salle: UC - Croft House

Abstracts/Résumés

TAOUFIK BOUEZMARNI (Université de Sherbrooke), **BENEDIKT FUNK** (Technische Universität Dortmund), **FÉLIX CAMIRAND LEMYRE** (Université de Sherbrooke)

[Tuesday May 27 / mardi 27 mai, 13:30]

Regression Estimation Based on Bernstein Density Copulas

Estimation de la régression en utilisant les densités de copules de Bernstein.

The regression function can be expressed in term of copula density and marginal distributions. In this paper, we propose a new method of estimating a regression function using the Bernstein copula density estimator for the copula density and the empirical distributions for the marginal distributions. The method is fully nonparametric and easy to implement. We provide some asymptotic results related to this copula-based regression modeling by showing the convergence in probability and the asymptotic normality of the estimator. We also study the finite sample performance of the estimator.

La fonction de régression peut être exprimée en terme de la densité de copule et les distributions marginales. Dans ce papier, nous proposons une nouvelle méthode d'estimation de la fonction de régression en utilisant l'estimateur de Bernstein pour la densité de copule et les distributions empiriques pour les distributions marginales. La méthode est entièrement non-paramétrique et facile à appliquer. Nous exposons des résultats asymptotiques de cet estimateur de régression : la convergence en probabilité et la normalité asymptotique. Nous montrons également avec les simulations la performance de l'estimateur.

CHARLES FONTAINE & TAOUFIK BOUEZMARNI (Université de Sherbrooke), **ANOUAR EL GHOUGH** (Université catholique de Louvain)

[Tuesday May 27 / mardi 27 mai, 13:45]

Semiparametric Estimation of the Regression Function in Situation of Right-Censored Data

Estimation semi-paramétrique de la fonction de régression dans une situation de données censurées à droite

This presentation discusses the semiparametric estimation of the regression function in situation of right-censored data. We present a new copula-based method to estimate the regression function. The key concept presented here is to write the regression function in terms of copula and marginals, and to substitute the marginal distribution of the response variable which is right-censored by an estimated distribution based on the product-limit estimator. We also provide a new method to avoid any problem of copula misspecification when the data come from an unknown distribution and, compare the performance of our estimator by simulations.

Dans cette présentation, l'estimation semi-paramétrique de la fonction de régression dans une situation de données censurées à droite sera examinée. Nous présentons une nouvelle méthode pour estimer la fonction de régression basée sur une copule. Le concept clé que nous présentons est d'écrire la fonction de régression en termes de copules et de marginales et de remplacer la distribution marginale de la variable réponse qui est censurée à droite par une distribution estimée basée sur l'estimateur produit-limite. Nous fournissons aussi une nouvelle méthode pour éviter tout problème d'erreur de spécification de la copule quand les données proviennent d'une distribution inconnue et nous comparons la performance de notre estimateur par des simulations.

2D-C4: Quantile, Spatial and Logistic Regression Régression quantile, spatiale et logistique

CHITHRAN VADAVERKKOT VASUDEVAN, ZHAO ZHI FAN & ASOKAN VARIYATH (Memorial University of Newfoundland)

[Tuesday May 27 / mardi 27 mai, 14:00]

Censored Quantile Regression with Auxiliary Covariates

Régression quantile censurée avec covariables auxiliaires

In quantile regression of survival data, the existence problem of estimating the regression coefficients for extreme quantiles can be triggered by severe censoring. A large sample is needed to have bigger event times included. In epidemiological studies, there is often only a small portion of the entire study cohort being accurately observed. The rest of the cohort has only some auxiliary information available. In this paper, we propose a nonparametric smoothing technique and conduct quantile regression analysis using the whole study cohort. The proposed estimator is consistent and asymptotically normal. The efficiency gain versus that solely based on the validation sample is remarkable. This method provides us with possibilities of looking into higher (lower) extreme quantiles of the failure distribution.

Dans la régression quantile des données de survie, le problème d'existence de l'estimation des coefficients de régression des quantiles extrêmes peut être déclenché par une censure sévère. Il faut un grand échantillon pour pouvoir inclure de plus grands temps d'événements. Dans les études épidémiologiques, souvent seule une petite portion de la cohorte étudiée est observée avec précision. Pour le reste de la cohorte, seules quelques informations auxiliaires sont disponibles. Dans cette présentation, nous proposons une technique de lissage non paramétrique et effectuons une analyse de régression quantile sur l'ensemble de la cohorte étudiée. L'estimateur proposé est cohérent et asymptotiquement normal. Le gain d'efficacité par rapport à une analyse fondée uniquement sur l'échantillon de validation est remarquable. Cette méthode nous offre des possibilités pour étudier les quantiles extrêmes supérieurs (inférieurs) de la distribution de faillite.

MARK WOLTERS & CHARMAINE DEAN (Western University)

[Tuesday May 27 / mardi 27 mai, 14:15]

Detection of Smoke in Satellite Images Using Autologistic Regression

Détection de la fumée dans les images satellite par régression auto-logistique

Satellite imagery provides a rich data source for studying the spread of smoke produced by forest fires. The present work considers hyperspectral images, where each of 36 image planes holds a measurement in a different spectral band. The objective is to develop an automatic image segmentation routine: a classifier that assigns a new image's pixels to either *smoke* or *nonsmoke* categories. To incorporate association between nearby pixels, the true scene is modelled as a binary random field. An autologistic regression approach is used to model the joint probability mass function of the scene, with the hyperspectral data as predictors. This model poses computational challenges for feature selection, estimation, and prediction; methods for overcoming these challenges will be discussed.

L'imagerie satellite offre une riche source de données pour l'étude de la diffusion de la fumée produite par les incendies de forêt. Dans cette présentation, nous étudions les images hyperspectrales, où chacune de 36 plans d'image donne une mesure dans une bande spectrale distincte. L'objectif est de mettre au point une routine de segmentation d'image automatique : un classificateur qui affecte les pixels d'une nouvelle image à la catégorie *smoke* ou *nonsmoke* (fumée ou non-fumée). Pour inclure l'association entre pixels voisins, la scène réelle est modélisée comme un champ aléatoire binaire. Nous utilisons une approche de régression autologistique pour modéliser la fonction de masse de probabilité commune de la scène, avec les données hyperspectrales comme prédicteurs. Ce modèle pose des problèmes de calcul quant à la sélection des caractéristiques, l'estimation et la prédiction ; nous discutons de méthodes pour surmonter ces problèmes.

2D-C4: Quantile, Spatial and Logistic Regression Régression quantile, spatiale et logistique

YIN XU, MEI LING HUANG & WAI KONG YUEN (Brock University)

[Tuesday May 27 / mardi 27 mai, 14:30]

On High Quantile Regression

Régression des quantiles supérieurs

Recently studies of heavy tailed distributions have rapidly developed. For multivariate heavy tailed distributions, estimation of conditional quantiles at very high or low tails is of interest in numerous applications. Quantile regression uses an L1-loss function, and the optimal solution of linear programming for estimating coefficients of regression. This paper proposes a weighted quantile regression method on high quantile regression for certain extreme value sets. The Monte Carlo simulations show good results of the proposed weighted method. The paper also investigates a real-world example of application on extreme events by using the proposed method.

Récemment, de nombreuses études ont été effectuées sur les distributions à queues lourdes. Pour les distributions à queues lourdes multivariées, l'estimation des quantiles conditionnels pour les queues très élevées ou faibles est intéressante pour de nombreuses applications. La régression quantile utilise une fonction de perte L1 et la solution optimale de la programmation linéaire pour estimer les coefficients de régression. Dans cette présentation, nous proposons une méthode de régression quantile pondérée sur la régression des quantiles supérieurs pour certains ensembles de valeurs extrêmes. Les simulations de Monte Carlo montrent de bons résultats pour cette méthode pondérée. Nous étudions également une application réelle de cette méthode sur les événements extrêmes.

XUEMAO ZHANG (Trent University)

[Tuesday May 27 / mardi 27 mai, 14:45]

Teaching Introductory Statistics Using RExcel

Enseigner l'introduction à la statistique à l'aide de RExcel

In recent years, professional statistical packages like SPSS and Stata are often used in teaching introductory statistics courses since they have user-friendly point-and-click interfaces. A more popular spreadsheet software Microsoft Excel is a great option for entering, organizing, and analyzing data. However, many statistical methods are not available in Excel. The free Excel add-in RExcel allows access to the statistics package R from within Excel. The use of RExcel as a GUI for R, making R functionality accessible through menus and dialog boxes like in SPSS. Moreover, it makes R easier to use for those not familiar with the command-line interface. I illustrate the use of RExcel in teaching introductory statistics courses in this talk.

Ces dernières années, on utilise de plus en plus de logiciels statistiques professionnels comme SPSS et Stata pour enseigner les cours d'introduction à la statistique car ils présentent des interfaces conviviales pointer-cliquer. Un logiciel de tableur populaire comme Microsoft Excel est très utile pour la saisie, l'organisation et l'analyse de données. Cependant, rares sont les méthodes statistiques disponibles sous Excel. RExcel, un add-in gratuit pour Excel, permet d'accéder au logiciel statistique R depuis Excel. RExcel est une interface graphique pour R, rendant les fonctions R accessibles via des menus et des boîtes de dialogue, comme sous SPSS. De plus, R devient plus facile à utiliser pour les étudiants qui ne connaissent pas bien l'interface de ligne de commande. J'illustre l'utilisation de RExcel dans l'enseignement des cours d'introduction à la statistique.

2D-C5: Statistical Inference and Applications 4 Inférence statistique et applications 4

Chair/Présidente: Leilei Zeng (University of Waterloo)

Room/Salle: UC - 144

Abstracts/Résumés

JULES JOSEPH SUNDA DE TIBEIRO & VARTAN CHOULAKIAN (Université de Moncton)

[Tuesday May 27 / mardi 27 mai, 13:30]

Graph Partitioning by Correspondence Analysis and Taxicab Correspondence Analysis

Partitionnement des graphes par l'analyse factorielle des correspondances et par l'analyse factorielle des correspondances du taxi

We consider correspondence analysis (CA) and taxicab correspondence analysis (TCA) of relational datasets that can mathematically be described as weighted loopless graphs. Such data appear in particular in network analysis. We present CA and TCA as relaxation methods for the graph partitioning problem. Examples of real datasets are provided.

Nous considérons ici l'analyse factorielle des correspondances (AFC) et l'analyse factorielle des correspondances du taxi (AFCT) sur la base des jeux de données relationnelles pouvant être mathématiquement décrites en termes de graphes pondérés «sans boucles». De telles données réelles apparaissent particulièrement dans l'analyse des réseaux. Nous confrontons à cette fin l'AFC et l'AFCT comme des méthodes de relaxation dans l'optique du problème de partitionnement des graphes.

GABRIEL ESCARELA (Universidad Autónoma Metropolitana - Iztapalapa)

[Tuesday May 27 / mardi 27 mai, 13:45]

Analysis and Prediction of Daily Maxima of Ozone Levels in Guadalajara City

Analyse et prédiction des maximums quotidiens des niveaux d'ozone dans la ville de Guadalajara

This paper develops an extreme value model to both analyzing and forecasting daily meteorologically adjusted tropospheric ozone maxima. The methods are applied to a 10-year follow-up of daily maxima of ozone levels for the metropolitan area of Guadalajara City, Mexico. The model adjusts the daily rate of change in ozone for concurrent impacts of seasonality and present and past meteorological conditions, which include surface temperature, wind speed, wind direction, relative humidity, and ozone. One-day-lagged and present meteorological variables, 1- and 2-day-lagged tropospheric ozone maxima, seasonality, and a curvilinear trend are important predictors of the daily tropospheric ozone maximum in Guadalajara. The method provides a reliable tool to predict ozone levels exceeding a relevant threshold.

Cet article développe un modèle de la valeur extrême pour analyser et prédire les maximums d'ozone troposphérique quotidiens ajustés météorologiquement. Les méthodes sont appliquées sur un suivi de 10 ans des niveaux d'ozone quotidiens pour la région métropolitaine de la ville de Guadalajara au Mexique. Le modèle ajuste le taux de variation quotidien de l'ozone selon les impacts saisonniers concurrents et les conditions météorologiques présentes et passées, qui incluent la température de surface, la vitesse du vent, la direction du vent, l'humidité relative et l'ozone. Les variables météorologiques présentes et décalées d'une journée, les maximums troposphériques d'ozone décalés d'une ou deux journées, la saisonnalité et la tendance curviligne sont des prédicteurs importants du maximum quotidien d'ozone à Guadalajara. La méthode offre un outil fiable pour prédire les niveaux d'ozone dépassants un seuil pertinent.

2D-C5: Statistical Inference and Applications 4

Inférence statistique et applications 4

FENG (FRANK) FENG (City of Edmonton - Office of Traffic Safety)

[Tuesday May 27 / mardi 27 mai, 14:00]

Time Series Analysis for Collision Forecasting

Analyse de séries chronologiques pour la prédiction de collisions

For years, Office of Traffic Safety of City of Edmonton has been trying to make long-term forecasts of city-wide intersection collisions to support strategic planning. A most common type of models used in the transportation safety engineering field for estimating and predicting collision counts is generalized linear models (GLM), which assume collisions occur as the result of some Poisson process. As traditional GLM models cannot effectively overcome some major technical challenges associated with this macro-level city-wide forecasting problem, a variety of time series models are calibrated including ARIMA, pure GLM, GLM combined with ARIMA, ARMAX, and support vector regression models. These techniques are for the first time systematically introduced into this research area.

Depuis des années, le bureau de la sécurité de la circulation de la ville d'Edmonton essaye de prédire à long terme les collisions à ses intersections à la grandeur de la ville pour supporter la planification stratégique. Un type de modèle fréquemment utilisé dans le domaine de l'ingénierie de la sécurité des transports pour l'estimation et la prédiction du nombre de collisions est le modèle linéaire généralisé (MLG), qui suppose que les collisions surviennent à la suite d'un processus de Poisson. Puisque les modèles MLG traditionnels ne peuvent surmonter efficacement certains défis techniques importants associés à ce problème de grande échelle de prédiction, une variété de modèles de séries chronologiques sont calibrés à l'aide de ARIMA, MLG purs, MLG combiné avec ARIMA, ARMAX et des modèles de soutien de régressions vectorielles. Ces techniques sont introduites de façon systématique pour la première fois dans ce champ de recherche.

DAVID HAN (University of Texas)

[Tuesday May 27 / mardi 27 mai, 14:15]

Time and Cost Constrained Optimal Designs of Constant-stress and Step-stress Accelerated Life Tests

Plans optimaux d'essais de vie accélérée sous contraintes de temps et de coût pour une tension constante ou échelonnée

To conduct an accelerated life test efficiently with constrained resources in practice, several decision variables such as allocation proportions and stress durations should be determined carefully at the design stage. These variables affect not only the experimental cost but also the estimate precision. In this work, under the constraint that the total experimental cost does not exceed a pre-specified budget, the optimal decision variables are determined based on $C/D/A$ -optimality criteria. In particular, constant-stress and step-stress accelerated life tests are considered with exponential failure data under time constraint. We illustrate the proposed methods using two case studies, and under the identical budget constraint, efficiencies of these stress loading schemes are compared in terms of the ratio of optimal objective functions.

Pour mener efficacement un essai de vie accélérée sous contraintes de ressources, il faut déterminer soigneusement à l'étape de conception bon nombre des variables décisionnelles, notamment les proportions de répartition et les durées de tensions. Ces variables influencent à la fois le coût expérimental et la précision de l'estimation. Compte tenu d'une contrainte de coût total n'excédant pas un budget préalablement déterminé, les variables de décision optimales de la présente étude sont déterminées en fonction des critères d'optimalité $C/D/A$. Nous analysons en particulier des essais de vie accélérée sous tension constante et sous tension échelonnée avec des données exponentielles de défaillance sous contrainte de temps. Nous illustrons les méthodes proposées à l'aide de deux études de cas, et compte tenu de contraintes de budget identiques, nous comparons l'efficacité de ces plans de tensions mesurée par le ratio des fonctions objectives optimales.

2D-C5: Statistical Inference and Applications 4 Inférence statistique et applications 4

RAJAT MALIK, ROB DEARDON & GRACE P.S. KWONG (University of Guelph)

[Tuesday May 27 / mardi 27 mai, 14:30]

Parameterizing Individual-Level Models of Infectious Disease Spread Using Sampling-Based Likelihood Approximations

Établissement des paramètres de modèles au niveau individuel pour la propagation de maladies infectieuses par des approximations de la vraisemblance basée sur l'échantillonnage

Individual-level models (ILMs), typically fitted in a Bayesian MCMC framework, are a class of discrete-time models used to model the spread of infectious diseases. They can account for spatial and temporal disease dynamics by modeling the infectious pressure exerted by infected individuals in a population on each susceptible individual. Unfortunately, for large populations, quantifying this infectious pressure can be computationally burdensome leading to a time-consuming likelihood calculation. Here, we introduce sampling methods to speed-up the calculation of the likelihood function while accounting for the accuracy of model parameterization and compare the performances of these methods via simulation studies. Our results suggest substantial computational savings can be attained with acceptable information loss.

Ajustés dans un cadre bayésien de MCMC, les modèles au niveau individuel (MNI) sont une classe de modèles à temps discret utilisés pour modéliser la propagation de maladies infectieuses. Ils peuvent tenir compte de la dynamique spatio-temporelle des maladies en modélisant la pression infectieuse exercée sur chaque individu sensible par les personnes infectées dans une population. Malheureusement, la quantification de cette pression infectieuse pour de grandes populations peut nécessiter de lourds calculs demandant ainsi beaucoup de temps pour le calcul de la vraisemblance. Nous présentons ici des méthodes d'échantillonnage qui permettent d'accélérer le calcul de la vraisemblance tout en tenant compte de la précision de paramétrage, et de comparer la performance de ces méthodes grâce à des études de simulation. Nos résultats montrent qu'un allègement substantiel des calculs peut en résulter avec un niveau acceptable de perte d'information.

G. ARTHUR MIHRAM (Princeton, NJ), **DANIELLE MIHRAM, PH. D.** (University of Southern California)

[Tuesday May 27 / mardi 27 mai, 14:45]

Re-examining Quinn's Conclusion: Don't Either Our Statistical Methodologies or Works in Applied Mathematics Qualify as Science?

Réexamen de la conclusion de Quinn : nos méthodes statistiques et nos travaux en mathématiques appliquées ne se qualifient-ils pas au titre de science?

Science is that human activity dedicated to providing the very explanation for (the truth about) any particular naturally-occurring phenomenon. Yet, mathematics is not science: Validity criteria differ [Quinn, NOTICES, AMS59:31(2012)]. The Method of (Modern) Science?: A six-stage, essentially sequential, model-building process [TEOREMA28(2):35(2009)]: No.0 EXTANT KNOWLEDGE; No.I OBSERVATION + MENTAL REFLEXION (thereon); No.II THE MODEL (Artwork); No.III PERSCRUTATION (Editorial Review); No.IV CONFIRMATION (Nature-to-Model Concordances); No.V KNOWLEDGE AUGMENTATION: yet, three requisite feedback correctives: III-->II; IV-->I; and/or IV-->II. Applied mathematicians believe Science

La science est une activité humaine consacrée à la recherche de l'explication même (de la vérité) de tout phénomène d'origine naturelle particulier. Pourtant, les mathématiques ne sont pas une science : Les critères de validité diffèrent [Quinn, NOTICES, AMS59 :31(2012)]. La méthode de la science (moderne) ? : Un processus d'élaboration en six étapes d'un modèle essentiellement séquentiel [TEOREMA28(2) :35(2009)] : No.0 CONNAISSANCE EXISTANTE ; No.I OBSERVATIONS + RÉFLEXION MENTALE (sur celles-ci) ; No.II LE MODÈLE (Illustration) ; No.III PERSCRUTATION (Revue éditoriale) ; No.IV CONFIRMATION (Concordance entre le modèle et la nature) ; No.V AUGMENTATION DE LA CONNAISSANCE : pourtant, trois nécessitent des rétroactions correctives : III-->II ; IV-->I ; et/ou IV-->II. Les mathématiciens appliqués estiment que la sci-

2D-C5: Statistical Inference and Applications 4

Inférence statistique et applications 4

mimes theorem-proving; yet, even if one's 'assumption(s)' are (each) well-founded scientific models, logical deductions therefrom can CONFIRM[No.IV] the original models' validity: yet, no NEW scientific model[No.V]. On statistical regression? Recall [ABSTRACTS,SSC2012]: ModelFormat[No.II] precedes StageNo.I], violating mathematician Cotes's 'Preface' [Newton's NATURAL PHILOSOPHY'S [Modern Science's] MATHEMATICAL PRINCIPLES, ed2, ca.1713.)

ence mimique la démonstration des théorèmes, pourtant si une ou des hypothèse(s) sont (chacune) des modèles scientifiques légitimes, des déductions logiques peuvent par conséquent CONFIRMER [No.IV] la validité des modèles originaux : pourtant, il n'est pas question de modèle scientifique NOUVEAU [No.V]. À propos de la régression statistique ? Rappelez-vous [RÉSUMÉS, SSC2012 : Format du modèle [No.II] précède le stade No.I], déroge à la préface du mathématicien Cotes [LA PHILOSOPHIE NATURELLE DE Newton EST L'ENSEMBLE DES PRINCIPES MATHÉMATIQUES DE [la science moderne], ed2, ca.1713.)

**2E-I1: Bridging the Gap between Methods Development and Implementation in
Biostatistics (A Panel Discussion)**
**Comment combler le fossé entre le développement et la mise en œuvre de méthodes en
biostatistique (table ronde)**

Organizer and Chair / Responsable et présidente: Eleanor Pullenayegum (Hospital for Sick Children)

Session sponsored by the Biostatistics Section
Séance parrainée par le Groupe de biostatistique

Room/Salle: Hart House - Debates Rm

Description

Biostatisticians, clinical epidemiologists and a knowledge translation expert will lead a discussion on how to improve the impact of our developments in statistical methodology. Biostatisticians often complain that statistical methodology in health research falls short of best (or even adequate) practices. If we struggle to get basic methods used correctly, we should also expect to struggle to get our advances in methodology implemented in practice. This session will look at the gap between best practice and actual practice, asking where the gap is, why it exists, and how we might bridge it.

You will shape the discussion! Please visit <https://www.surveymonkey.com/s/36VKG7H> to fill out a short, anonymous survey before the session so we can target the discussion around the important issues.

Des biostatisticiens, des épidémiologistes cliniques et un expert en application des connaissances vont mener une discussion sur comment améliorer l'impact des nos développements en méthodologie statistique. Les biostatisticiens se plaignent souvent que la méthodologie statistique appliquée en recherche dans le domaine de la santé est rarement la meilleure (ou même simplement adéquate). S'il est difficile de faire appliquer correctement les méthodes de base, on devrait s'attendre à ce qu'on ait de la misère à mettre en œuvre nos avancées méthodologiques. Cette séance jettera un regard au fossé qui existe entre la meilleure pratique et la pratique réelle, en examinant où est le fossé, pourquoi il existe, et comment nous pouvons le combler.

Participez à l'élaboration de la discussion! SVP, allez sur le site <https://www.surveymonkey.com/s/36VKG7H> pour remplir un court sondage anonyme avant la séance pour nous aider à cibler la discussion autour des questions importantes.

2E-I2: Challenges and New Developments in Complex Multiple Comparison Procedures Défis et progrès récents pour les procédures de comparaisons multiples complexes

Chair/Président: Wenqing He (Western University)
Organizer/Responsable: Jianan Peng (Acadia University)

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

FRANK BRETZ & WILLI MAURER (Novartis Pharma AG)

[Tuesday May 27 / mardi 27 mai, 15:30]

Multiple Testing in Group Sequential Trials using Graphical Approaches

Tests multiples dans des essais séquentiels par groupes à l'aide de méthodes graphiques

We consider the situation of testing multiple hypotheses repeatedly in time using recently developed graphical approaches. We focus on closed testing procedures using weighted group sequential Bonferroni tests for the intersection hypotheses. Under mild monotonicity conditions on the error spending functions, this allows the use of sequentially rejective graphical procedures in group sequential trials. The methodology is illustrated with a numerical example from a real clinical trial.

Nous testons des hypothèses multiples à répétition à l'aide de méthodes graphiques développées récemment. Nous nous concentrons sur des procédures de test fermées au moyen de tests de groupe de Bonferroni séquentiels pondérés pour les hypothèses d'intersection. Dans des conditions légères de monotonie des fonctions de dépense basée sur l'erreur, il est possible d'utiliser des procédures graphiques réjectives de façon séquentielle dans des essais séquentiels par groupe. La méthodologie est illustrée à l'aide d'un exemple numérique provenant d'un essai clinique réel.

SIU HUNG CHEUNG (Chinese University of Hong Kong)

[Tuesday May 27 / mardi 27 mai, 16:00]

Non-inferiority Studies with Multiple New Treatments

Études de non-infériorité en présence de nouveaux traitements multiples

The objective of a non-inferiority (NI) trial is to affirm the efficacy of a new treatment compared with an active control by verifying that the new treatment maintains a considerable portion of the treatment effect of the control. Compensation by benefits other than efficacy is usually the justification for using a new treatment, as long as the loss of efficacy is within an acceptable margin from the standard treatment. A popular approach is to express this margin in terms of the efficacy difference between the new treatment and the active control. Based on this approach and the realization that NI trials often comprise several new treatments, statistical procedures that simultaneously conduct NI tests of several new treatments are discussed.

L'objectif d'un essai de non-infériorité consiste à affirmer l'efficacité d'un nouveau traitement par rapport à un groupe témoin actif en vérifiant qu'il conserve une portion importante de l'effet du traitement du groupe témoin. L'utilisation d'un nouveau traitement est habituellement justifiée par des bénéfices autres que l'efficacité, en autant que la perte d'efficacité soit à l'intérieur d'une marge acceptable par rapport au traitement standard. Une approche populaire consiste à exprimer cette marge comme la différence d'efficacité entre le nouveau traitement et le groupe témoin actif. D'après cette approche et compte tenu que les essais de non-infériorité comprennent souvent plusieurs nouveaux traitements, nous analysons des procédures statistiques qui effectuent simultanément des essais de non-infériorité pour plusieurs nouveaux traitements.

2E-I2: Challenges and New Developments in Complex Multiple Comparison Procedures Défis et progrès récents pour les procédures de comparaisons multiples complexes

JIANAN PENG (Acadia University)

[Tuesday May 27 / mardi 27 mai, 16:30]

Multiple Confidence Intervals for Selected Parameters Adjusted for the False Coverage-Statement Rate in Dose-Response Microarray Experiments

Intervalle de confiance multiples pour des paramètres sélectionnés et ajustés en fonction du taux de fausse couverture dans le cadre d'expériences avec biopuces portant sur la relation dose-effet

Benjamini and Yekutieli (2005) introduced a general procedure to control the false coverage-statement rate (FCR) in constructing confidence intervals and highlighted that the use of the FCR-adjusted confidence intervals is needed as an analogue to the control of FDR in addressing for multiple testing problems. Dose-response analysis in dose-response microarray experiments is provided only for genes that are found to have monotone dose-response relationship, which is a selection-adjusted problem. Lin et al. (2012) considered multiple confidence intervals for selected ratio parameters adjusted for FCR. In this talk, we consider multiple confidence intervals for mean differences under monotonicity for the subset of selected genes adjusted for the FCR.

Benjamini et Yekutieli (2005) ont présenté une procédure générale de contrôle du taux de fausse couverture (TFC) dans la construction d'intervalles de confiance et ont souligné que l'utilisation d'intervalles de confiance ajustés en fonction du TFC est nécessaire comme méthode analogue de contrôle du taux de fausse découverte (TFD) pour la résolution de problèmes de tests multiples. L'analyse dose-effet dans le cadre d'expériences avec biopuces est fournie uniquement pour les gènes présentant une relation dose-effet monotone, qui constitue un problème d'ajustement à la sélection. Lin et coll. (2012) ont étudié les intervalles de confiance multiples pour les paramètres du rapport sélectionnés ajustés en fonction du TFC. Dans cet exposé, nous considérons des intervalles de confiance multiples pour des différences de moyenne sous le critère de monotonie pour le sous-ensemble de gènes sélectionnés ajustés en fonction du TFC.

2E-I3: Copula Theory and Methods Théorie et méthodes liées aux copules

Organizer and Chair / Responsable et président: Ruodu Wang (University of Waterloo)

Room/Salle: Hart House - South Dining

Abstracts/Résumés

CLAUDIA CZADO (Technische Universität München)

[Tuesday May 27 / mardi 27 mai, 15:30]

Vines: Building Multivariate Copulas Using Pair Copula Constructions

Vignes: Construction de copules multivariées au moyen de copules construites par paires

Standard multivariate copula classes such as the elliptical or Archimedean one are restricted in their tail and symmetry behavior, since they are closed under margins. These restrictions might not be satisfied in real data applications. In contrast vine copula models are very flexible. They are constructed using only bivariate copula building blocks called pair copulas. The full specification of a vine model requires the choice of vine tree structure, copula families for each pair copula term and their corresponding parameters. I will introduce this class of copulas and discuss their statistical inference including model selection. Approaches will be illustrated using financial data.

Les classes de copules multivariées classiques, comme les copules elliptiques et archimédiennes, présentent un comportement caudal et symétrique limité en raison de leur fermeture sous les marges. Il n'est pas nécessairement possible de répondre à ces limites dans le cadre d'applications à des données réelles. En revanche, les modèles de copules en vignes sont très flexibles. Ils sont construits en utilisant uniquement des composantes bivariées appelées copules par paires. La spécification complète d'un modèle en vignes nécessite le choix de la structure de l'arbre en vignes, de la famille de copule pour chaque paire et de leurs paramètres respectifs. Je présente cette classe de copules et discute leur inférence statistique, y compris le choix du modèle. La méthode proposée est illustrée à l'aide de données financières.

HAIJUN LI (Washington State University)

[Tuesday May 27 / mardi 27 mai, 16:00]

Tail Densities of Copulas

Densités des queues de copules

The tail density of a copula, if exists, is a limiting density of the copula with proper scaling, and can be used to study multivariate extremes. In this talk, we will use tail densities to study mixtures of copulas, and show how tails of copulas can be changed from tail independence to tail dependence (and vice versa), via random mixture. The application to tail distortion is also discussed.

Si elle existe, la densité de la queue d'une copule est une densité limite de copule mise à l'échelle de façon appropriée qui permet d'étudier les valeurs extrêmes multivariées. Dans cet exposé, nous étudions les mélanges de copules à l'aide des densités de queues et nous montrons comment transformer la dépendance caudale des copules de l'indépendance vers la dépendance (et vice-versa) par le biais de mélanges aléatoires. Nous abordons aussi une application relative à la déformation de la queue.

EMILIANO VALDEZ (Michigan State University), **EDWARD W. FREES** (University of Wisconsin - Madison)

[Tuesday May 27 / mardi 27 mai, 16:30]

A Revisit of Hierarchical Insurance Claims Modeling

Réexamen de la modélisation hiérarchique de réclamations d'assurance

2E-I3: Copula Theory and Methods Théorie et méthodes liées aux copules

This work describes statistical modeling of detailed, microlevel auto insurance records to analyze experience at the individual vehicle level. We propose a hierarchical model for three components corresponding to the frequency, type, and severity of claims. The first is a negative binomial regression model for assessing claim frequency, with gender, age, and no claims discount as important predictors. The second is a multinomial logit model to predict the type of insurance claim. The third model is for the severity component where we use GB2 distribution for claim amounts and incorporate predictor variables. Using a t-copula, we show a significant dependence among different claim types. This integrated model allows a more efficient prediction of claims than traditional methods.

Cet article porte sur la modélisation statistique de registres d'assurance automobile détaillés à des microniveaux en vue d'analyser l'expérience au niveau de chaque véhicule. Nous proposons un modèle hiérarchique pour trois éléments correspondant à la fréquence, au type et à la gravité des réclamations. Le premier est un modèle de régression de type binomiale négative qui permet d'évaluer la fréquence des réclamations et dont le sexe, l'âge et la prime au bon conducteur constituent des variables prédictives importantes. Le deuxième est un modèle de régression multinomial qui prédit le type de réclamation. Le troisième modèle portant sur la gravité des sinistres fait appel à la loi bêta prime (GB2) pour le montant des réclamations et il intègre les variables prédictives. À l'aide d'une copule de Student, nous montrons l'existence d'une dépendance significative parmi les différents types de réclamations. Ce modèle intégré permet d'obtenir une prédiction plus efficace des réclamations par rapport aux méthodes traditionnelles.

2E-I4: Recent Advances in the Analysis of Complex Surveys Récents progrès en analyse d'enquêtes complexes

Chair/Présidente: Wendy Lou (University of Toronto)

Organizers/Responsables: Victoria Landsman and/et Wendy Lou (University of Toronto)

Session sponsored by the ICOSA – Canada Chapter

Séance parrainée par le chapitre canadien de l'ICOSA

Room/Salle: Hart House - East Common

Abstracts/Résumés

BARRY GRAUBARD (National Institute of Health), **MONROE SIRKEN** (Retired)

[Tuesday May 27 / mardi 27 mai, 15:30]

Estimating Sibling Recurrence Risk in Population Sample Surveys

Estimation du risque de récurrence chez la fratrie dans des enquêtes sur la population

Sibling recurrence risk (SRR), a measure of familial aggregation of a disease, is used in genetic epidemiology to indicate possible genetic causes of disease. Estimating SRR requires information about the disease status of sibships of families with two or more siblings where at least one sibling is affected. Network sampling often used in survey research to estimate rare population characteristics also offers a way to ascertain disease status of sibships in household surveys samples, and obtain (approximately) unbiased estimators of SRR. Reported diabetes for siblings from the 1976 National Health Interview is used to illustrate that network sampling estimators can provide population estimates of SRR for diabetes.

Le risque de récurrence chez la fratrie (SRR) est une mesure d'agrégation familiale d'une maladie qui est utilisée en épidémiologie génétique pour indiquer les causes génétiques possibles d'une maladie. L'estimation du SRR nécessite de l'information ayant trait à l'état de maladie de deux frères ou soeurs et plus, où au moins l'un d'eux est atteint. L'échantillonnage par réseau, souvent utilisé dans les enquêtes pour estimer des caractéristiques rares d'une population, offre aussi une façon d'évaluer l'état de maladie de la fratrie dans des échantillons d'enquête auprès des ménages, et d'obtenir approximativement des estimateurs non biaisés de SRR. Les cas de diabète signalés dans l'enquête du National Health Interview de 1976 servent à illustrer que les estimateurs d'échantillonnage par réseau peuvent produire des prévisions démographiques du SRR pour le diabète.

VICTORIA LANDSMAN, LOU WENDY & FAN LINGLING (University of Toronto)

[Tuesday May 27 / mardi 27 mai, 16:00]

Methods to Address Bias in Web Surveys

Méthodes pour tenir compte du biais dans les enquêtes en ligne

With the rise of the internet, web surveys have become an easy and inexpensive means to collect data for population studies, e.g. to gauge consumer attitudes or political opinions. Statistical inference based on web-survey samples often requires cautious treatment to account for potential biases due to lack of representation of the respon-

Avec l'essor de l'Internet, les enquêtes en ligne sont devenues un moyen facile et peu coûteux de recueillir des données à des fins d'études sur la population, notamment pour mesurer les attitudes des consommateurs ou les opinions politiques. Dans l'application des méthodes statistiques classiques pour les données d'enquête, l'inférence statistique basée sur des échantillons d'enquête en ligne nécessite sou-

2E-I4: Recent Advances in the Analysis of Complex Surveys Récents progrès en analyse d'enquêtes complexes

dent sample for the reference population when applying conventional statistical methods for survey data. In this talk, we focus on non-coverage bias and discuss several approaches, including inverse-probability-weighting and GREG-type calibration estimation, for bias adjustment. We present simulation results to illustrate the performance of the methods under different scenarios depending on the availability of additional information for the reference population. Possible directions for future work are discussed.

vent un traitement prudent pour tenir compte des biais potentiels résultant d'un déficit de représentativité de l'échantillon de répondants dans la population de référence. Dans le présent article, nous examinons le biais de non-couverture, et nous discutons diverses approches, y compris la pondération de probabilité inverse et l'estimation de calage (GREG) pour le redressement de biais. Nous présentons des résultats de simulation pour illustrer la performance des méthodes selon différents scénarios en fonction de la disponibilité d'information supplémentaire sur la population de référence. Nous discutons aussi des orientations possibles pour des travaux futurs.

SUNGHEE LEE, TUBA SUZER GURTEKIN & MICHAEL ELLIOTT (University of Michigan)

[Tuesday May 27 / mardi 27 mai, 16:30]

Sensitivity Analysis of Seed Selection in Respondent Driven Sampling

L'analyse de sensibilité de la sélection des répondants initiaux dans un échantillonnage fondé sur les répondants

Respondent-driven sampling (RDS) is a new sampling method proposed for rare populations. Unlike traditional sampling which requires extensive screening efforts, RDS exploits the population networks from selected seed respondents through incentivized coupons and samples from them in waves without screening. Due to pressing needs for studying rare populations, RDS is widely practiced. However, its formal base for inferences is relatively weak and sits on strong assumptions.

This paper tests sensitivity of the Markov process assumed for RDS, which states that as recruitment waves continue, the sample characteristics become independent of the seeds. We use a publicly-available RDS dataset targeting those at risk for HIV. Estimates of key variables from the full sample will be compared to those from simulated subsamples.

L'échantillonnage fondé sur les répondants (RDS) est une nouvelle méthode d'échantillonnage proposée pour les populations rares. Contrairement à l'échantillonnage traditionnel qui nécessite des efforts de dépistage, le RDS utilise les réseaux de la population à partir des répondants initiaux et au moyen de coupons d'incitation, et prend des vagues d'échantillons sans dépistage. En raison des besoins pressants, le RDS est fort répandu pour l'étude de populations rares. Pour l'inférence, par contre, sa base formelle est relativement faible et s'appuie sur des hypothèses fortes.

Le présent article décrit des tests de sensibilité du processus de Markov présumé pour le RDS, qui établit qu'à mesure que se poursuivent les vagues de recrutement, les caractéristiques d'échantillon deviennent indépendantes des valeurs initiales. Nous utilisons un ensemble de données RDS accessibles au public qui cible des sujets à risque pour le VIH. Des estimations de variables clés de l'échantillon complet sont comparées à celles de sous-échantillons simulés.

2E-I5: Statistical Challenges in Modern Astronomy Problèmes statistiques en astronomie moderne

Organizer and Chair / Responsable et président: Jason Loeppky (University of British Columbia)

Session sponsored by the Business and Industrial Statistics Section
Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: UC - 140

Abstracts/Résumés

EARL LAWRENCE & SCOTT VANDER WIEL (Los Alamos National Laboratory), **CASEY LAW** (University of California), **GEOFFREY BOWER** (Academia Sinica)

[Tuesday May 27 / mardi 27 mai, 15:30]

Searching for Fast Radio Bursts with the Very Large Array

À la recherche de sursauts radioélectriques rapides à l'aide du Very Large Array

Recently, astronomers have discovered the existence of fast radio bursts: single bright pulses of very short duration (5ms). The origin of these bursts is unknown. So far, they have only been detected on single-dish radio telescopes. Interferometers, like the Very Large Array, could provide better location accuracy, more efficient survey speeds, and more effective interference mitigation, but produce huge volumes of data at the time scales of interest (1TB/hour). Currently, this data is turned into images for offline detection. We describe the background and statistical work done to support this imaging. We will also touch on a different technique that handles the data stream in a more efficient manner and bypasses the computationally demanding imaging step.

Récemment, des astronomes ont découvert l'existence de sursauts radioélectriques rapides, soit de vives monoimpulsions de très courte durée (5 ms), dont l'origine est inconnue. Jusqu'à maintenant, ils ont seulement été détectés au moyen de radiotélescopes à réflecteur unique. Les interféromètres, comme le Very Large Array (VLA), permettent d'obtenir une meilleure précision d'emplacement, des vitesses de balayage plus rapides et une atténuation des interférences plus efficace, mais ils produisent d'énormes volumes de données à l'échelle de temps souhaitée (1 To/heure). Actuellement, ces données sont transformées en images permettant la détection hors ligne. Nous décrivons le contexte et les travaux statistiques nécessaires à la création de cette imagerie. Nous présentons aussi une autre technique qui gère le flux de données de façon plus efficace et contourne l'étape de l'imagerie très exigeante en matière de calculs.

SHANE REESE & BRITTANY SPENCER (Brigham Young University)

[Tuesday May 27 / mardi 27 mai, 16:00]

Hierarchical Gaussian Process Modeling of Supernova Light Curves

Modélisation gaussienne hiérarchique de courbes de lumière de supernova

With data collection projects such as the Dark Energy Survey underway, an area of great interest is the development of a strictly photometric classification mechanism. The first step in the advancement of modern photometric classification is the estimation of individual supernova light curves. We fo-

Compte tenu des projets de collecte de données en cours, comme le Dark Energy Survey (DES), la création d'un mécanisme de classification strictement photométrique suscite un grand intérêt. La première étape dans l'avancement de la classification photométrique moderne réside dans l'estimation des courbes de lumière des supernovas. Nous nous

2E-15: Statistical Challenges in Modern Astronomy Problèmes statistiques en astronomie moderne

cus on the use of hierarchical Gaussian processes to model light curves. Properties inherent in this Bayesian semiparametric form of modeling allow curve definition at a specific point to borrow information from neighboring points. Hierarchical modeling not only allows for individual curve definition, it also yields a meaningful representation of supernova class differences via mean curve estimation. The differences inherent in these mean curves may eventually contribute to a possible photometric classification mechanism for supernovae.

intéressons aux processus gaussiens hiérarchiques pour modéliser les courbes de lumière. Les propriétés inhérentes à ce type de modélisation semi-paramétrique bayésienne font en sorte que la définition de la courbe à un point spécifique peut capter de l'information des points voisins. En plus de permettre la définition des courbes de façon individuelle, la modélisation hiérarchique mène aussi à une représentation précise des différences entre les classes de supernovas par le biais de l'estimation moyenne de la courbe. Les différences inhérentes à ces courbes moyennes pourraient éventuellement contribuer à l'établissement d'un mécanisme de classification photométrique des supernovas.

ERIK ROSOLOWSKY (University of Alberta), **JASON LOEPPKY** (University of British Columbia Okanagan)
[Tuesday May 27 / mardi 27 mai, 16:30]

Critiquing the Statistical Analysis of Star Formation
Critique de l'analyse statistique de la formation d'étoiles

In this talk, we will introduce using designed experiments to critically evaluate the statistical tools used in the study of star formation. Star formation drives galaxy evolution, and one window into this process is the structure of the molecular clouds hosting the star formation. Cloud structure is shaped by many competing physical processes. Astronomers have employed a plethora of statistical tools to isolate the dominant physics. However, not all these tools are as effective as claimed. We have completed a designed numerical experiment, modelling the structure of star forming clouds subject to different physics. By testing which tools respond to the varying physical inputs, we can isolate those techniques best able to unravel this problem.

Dans cet exposé, nous proposons l'utilisation de plans d'expériences pour évaluer de façon critique les outils statistiques utilisés dans l'étude de la formation d'étoiles. La formation d'étoiles est à l'origine de l'évolution des galaxies et la structure des nuages moléculaires à la base de la formation stellaire constitue une fenêtre ouverte sur ce processus. La structure des nuages est façonnée par une foule de processus physiques concurrents. Les astronomes ont employé une pléthore d'outils statistiques pour isoler les aspects physiques dominants. Cependant, certains de ces outils ne sont pas aussi efficaces qu'ils le devraient. Nous avons réalisé un plan d'expérience numérique en modélisant la structure des nuages responsables de la formation stellaire qui sont soumis à différentes variables physiques. En évaluant quels sont les outils qui répondent à ces entrées physiques variables, nous pouvons identifier les meilleures techniques pour résoudre cette problématique.

2E-I6: Statistics of Sea Surface Winds and Non-Gaussianity in Climate Variability Statistique des vents de surface de la mer et non-gaussianité de la variabilité climatique

Organizer and Chair / Responsable et présidente: Priscilla Greenwood (University of British Columbia)

Room/Salle: UC - 179

Abstracts/Résumés

ADAM MONAHAN (University of Victoria)

[Tuesday May 27 / mardi 27 mai, 15:30]

Stochastic Dynamics of Near-Surface Winds

Les données stochastiques dynamiques des vents de surface

Understanding the probability distribution of land surface wind speeds is important for characterizing wind extremes; calculating exchanges of momentum, mass, and energy between the atmosphere and the underlying surface; and managing wind power resources. Consequently, there is a long history of modelling surface winds with empirical distributions fit to (often quite limited) observations. The past decade has seen the beginning of a sustained effort to develop physically-based models of wind speed probability distributions, informed by long-duration and high-resolution data which have recently become available. This talk will describe how statistics, stochastic dynamics, and atmospheric boundary-layer physics can be combined to deepen our understanding of near surface wind variability over both land and the ocean.

La connaissance de la distribution des vents de surface est importante pour caractériser les vents extrêmes, calculer les échanges de mouvement, de masse et d'énergie entre l'atmosphère et la surface sous-jacente, et gérer les ressources éoliennes. Par conséquent, les observations comprennent un long historique de modélisation des vents de surface avec des distributions empiriques ajustées aux observations (souvent limitées). La dernière décennie a été témoin d'un effort soutenu de développement de modèles basés sur la physique des lois de probabilité des vitesses de vent, appuyés sur des données à forte résolution et de longue durée qui sont devenues accessibles. Le présent exposé décrit la façon dont la statistique, la dynamique stochastique et la physique de couche limite atmosphérique peuvent être combinées pour approfondir notre compréhension de la variabilité des vents près de la surface par rapport aux vents provenant de la terre ou de l'océan.

PHILIP SURA (Florida State University), **ABDEL HANNACHI** (Stockholm University)

[Tuesday May 27 / mardi 27 mai, 16:00]

Perspectives of Non-Gaussianity in Weather and Climate Variability

Perspectives des statistiques non gaussiennes sur la variabilité des phénomènes météorologiques et climatiques

Understanding non-Gaussian statistics of weather and climate variability has important consequences in the atmospheric and ocean sciences because weather and climate risk assessment depends on knowing and understanding the exact shape of the system's probability density function. While there is no doubt that many atmospheric variables exhibit non-Gaussian statistics on many time (and spatial) scales a full and complete understanding of this phenomenon remains a challenge.

La compréhension des statistiques non gaussiennes sur la variabilité des phénomènes météorologiques et climatiques a des conséquences importantes sur les sciences atmosphérique et océanographique, étant donné que l'évaluation des risques météorologiques et climatiques dépend de la connaissance et de la compréhension de la forme exacte de la fonction de densité du système. Même s'il est certain que bon nombre de variables atmosphériques présentent des statistiques non gaussiennes à plusieurs échelles temporelles (et spatiales), une compréhension approfondie et

2E-I6: Statistics of Sea Surface Winds and Non-Gaussianity in Climate Variability Statistique des vents de surface de la mer et non-gaussianité de la variabilité climatique

Various mechanisms behind the observed non-Gaussian statistics have been proposed but re-

main, however, multifaceted and scattered in the literature. Given the importance of this subject for climate research, a thorough review and discussion of the different mechanisms that lead to non-Gaussian weather and climate variability are presented and an outlook is given.

complète de ce phénomène reste encore un défi.

Divers mécanismes sous-jacents aux statistiques observées non gaussiennes ont été proposés, mais ils comportent de multiples facettes et restent très épars dans la littérature spécialisée. Compte tenu de l'importance de ce sujet dans la recherche climatique, l'exposé présente un examen approfondi des différents mécanismes qui mènent à une variabilité non gaussienne des phénomènes météorologiques et climatiques, suivi d'une discussion et de perspectives sur le sujet.

WILLIAM THOMPSON (University of British Columbia), **ADAM H. MONAHAN** (University of Victoria), **DAAN CROMMELIN** (Centrum Wiskunde and Informatica)

[Tuesday May 27 / mardi 27 mai, 16:30]

Parametric Estimation of Sea Surface Wind Dynamics via the Infinitesimal Generator

Estimation paramétrique de la dynamique des vents de surface de la mer par le générateur infinitésimal

Estimating parameters for dynamical models from data is a ubiquitous problem in the physical sciences. When stochasticity is an intrinsic property of the assumed dynamics, the problem is more com-

plicated.

This talk will present a method for parametrically fitting time series data to a particular stochastic differential equation model. This method operates on variational principles, finding the parameters for a particular stochastic model giving the best correspondence between the model and the data using

the eigenstructure of the infinitesimal generator.

After presenting this method, I will display some results we have obtained by applying it to global sea surface wind data to generate parameter fields for a particular stochastic wind model.

L'estimation des paramètres pour des modèles dynamiques est un problème omniprésent dans les sciences physiques. Si la stochasticité est une caractéristique de la dynamique présumée, le problème est plus compliqué.

Cette présentation discute d'une méthode pour l'estimation des paramètres pour un modèle particulier d'équation dif-

férentielle stochastique. Cette méthode est variationnelle, utilisant la structure propre du générateur infinitésimal pour trouver les paramètres pour un modèle stochastique particulier qui donne la meilleure correspondance entre le modèle et les données.

Suivant la présentation de la méthode, je présente ensuite des résultats que nous avons obtenus par l'application aux

données globales des vents de surface de la mer pour produire des champs de paramètres pour un modèle stochastique particulier du vent.

2E-C1: Actuarial Science and Finance 2 Actuariat et finance 2

Chair/Présidente: Alison Gibbs (University of Toronto)

Room/Salle: UC - 152

Abstracts/Résumés

DEZHAO HAN, KEN SENG TAN & CHENGGUO WENG (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 15:30]

Sparse Vine Copula with Applications to Financial Data

Applications des copules en vignes parcimonieuses à des données financières

Recently vine copula is receiving increasing attention due to its flexibility in modeling high dimensional dependence of financial data. It can be used to capture the asymmetry and heavy-tailedness in a joint distribution. Nevertheless, its usability comes along with an exponentially increasing complexity in dimension. To develop more parsimonious models, we propose a regularization method to reduce the number of parameters, where those bivariate copulas with insignificant dependence are shrunk to be independence. The resulting model is called "sparse vine copula". An application to several financial datasets consistently confirms the superiority of our sparse vine copula to the normal vine copula.

Les copules en vignes ont récemment reçu une attention grandissante en raison de leur flexibilité dans la modélisation de la dépendance en grande dimension, notamment pour les données financières. Elles peuvent être utilisées pour illustrer l'asymétrie et la lourdeur de queue d'une distribution conjointe. Pourtant, leur facilité d'utilisation s'accompagne d'une complexité s'accroissant exponentiellement en fonction du nombre de dimensions. Pour élaborer des modèles plus parcimonieux, nous proposons une méthode de régularisation pour réduire le nombre de paramètres, où les copules bivariées avec dépendance non significative sont réduites à l'état d'indépendance. Le modèle qui en résulte est désigné «copule en vignes parcimonieuse». Une application à divers ensembles de données confirme de façon constante la supériorité de notre copule en vignes parcimonieuse par rapport à la copule en vignes normale.

ZIED BEN SALAH (Concordia University)

[Tuesday May 27 / mardi 27 mai, 15:45]

On a Generalization of the Expected Discounted Penalty Function to Include Deficits at and Beyond Ruin

Généralisation de la fonction de pénalité escomptée prévue pour tenir compte des déficits au moment de la faillite et au-delà

In this paper we propose an extended concept of the Expected Discounted Penalty Function (EDPF) that takes into account new ruin-related random variables. We add to the EDPF, which was introduced in classical papers [Gerber and Shiu (1997), (1998) and Gerber and Landry (1998)], a sequence of expected discounted functions of new record minima reached by a jump of the risk process after ruin. Inspired by results in fluctuation theory for Lévy processes, we provide a characterization for this extended EDPF in a setting involving a cu-

Dans cet article, nous proposons un concept étendu de la fonction de pénalité escomptée prévue (Expected Discounted Penalty Function, ou EDPF) qui tient compte de nouvelles variables aléatoires liées à la défaillance. Nous ajoutons à la fonction EDPF, introduite dans des articles classiques [Gerber et Shiu (1997), (1998) et Gerber et Landry (1998)], une séquence de fonctions escomptées prévues de nouveaux minimums absolus obtenus par un saut du processus de risque après la faillite. Inspirés par des résultats en théorie de la fluctuation pour les processus de Levy, nous proposons de caractériser cette fonction EDPF

2E-C1: Actuarial Science and Finance 2

Actuariat et finance 2

mulative claims modeled by a subordinator, and Brownian perturbation. We illustrate how the extended EDPF can be used to compute the expected discounted value of capital injections (EDVCI) for Brownian perturbed risk model.

étendue dans un environnement avec des réclamations cumulées modélisées par un subordonateur et une perturbation brownienne. Nous illustrons comment cette fonction EDPF étendue peut servir à calculer la valeur escomptée prévue d'injections de capital pour un modèle de risque perturbé brownien.

JIANXI SU & EDWARD FURMAN (York University)

[Tuesday May 27 / mardi 27 mai, 16:15]

On a Multiple-Shock Dependence Structure
Structure de dépendance multi-chocs

Let $E_j \sim \text{Exp}(\lambda_j (\in \mathbf{R}_+))$, $j = 1, \dots, n (n \in \mathbf{N}) + 1$ denote independent 'hitting' times of one systemic and n residual risk factors. Also, imagine a risk portfolio of lifetimes $\mathbf{X} = (X_1, \dots, X_n)'$, such that $X_i = E_i \wedge E_{n+1}$, $i = 1, \dots, n$. I will start off by observing, in the context of the aforementioned dependence structure, that: 1) hitting times are either fully co-monotonic or independent, and 2) Pearson's and Kendall's correlations are both equal to the probability of simultaneous death (> 0). I will then introduce a new multiple-shock (MS) dependence structure that overcomes these issues. The proposed MS formation is able to model non-exponential dependent hitting times, as well as allows for zero and non-zero probabilities of simultaneity death. Furthermore, this approach gives birth to a new family of copula functions.

Soient $E_j \sim \text{Exp}(\lambda_j (\in \mathbf{R}_+))$, $j = 1, \dots, n (n \in \mathbf{N}) + 1$ les cas de réalisations indépendantes d'un facteur de risque systémique et de n facteurs de risque résiduels. Soit aussi un portefeuille de risques de durées de vie $\mathbf{X} = (X_1, \dots, X_n)'$, tel que $X_i = E_i \wedge E_{n+1}$, $i = 1, \dots, n$. Je commence par observer que dans le contexte de la structure de dépendance susmentionnée : 1. Les cas de réalisation sont soit entièrement comonotoniques, soit indépendants, et 2. Les corrélations de Pearson et de Kendall sont égaux à la probabilité de mort simultanée (> 0). J'introduis ensuite une structure de dépendance multi-chocs (MC) qui surmonte ces problèmes. La formation MC proposée est capable de modéliser des cas de réalisations dépendants non exponentiels et permet des probabilités de mort simultanée nulles et non nulles. De plus, l'approche donne lieu à une nouvelle famille de fonctions de copules.

JULIEN TRUFIN & ILIE-RADU MITRIC (Laval University)

[Tuesday May 27 / mardi 27 mai, 16:30]

On a Risk Measure Inspired from the Ruin Probability and the Expected Deficit at Ruin
Une mesure du risque inspirée de la probabilité de faillite et du déficit prévu à la faillite

In this talk, we study a risk measure derived from ruin theory, defined as the amount of capital needed to cope in expectation of the first occurrence of a ruin event. Specifically, within the compound Poisson model, we investigate some properties of this risk measure with respect to the stochastic ordering of claim severities. Particular situations, where combining risks yield diversification benefits, are identified. Closed form expressions and upper bounds are also provided for certain claim severities.

Dans cette présentation, nous étudions une mesure du risque dérivée de la théorie de la faillite, définie comme le capital requis pour faire face à l'attente de la première occurrence d'un événement de faillite. Spécifiquement, dans le cadre du modèle de Poisson composé, nous étudions certaines propriétés de cette mesure du risque ayant trait à l'ordonnement stochastique de la gravité des réclamations. Nous identifions des situations particulières où la combinaison de risques donne des avantages en termes de diversification. Nous proposons des expressions closes et des limites supérieures pour certaines gravités de réclamations.

2E-C2: Biostatistics - New Inferential Strategies Biostatistique - nouvelles stratégies inférentielles

Chair/Président: Paul Peng (Queen's University)

Room/Salle: UC - 244

Abstracts/Résumés

PARISA GAVANJI, BINGSHU CHEN & WENYU JIANG (Queen's university)

[Tuesday May 27 / mardi 27 mai, 15:30]

Residual Bootstrap for Testing the Treatment and Biomarker Interaction for Unknown Cut-off Point

Bootstrap résiduel pour tester le traitement et l'interaction des biomarqueurs pour une limite inconnue

Many molecularly targeted treatments tend to benefit a subset of patients, known as the sensitive subset, with certain biomarker features. We are interested in testing the treatment and biomarker interaction effect on survival responses based on the Cox's proportional hazards model when a biomarker is identified, but the cut-point defining the sensitive patient subset is unknown. We propose using residual bootstrap method that resamples the residuals from the model allowing joint consideration of treatment and biomarker in the Cox regression context. The residual bootstrap and permutation methods are evaluated in a simulation study by comparing the size and power of the test. The simulation study shows that the residual bootstrap works better when treatment assignments vary with patients' biomarker values.

Plusieurs traitements moléculaires ciblés ont tendance à bénéficier à un sous-ensemble de patients, connu sous le nom de sous-ensemble sensible, possédant certains biomarqueurs. Nous sommes intéressés à tester l'effet du traitement et l'interaction des biomarqueurs sur la survie en nous basant sur le modèle à risques proportionnels de Cox quand un biomarqueur est identifié, mais le point de séparation qui définit le sous ensemble de patients sensibles est inconnu. Nous proposons l'utilisation d'une méthode de bootstrap résiduel qui fait un rééchantillonnage des résiduels du modèle ce qui permet un examen conjoint du traitement et du biomarqueur dans le contexte de régression de Cox. Les méthodes de bootstrap résiduel et de permutation sont évaluées dans une étude de simulation en comparant la taille et la puissance du test. L'étude de simulation démontre que le bootstrap résiduel fonctionne mieux quand les affectations du traitement changent avec les valeurs des biomarqueurs des patients.

NATHALIE MOON, RICHARD COOK & LEILEI ZENG (University of Waterloo)

[Tuesday May 27 / mardi 27 mai, 15:45]

The Impact Of Prevalent Cohort Sampling on Multistate Transition Probabilities in Chronic Disease Processes

L'impact de l'échantillonnage de cohortes prévalentes sur les probabilités de transition dans des modèles multi-états de maladies chroniques

Prevalent cohort studies sample individuals from a population with a probability proportional to their survival time. Considerable attention has been given to the impact of this sampling scheme on inferences about survival, but relatively little work has been done on more general multistate models of complex disease processes. We derive the length-biased transition probability matrix for a K-state Markov disease process and compare vari-

Dans une étude de cohorte prévalente, la probabilité qu'un individu soit sélectionné pour l'échantillon est proportionnelle à son temps de survie. Une attention considérable a été accordée à l'impact de ce plan d'échantillonnage sur des inférences concernant la survie, mais relativement peu d'attention a été accordée à de modèles multi-états plus généraux pour la progression de maladies. Nous dérivons les probabilités de transition pour un modèle de Markov multi-état avec biais de longueur et comparons plusieurs caractéristiques.

2E-C2: Biostatistics - New Inferential Strategies Biostatistique - nouvelles stratégies inférentielles

ous features to the population-level counterparts. The effect of length bias on several quantities of interest will be discussed along with suitable methods of analysis to address this problem.

téristiques avec leurs homologues au niveau de la population. L'effet du biais de longueur sur différentes caractéristiques sera discuté, ainsi que des méthodes appropriées pour adresser ce problème.

QIAN SHI & LINGLONG KONG (University of Alberta)

[Tuesday May 27 / mardi 27 mai, 16:00]

Variable Screening Based On Combining Quantile Regression

Filtrage de variables par régression quantile combinée

We introduce a quantile regression framework for linear and nonlinear variable screening with high-dimensional heterogeneous data. Inspired by the efficiency of various variable screening methods, especially the quantile-adaptive framework, we propose a more efficient variable screening procedure by combining the information from different quantile levels. There are two ways to do it: one is to simply take (weighted) average across different levels of quantile regression; the other one is to use (weighted) composite quantile regression. Asymptotically, these two methods are equivalent with respect to efficiency. The fine performance of the proposed methods for linear and nonlinear models are confirmed by numerical studies. This is joint work with Linglong Kong.

Nous introduisons un cadre de régression quantile pour filtrer des variables linéaires et non linéaires avec des données hétérogènes de grandes dimensions. Inspirés par l'efficacité de diverses méthodes de filtrage de variables, notamment le cadre s'adaptant aux quantiles, nous proposons une procédure de filtrage de variables plus efficace en combinant les informations de plusieurs niveaux de quantiles. On peut procéder de deux manières : prendre une moyenne (pondérée) sur différents niveaux de régression quantile ou utiliser une régression quantile composée (pondérée). D'un point de vue asymptotique, les deux méthodes sont équivalentes en ce qui concerne l'efficacité. La bonne performance des méthodes proposées pour les modèles linéaires et non linéaires est confirmée par des études numériques. Ces travaux ont été effectués en collaboration avec Linglong Kong.

YUKUN ZHANG & XUEWEN LU (University of Calgary)

[Tuesday May 27 / mardi 27 mai, 16:15]

Variable Selection in Log-linear Birnbaum-Saunders Regression Model for High-dimensional Survival Data via the Elastic Net and Stochastic EM

Sélection de variables dans le modèle de régression de Birnbaum-Saunders loglinéaire pour les données de survie de grandes dimensions via le filet élastique et Stochastic EM

Birnbaum-Saunders (BS) distribution is broadly used to model failure times in reliability and survival analysis. In this paper, we propose a variable selection procedure in a log-linear BS regression model for high-dimensional survival data. We introduce a path wise algorithm via cyclical coordinate descent based on the elastic net penalty. To deal with censored survival data, we iteratively run a combination of stochastic EM algorithm (StEM) and variable selection procedure to generate pseudo-complete data and select variables until convergence. Treating pseudo-complete data as uncensored data via StEM simplifies computa-

La distribution de Birnbaum-Saunders (BS) est communément utilisée pour modéliser les temps de défaillance dans l'analyse de fiabilité et de survie. Dans cette présentation, nous proposons une procédure de sélection de variables dans un modèle de régression BS loglinéaire pour les données de survie de grandes dimensions. Nous introduisons un algorithme par chemin de descente cyclique des coordonnées basé sur la pénalité en filet élastique. Pour traiter les données de survie censurées, nous exécutons de manière itérative une combinaison d'algorithme stochastique EM (StEM) et de procédure de sélection de variables pour générer des données pseudo-complètes et sélectionner des variables jusqu'à la convergence. Les données pseudo-

2E-C2: Biostatistics - New Inferential Strategies Biostatistique - nouvelles stratégies inférentielles

tion and makes it possible to incorporate iterative least squares for parameter estimation and variable selection simultaneously. We demonstrate the efficacy of our method using simulated and real data sets.

complètes sont traitées comme des données non censurées via StEM pour simplifier le calcul et inclure les moindres carrés itératifs pour l'estimation des paramètres et la sélection de variables en même temps. Nous prouvons l'efficacité de notre méthode à l'aide d'ensembles de données simulés et réels.

MICHAEL WALLACE & ERICA MOODIE (McGill University)

[Tuesday May 27 / mardi 27 mai, 16:30]

A New Method for Identifying Dynamic Treatment Regimes: Introducing weighted Q-learning

Nouvelle méthode d'identification des régimes de traitement dynamiques : Q-learning pondéré

As populations age and chronic disease prevalence increases, there is a growing demand for long-term, personalized, health management plans. Dynamic treatment regimes cater to this by using patient covariates and treatment history to recommend sequences of treatment decision rules which optimize patient outcome. Methods including Q-learning, G-estimation and augmented inverse probability of treatment weighted (IPTW) estimators have been developed for identification of such strategies. Building on these, we present a new alternative: 'weighted Q-learning', boasting a double-robustness similar to that of G-estimation and augmented IPTW, but with easier implementation akin to standard Q-learning. Along with outlining the underlying theory, we illustrate the method via simulation.

À mesure que les populations vieillissent et que les maladies chroniques deviennent plus fréquentes, la demande de plans de gestion de la santé personnalisés à long terme augmente. Les régimes de traitement dynamiques répondent à cette demande en incluant les covariables et les antécédents de traitement du patient pour recommander des séquences de règles de décision de traitement afin d'optimiser les résultats pour le patient. Diverses méthodes, comme le Q-learning, la G-estimation et les estimateurs IPTW (inverse probability of treatment weighted) augmentés ont été mises au point pour identifier ces stratégies. Nous présentons une nouvelle alternative à celles-ci : le Q-learning pondéré, dont la double robustesse est similaire à celle de la G-estimation et des estimateurs IPTW augmentés, mais qui est aussi facile à mettre en œuvre que le Q-learning standard. Nous en proposons la théorie sous-jacente et l'illustrons via simulation.

TOLULOPE SAJOBI (University of Calgary), **LISA M. LIX** (University of Manitoba), **GURBAKHSHASH SINGH & GORDON H. FICK** (University of Calgary)

[Tuesday May 27 / mardi 27 mai, 16:45]

Statistical Inference on Discriminant Function Coefficients in the Presence of Non-Normality and Covariance Heterogeneity

Inférence statistique sur les coefficients de fonctions discriminantes en la présence de non-normalité et d'hétérogénéité de la covariance

Discriminant function coefficients are used to describe the relative importance of multiple variables in distinguishing two independent groups. Test procedures due to Das Gupta (1968), Bodnar & Okhirin (2011), and Rao (1970), as well as the bootstrap have been proposed for evaluating statistical significance of variable importance. However, these procedures assume multivariate normality and equal group covariances, two assump-

Les coefficients de fonctions discriminantes sont utilisés pour décrire l'importance relative de multiples variables pour distinguer entre deux groupes indépendants. Des procédures de vérification proposées par Das Gupta (1968), Bodnar et Okhirin (2011) et Rao (1970), ainsi que le bootstrap, permettent d'évaluer la pertinence statistique de l'importance des variables. Toutefois, ces procédures supposent une normalité multivariée et des covariances de groupes égales, hypothèses peu susceptibles d'être satisfaites dans

2E-C2: Biostatistics - New Inferential Strategies Biostatistique - nouvelles stratégies inférentielles

tions which are unlikely to be satisfied in practice. This study compares Type I error and statistical power rates of these procedures under non-normal data distributions and heterogeneous covariances using Monte Carlo methods and offers recommendations on the optimal procedure.

la pratique. Cette présentation compare l'erreur de type I et le taux d'efficacité statistique de ces procédures en cas de distribution des données non-normale et de covariances hétérogènes à l'aide de méthodes de Monte Carlo ; elle propose des recommandations concernant la procédure optimale.

2E-C3: Business and Industrial Statistics 2 Statistique industrielle et de gestion 2

Chair/Président: Derek Bingham (Simon Fraser University)

Room/Salle: UC - 256

Abstracts/Résumés

ARDEN EWAN MILLER (University of Auckland)

[Tuesday May 27 / mardi 27 mai, 15:30]

Error Rates for the Analysis of Unreplicated Experiments

Taux d'erreurs pour l'analyse de plans sans réplicats

Many model selection procedures that have been proposed for the analysis of unreplicated experiments have been devised to control either the individual error rate (IER) or the experimentwise error rate (EER) at a stated level under the null model. In general, the values of the IER and the EER for a given selection procedure will depend on the number of active effects and on the magnitudes of these effects relative to the magnitude of experimental variability. This presentation discusses establishing an upper limit on the error rate for a given method.

Bon nombre des procédés de sélection de modèles qui ont été proposés pour l'analyse de plans sans réplicats ont été conçus pour contrôler soit le taux d'erreur individuel (TEI) ou le taux d'erreur inhérent à l'expérience (TEE) et le maintenir à un niveau stipulé sous l'hypothèse nulle. En général, les valeurs de TEI ainsi que le TEE d'un processus de sélection donné dépendront du nombre d'effets actifs et de leur ordre de grandeur en comparaison avec celui de la variabilité expérimentale. Cet exposé aborde l'établissement d'une limite supérieure de taux d'erreur pour une méthode quelconque.

PAULE BODSON-CLERMONT (Université de Montréal), **FRANÇOIS LAFLEUR** (Institut de recherche d'Hydro-Québec), **JEAN-FRANÇOIS ANGERS** (Université de Montréal), **LUC PERREAULT** (Institut de recherche d'Hydro-Québec)

[Tuesday May 27 / mardi 27 mai, 15:45]

Statistical Modeling of Cavitation Erosion within the Operation Parameters of a Hydraulic Turbine

Modélisation de l'érosion de cavitation d'une turbine hydraulique selon les paramètres d'opération

Cavitation resulting from the implosion of vapor cavities leads to severe erosion of the materials that compose hydraulic turbines. A system to detect erosion using vibration measurements has been installed on 4 hydropower units and the data was collected. The preliminary analysis enhanced the complexity of the phenomenon. Indeed, changes in the relationship between cavitation and various operating variables were observed and could be due to a seasonal behavior or different operating conditions. Using hierarchical clustering and regression models, we formalize this heterogeneity by developing a model which includes operating variables such as active power, tail water level and gate opening.

Dans une turbine hydraulique, la rotation des aubes dans l'eau crée une zone de basse pression, amenant l'eau à passer de l'état liquide à l'état gazeux. L'implosion des cavités de vapeur qui en résultent mène à une érosion sévère des matériaux. Un système de détection d'érosion à l'aide de mesures vibratoires a donc été installé sur 4 groupes turbine-alternateur et les données ont été recueillies. Les premières analyses ont mis en relief l'hétérogénéité des observations de cavitation, qui semblent provenir de populations statistiques distinctes. Pour expliciter formellement cette hétérogénéité, un modèle statistique liant les variables opératoires et l'érosion de la turbine est développé, en s'appuyant entre autres sur le regroupement hiérarchique et les modèles de régression.

2E-C3: Business and Industrial Statistics 2

Statistique industrielle et de gestion 2

JANIE COULOMBE (Université de Montréal)

[Tuesday May 27 / mardi 27 mai, 16:00]

Assessment of Wind Speed Prediction Error in Wind Energy

Évaluation de l'erreur de prévision de la vitesse du vent en énergie éolienne

Wind power is generated by wind turbines. To assess the wind power potential of a particular site, wind data is collected from anemometers during a short period of time (e.g. one year) and the wind turbine engineers use long-term reference datasets to predict wind speed over a longer time period. Some of these datasets represent real data, collected far from the site and others, simulated data from a mesoscale model, specifically tailored for the site. We will assess the usefulness of mesoscale data in wind speed prediction using cross-validation and will use block bootstrap to assess the accuracy of the ten-year wind speed average prediction.

L'énergie éolienne est produite à partir du vent. Pour évaluer le potentiel éolien d'un site, des données de vent sont récoltées par des anémomètres sur une courte période de temps et les ingénieurs utilisent d'autres jeux de données de référence à plus long terme afin de prédire les vitesses de vent sur une plus grande échelle de temps. Certains de ces jeux représentent de vraies données collectées à des mâts éloignés et d'autres, des données de simulation méso-échelle pour l'emplacement exact du site. On évaluera l'utilité des données de type méso-échelle dans la prévision des vents à partir de la validation croisée et le bootstrap par bloc permettra d'évaluer la précision de la prévision du vent moyen sur dix ans.

PAVEL KRUPSKIY & HARRY JOE (University of British Columbia)

[Tuesday May 27 / mardi 27 mai, 16:15]

Structured Factor Copula Models for Multivariate Data

Modèles de copule pour des données multivariées avec facteurs structurés

We propose two structured factor copula models for the case where variables can be split into non-overlapping groups such that there is homogeneous dependence within each group. The models are general conditional independence models for d observed variables when dependence is modelled via common and group latent factors. The structured factor copula models can handle multivariate data with tail dependence and tail asymmetry, properties that the multivariate normal copula does not possess. It is a good choice for modelling high-dimensional data as a parametric form can be specified to have $O(d)$ dependence parameters instead of $O(d^2)$ parameters.

Nous proposons deux modèles de copule avec des facteurs structurés pour le cas où des variables peuvent être divisées en groupes qui ne se recoupent pas, de sorte qu'il en résulte une dépendance homogène à l'intérieur de chaque groupe. Les modèles sont des structures d'indépendance conditionnelle générales pour d variables observées dont la dépendance est modélisée au moyen de facteurs latents communs et de groupes. Les modèles de copule à facteurs structurés peuvent traiter des données multivariées avec une dépendance codale et une asymétrie codale, contrairement à la copule normale multivariée. Ces modèles constituent de bons choix pour la modélisation de données de grande dimension, étant donné qu'il est possible de spécifier une forme paramétrique pour obtenir des paramètres de dépendance $O(d)$ au lieu des paramètres $O(d^2)$.

PAULA MURRAY, RYAN BROWNE & PAUL MCNICHOLAS (University of Guelph)

[Tuesday May 27 / mardi 27 mai, 16:30]

The Hidden Truncation Hyperbolic Distribution

Distribution hyperbolique d'une troncature dissimulée

2E-C3: Business and Industrial Statistics 2

Statistique industrielle et de gestion 2

We introduce a novel formulation of the hyperbolic distribution which we refer to as the hidden truncation hyperbolic (HTH) distribution. The HTH distribution encapsulates several distributions as special or limiting cases including the skew- t and skew-normal distributions. An expectation-maximization (EM) algorithm for the estimation of model parameters is presented. We discuss the ability of the HTH distribution to model asymmetric data as compared to other skewed distributions which have appeared in the literature.

Nous présentons une formulation nouvelle de la distribution hyperbolique, que nous appelons distribution hyperbolique de troncature dissimulée (HTD). La distribution HTD englobe bon nombre de distributions comme cas limites ou particuliers, y compris la distribution asymétrique de Student et la normale. Un algorithme espérance-maximisation (EM) est présenté pour l'estimation des paramètres du modèle. Nous comparons la capacité de la distribution HTD de modéliser des données asymétriques avec d'autres distributions asymétriques présentes dans la littérature sur le sujet.

2E-C4: Probability Theory and Finance Théorie des probabilités et finance

Chair/Président: Deniz Sezer (University of Calgary)

Room/Salle: UC - Croft House

Abstracts/Résumés

KATHERINE DAVIES (University of Manitoba), **N. BALAKRISHNAN** (McMaster University)

[Tuesday May 27 / mardi 27 mai, 15:30]

Pitman Closeness Results for Type-I Censored Data from Exponential Distribution

Résultats sur le critère de proximité de Pitman pour des données censurées de type-I provenant d'une distribution exponentielle

Recent work on Pitman closeness has compared estimators under Type-II censored samples from exponential distribution. It was shown that a Type-II censored sample of size r' from n units results in a Pitman closer estimator of the exponential mean θ than a Type-II censored sample of size r , where $r < r'$. The natural question that arises is whether an analogous result holds for the Type-I censoring scenario, i.e., whether the estimator of θ based on a Type-I censored experiment with censoring time T^* is Pitman closer than the estimator based on the same experiment but with censoring time $T < T^*$. In this talk we answer this by deriving the Pitman closeness probability between two estimators based on the two Type-I censoring schemes.

Une étude récente sur le critère de proximité de Pitman a comparé des estimateurs d'échantillons censurés de type-II provenant d'une distribution exponentielle. Il a été démontré qu'un échantillon censuré de type-II de taille r' de n unités produit une estimation de la moyenne exponentielle θ plus précise au sens de Pitman qu'un échantillon censuré de type-II de taille r , où $r < r'$. Il convient alors de s'interroger pour déterminer si un scénario censuré de type-I produit un résultat analogue, c'est-à-dire : l'estimateur de θ basé sur une expérience censurée de type-I avec censure de temps T^* est-il plus précis au sens de Pitman que l'estimateur basé sur une même expérience mais avec une censure de temps $T < T^*$. Dans cet exposé, nous répondons à cette interrogation en dérivant la probabilité de proximité de Pitman entre deux estimateurs basés sur les deux schémas de censure de type-I.

DELI LI (Lakehead University)

[Tuesday May 27 / mardi 27 mai, 15:45]

A Characterization of Chover-Type Law of Iterated Logarithm

Caractérisation de la loi du logarithme itéré de type Chover

Let $0 < \alpha \leq 2$ and $-\infty < \beta < \infty$.

Let $\{X_n; n \geq 1\}$ be a sequence of independent copies of a real-valued random variable X and set $S_n = X_1 + \dots + X_n$, $n \geq 1$. We say X satisfies the (α, β) -Chover-type law of the iterated logarithm (and write $X \in CTLIL(\alpha, \beta)$) if $\limsup_{n \rightarrow \infty} \left| \frac{S_n}{n^{1/\alpha}} \right|^{(\log \log n)^{-1}} = e^\beta$ almost surely. This paper is devoted to a characterization of $X \in CTLIL(\alpha, \beta)$.

Soient $0 < \alpha \leq 2$ et $-\infty < \beta < \infty$.

Soit $\{X_n; n \geq 1\}$, une séquence de copies indépendantes d'une variable aléatoire à valeurs réelles X , et définissons $S_n = X_1 + \dots + X_n$, $n \geq 1$. Alors nous disons que X satisfait à la loi du logarithme itéré de type Chover (α, β) (et écrivons que $X \in CTLIL(\alpha, \beta)$) si $\limsup_{n \rightarrow \infty} \left| \frac{S_n}{n^{1/\alpha}} \right|^{(\log \log n)^{-1}} = e^\beta$ presque sûrement. Cet exposé porte sur la caractérisation de $X \in CTLIL(\alpha, \beta)$.

2E-C4: Probability Theory and Finance Théorie des probabilités et finance

ÉRIC MARCHAND & DJILALI AIT AOUDIA (Université de Sherbrooke), **FRANÇOIS PERRON** (Université de Montréal), **LATIFA BEN HADJ SLIMENE** (Université de Sherbrooke)

[Tuesday May 27 / mardi 27 mai, 16:00]

On Runs, Bivariate Poisson Mixtures and Distributions That Arise in Bernoulli Arrays

À propos du nombre de succès consécutifs, de mélanges de lois de Poisson bivarié et de tableaux de Bernoulli

Distributional findings are obtained for quantities arising in Bernoulli arrays $\{X_{k,j}, k \geq 1, j = 1, \dots, r+1\}$, with rows $(X_{k,1}, \dots, X_{k,r+1})$ independently distributed as Multinomial $(1, p_{k,1}, \dots, p_{k,r+1})$ for $k \geq 1$ and homogeneity across the first r columns $p_{k,1} = \dots = p_{k,r}$. The quantities of interest relate to the number of runs of length 2 and are $\underline{S}_n = (S_{n,1}, \dots, S_{n,r})$, $\underline{S} = \lim_{n \rightarrow \infty} \underline{S}_n$, $T_n = \sum_{j=1}^r S_{n,j}$, and $T = \lim_{n \rightarrow \infty} T_n$, where

$S_{n,j} = \sum_{k=1}^n X_{k,j} X_{k+1,j}$. With various known results applicable to the marginals $S_{n,j}$'s and to their limiting quantities $S_j = \lim_{n \rightarrow \infty} S_{n,j}$, we investigate joint distributions in the bivariate ($r = 2$) case and the distributions of their totals T_n and T for $r \geq 2$.

Nous étudions des lois de probabilité survenant pour des tableaux de Bernoulli $\{X_{k,j}, k \geq 1, j = 1, \dots, r+1\}$, avec lignes $(X_{k,1}, \dots, X_{k,r+1})$ indépendamment distribuées de lois Multinomiale $(1, p_{k,1}, \dots, p_{k,r+1})$ pour $k \geq 1$ et sous l'homogénéité des colonnes : $p_{k,1} = \dots = p_{k,r}$. Les variables d'intérêt se rapportent aux nombres de succès consécutifs de longueur 2 sont : $\underline{S}_n = (S_{n,1}, \dots, S_{n,r})$, $\underline{S} = \lim_{n \rightarrow \infty} \underline{S}_n$, $T_n = \sum_{j=1}^r S_{n,j}$ et $T = \lim_{n \rightarrow \infty} T_n$, où $S_{n,j} = \sum_{k=1}^n X_{k,j} X_{k+1,j}$.

Avec plusieurs résultats existants pour les lois marginales des $S_{n,j}$'s et leurs lois limites $S_j = \lim_{n \rightarrow \infty} S_{n,j}$, nous étudions les lois conjointes pour le cas bivarié ($r = 2$), ainsi que les lois de T_n et T pour $r \geq 2$.

FELIX CAMIRAND & TAOUFIK BOUEZMARNI (Université de Sherbrooke), **JEAN-FRANÇOIS QUESSY** (Université du Québec à Trois-Rivières)

[Tuesday May 27 / mardi 27 mai, 16:15]

Empirical Copula Processes for Time Series Data

Processus de copule empirique pour des séries chronologiques

The dependence structure between two random variables is sometimes subject to the influence of covariates. To model conditional dependence, two estimators of the conditional copula were first proposed by Gijbels (2011) and their asymptotic properties were studied by Veraverbeke (2011) for i.i.d. data. In this presentation, some of the asymptotic properties will be examined in the context of mixing data.

La structure de dépendance entre deux variables aléatoires est parfois sous l'influence de covariables. Pour modéliser la dépendance conditionnelle, deux estimateurs de la copule conditionnelle ont d'abord été proposés par Gijbels (2011) et leurs propriétés asymptotiques ont été étudiées par Veraverbeke (2011) pour des données i.i.d. Dans cette présentation, il sera question de certaines de leurs propriétés asymptotiques dans le cadre de données mélangeantes.

ALEXANDRE SCOTT (University of Western Ontario), **ADAM METZLER** (Wilfrid Laurier University)

[Tuesday May 27 / mardi 27 mai, 16:30]

Simulating Rare Event Using a Two Stage IS Estimator

Échantillonnage d'importance en deux étapes pour simuler des événements rares

2E-C4: Probability Theory and Finance Théorie des probabilités et finance

Rare event simulation is a crucial part of financial mathematics. If one wants to estimate accurately tails of a distribution, loss function or conditional value-at-risk, one needs to consider implementing a variance reduction method, often based on Importance Sampling (IS). In this talk, we propose a two stages IS estimator for diffusion processes; the first stage eliminates discretization error while the second stage provides variance reduction. Examples will be provided for the CIR and OU processes.

Simuler des événements rares est une étape cruciale des mathématiques financières. Par exemple, si l'on désire estimer avec précision les queues d'une distribution, une fonction de perte ou la CVAR, il faut développer une méthode visant à réduire la variance, ce qui est souvent fait en utilisant l'échantillonnage d'importance. Nous proposons d'utiliser deux étapes d'échantillonnage d'importance pour les processus de diffusion ; la première étape élimine l'erreur de discrétisation alors que la deuxième vise à réduire la variance. Nous présenterons aussi des exemples pour les modèles CIR et OU.

FRANCOIS WATIER & IMÈNE ALLAB (Université du Québec à Montréal)

[Tuesday May 27 / mardi 27 mai, 16:45]

First Passage Time Estimation of Diffusion Processes Through Time-varying Boundaries

Estimation du temps de premier passage dans les processus de diffusion à travers des limites variant dans le temps

We construct a Monte Carlo-based algorithm for estimating the FPT density of a time-homogeneous SDE through a time-dependent frontier. Brownian bridges are considered, as well as localized Daniels curve approximations, to obtain tractable estimations of upcrossing probabilities between successive points of a simulated path of the process. Finally we will apply our technique to a portfolio management problem.

Nous construisons un algorithme de Monte Carlo pour estimer la densité du temps de premier passage d'une équation différentielle stochastique homogène dans le temps à travers une limite variant dans le temps. Nous considérons des ponts browniens ainsi que des approximations de courbes de Daniels localisées pour obtenir des estimations tractables de la probabilité de croisements en montant entre points successifs d'un chemin du processus simulé. Enfin, nous appliquons notre technique à un problème de gestion de portefeuille.

2E-C5: Robust and Shrinkage Estimation Estimation robuste et par rétrécissement

Chair/Président: Alex Leblanc (University of Manitoba)

Room/Salle: UC - 144

Abstracts/Résumés

FUQI CHEN & SÉVÉRIEN NKURUNZIZA (The University of Windsor)

[Tuesday May 27 / mardi 27 mai, 15:30]

A Class of Shrinkage-Type Estimators in Multivariate Linear Regression with Change-Points

Classe d'estimateurs à rétrécissement dans un modèle de régression multivarié avec points de rupture

In this talk, we present an estimation problem in multivariate regression with several change-points, when the parameter matrix may satisfy some restrictions. We relax some assumptions which are commonly given in literature about the models with change-points and we derive the joint asymptotic normality between the unrestricted (UE) and the restricted (RE) estimators. We also propose a class of estimators which includes shrinkage estimators (SEs) as well as the UE and RE. We also generalize some identities which are useful in computing the bias and risk functions of shrinkage-type estimators. Further, we derive a more general condition for the SEs to dominate the UE. Finally, in order to illustrate the performance of the proposed method, we present some simulation studies.

Dans cet exposé, on s'intéresse au problème d'estimation matriciel en régression multivariée avec plusieurs points de rupture, lorsque la matrice pourrait satisfaire une restriction. On allège quelques présuppositions communément imposées et on établit la normalité asymptotique des estimateurs sans et avec restriction. Ensuite, on propose une classe d'estimateurs à rétrécissement qui comprend les estimateurs avec et sans restriction ainsi que ceux à rétrécissement. On généralise quelques identités applicables dans le calcul du biais et du risque. De plus, on établit une condition suffisante pour que l'estimateur à rétrécissement domine l'estimateur sans restriction. Cette condition est plus générale que celle communément donnée en littérature. Finalement, pour illustrer les performances de la méthodologie proposée, on présente quelques résultats de simulations.

DEBBIE DUPUIS, NICOLAS PAPAGEORGIOU & BRUNO RÉMILLARD (HEC Montréal)

[Tuesday May 27 / mardi 27 mai, 15:45]

Robust Conditional Variance and Value-at-Risk Estimation

Estimation robuste de la variance conditionnelle et de la valeur à risque

Losses due to idiosyncratic events can have a disproportionate impact on traditional value-at-risk estimates, upwardly biasing these estimates, increasing capital requirements and unnecessarily reducing the available capital and profitability of financial institutions. We propose new bias-robust conditional variance estimators based on weighted likelihood at heavy-tailed models, as well as value-at-risk estimators based on the latter and on volatility updated historical simulation. The new value-at-risk estimators also use optimally chosen rolling window length and smoothing parameter value.

Les pertes dues à des événements particuliers peuvent avoir un impact disproportionné sur les estimateurs traditionnels de la valeur à risque. Les estimations sont biaisées vers le haut et ceci augmente les exigences de fonds propres, réduisant inutilement le capital disponible et la rentabilité des institutions financières. Nous proposons des nouveaux estimateurs robustes de la variance conditionnelle basés sur une vraisemblance pondérée des modèles à queue lourde. Nous utilisons ces nouveaux estimateurs ainsi qu'une simulation historique avec mise à jour de la volatilité pour construire des estimateurs robustes de la valeur à risque. Ces nouveaux estimateurs utilisent également un choix optimal

2E-C5: Robust and Shrinkage Estimation Estimation robuste et par rétrécissement

de la longueur de la fenêtre dynamique et de la valeur du paramètre de lissage.

DAVID HAZIZA (Université de Montréal), **CYRIL FAVRE MARTINOZ** (CREST-ENSAI), **JEAN-FRANÇOIS BEAUMONT** (Statistique Canada)

[Tuesday May 27 / mardi 27 mai, 16:00]

Robust Estimation for Asymmetric Infinite Populations

Estimation robuste pour des populations infinies asymétriques

Estimation of the mean in the case of an asymmetric distribution is a practical problem. Indeed, it is quite common to observe variables with asymmetric distribution. In practice, the sample contains units which are very influential on the empirical mean, which is often the chosen estimator. Rivest (1994) proposed a winsorised estimator, obtained by winsorizing the largest observation in the sample, and showed that this estimator has good properties in terms of the mean squared error. We present an alternative estimator that is based on the concept of conditional bias of a unit. We will give the asymptotic properties of this estimator in terms of mean squared error. Finally, we present the results of a simulation study.

L'estimation de la moyenne dans le cas d'une population asymétrique est un problème en pratique. En effet, il est très courant d'observer des variables dont la distribution est asymétrique. En pratique, l'échantillon contient des unités qui sont très influentes sur la moyenne empirique, qui est l'estimateur souvent privilégié. Rivest (1994) a proposé l'estimateur winsorisé une fois obtenu en winsorisant la plus grande observation de l'échantillon et a montré que cet estimateur possède de bonnes propriétés en termes d'erreur quadratique moyenne. Nous présentons un estimateur alternatif qui repose sur le concept de biais conditionnel d'une unité. Nous donnerons les propriétés de cet estimateur en termes d'erreur quadratique moyenne. Finalement, nous présenterons les résultats d'une étude par simulation.

MARCOS SANCHES (Centre for Addiction and Mental Health (CAMH))

[Tuesday May 27 / mardi 27 mai, 16:15]

Regression Discontinuity with Age as Running Variable – Issues with Statistical Tests and Comparison with Time Series Intervention Analysis

Discontinuité par régression avec l'âge comme variable d'ordre, tests statistiques et comparaison avec l'analyse des séries chronologiques

Regression Discontinuity Design has become mainstream in studies where age is used as the running variable. In these cases age is not a time variable and Least Square Regression is used to estimate $E(Y|X)$ and to make inference about the causal effect at the discontinuity point. In independent samples, statistics for different age ranges are usually considered independent as they are calculated from different individuals. We show that when age ranges are finely created (e.g. age in months) the sequence of observations will many times have a strong auto-correlation structure, making the usual regression inference not appropriated. In these cases we investigate the use Time Series Intervention Analysis and compare both approaches.

Le concept de discontinuité par régression est devenu commun dans les études où l'âge est utilisé comme variable d'ordre. Dans ces cas, l'âge n'est pas une variable de temps et la régression des moindres carrés est utilisée pour estimer $E(Y|X)$ et mener des inférences à propos de l'effet causal au point de discontinuité. Dans des échantillons indépendants, les statistiques sur les tranches d'âge sont généralement considérées indépendantes car elles sont calculées à partir de différents individus. Nous montrons que lorsque les tranches d'âge sont fines (p. ex., âge en mois), la séquence des observations présente souvent une forte structure d'autocorrélation, rendant l'inférence de régression habituelle inappropriée. Dans ces cas, nous étudions l'utilisation de l'analyse des séries chronologiques et comparons les deux approches.

3A-A1: CRM-SSC Prize in Statistics Address
Allocution du récipiendaire du Prix CRM-SSC en statistique

Organizer and Chair / Responsable et président: Jose Garrido (Concordia University)

Room/Salle: Hart House - Music Rm

Abstract/Résumé

FANG YAO (University of Toronto)

[Wednesday May 28 / mercredi 28 mai, 8:45]

Functional Data Analysis and Related Topics

Analyse de données fonctionnelles et sujets connexes

Functional data analysis (FDA) has received substantial attention, with applications arising from various disciplines, such as engineering, public health, finance etc. In general, the FDA approaches focus on nonparametric underlying models that assume the data are observed from realizations of stochastic processes satisfying some regularity conditions, e.g., smoothness constraints. The estimation and inference procedures usually do not depend on merely a finite number of parameters, which contrasts with parametric models, and exploit techniques, such as smoothing methods and dimension reduction, that allow data to speak for themselves. In this talk, I will give an overview of FDA methods and related topics developed in recent years.

L'analyse de données fonctionnelles (ADF) a reçu une attention considérable, avec des applications provenant de diverses disciplines, notamment l'ingénierie, la santé publique, les finances, etc. En général, l'approche de l'ADF se concentre sur des modèles sous-jacents non paramétriques qui présupposent que les données sont tirées de réalisations de processus stochastiques qui suivent certaines conditions de régularité, telles que des contraintes d'ajustement. Les processus d'estimation et d'inférence ne dépendent généralement pas seulement d'un nombre fini de paramètres, ce qui contraste avec les modèles paramétriques, et exploitent des techniques telles que les méthodes de lissage et de réduction de dimension, qui permettent aux données de parler d'elles-mêmes. Dans cette allocution, je présenterai un survol des méthodes d'ADF et de sujets connexes développés au cours des dernières années.

3A-A2: SSC Impact Award Address
Allocution du récipiendaire du Prix pour impact de la SSC

Organizer and Chair / Responsable et présidente: Thérèse Stukel (Institute for Clinical Evaluative Sciences)

Room/Salle: Hart House - Debates Rm

Abstract/Résumé

DAVID THOMSON (Queen's University)

[Wednesday May 28 / mercredi 28 mai, 8:45]

A Time Series Analyst's Look at Climate Data

Regard d'un analyste de séries chronologiques sur les données climatiques

Information about many processes, from those created by man to those at the core of our existence, appear as time series. A statistician attempting to analyze one of these time series, however, has been faced with a bewildering array of what are often ad-hoc methods cobbled together from different fields. Many of these do not extract information from real data efficiently. Multitaper methods for analyzing time series have improved this situation in several ways: there is theory explaining data tapers; multitaper estimates of spectra and autocovariances generally outperform standard estimates, and they allow estimates of quantities that were previously indeterminate. These are illustrated these climate and related data.

Dans une foule de procédés, tant ceux créés par l'être humain que d'autres qui sont au coeur de notre existence, l'information se présente sous forme de séries chronologiques. En cherchant à analyser l'une de ces séries chronologiques, un statisticien s'est toutefois confronté à une mosaïque surprenante de méthodes ad hoc issues de disciplines diverses. Bon nombre de ces méthodes ne peuvent extraire efficacement l'information à partir de données réelles. Les méthodes à fenêtres de pondération orthogonales utilisées dans l'analyse des séries chronologiques ont grandement amélioré la situation de plusieurs façons : les fenêtres de pondération de données possèdent un fondement théorique ; l'estimation du spectre et de l'autocovariance basée sur des fenêtres de pondération orthogonales offre une performance supérieure aux estimations standards et elle permet d'estimer des quantités qui restaient auparavant indéterminées. Ces méthodes sont illustrées avec des données climatiques et d'autres données afférentes.

3B-I1: Bioinformatics and Copula Models Bioinformatique et modèles de copules

Chair/Présidente: Ying Zhang (Acadia University)
Organizer/Responsable: Laura Cowen (University of Victoria)

Session sponsored by the Committee on Women in Statistics
Séance parrainée par le Comité des femmes en statistique

Room/Salle: Hart House - East Common

Abstracts/Résumés

KELLY BURKETT (University of Ottawa)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Using Gene Genealogies to Detect Rare Disease-Predisposing Genetic Variants

Détection de variantes génétiques prédisposant à des maladies rares à l'aide de la généalogie génétique

The gene genealogy is a tree that describes the relationships between DNA sequences sampled from a population. We consider the potential application of gene genealogies to the discovery of rare disease-predisposing genetic variants. Sequences from affected individuals that all carry the same disease-predisposing mutation will be clustered together in the ancestral tree corresponding to the location of the mutation. In this talk, we describe tree-based statistics that quantify the clustering of affected individuals in the gene genealogy and we describe simulations that show how the proposed statistics can be used to find regions harbouring rare disease-predisposing genetic variants. Since the true ancestral trees are unknown, we also give a brief discussion on how to handle tree uncertainty.

La généalogie génétique est un arbre qui décrit les relations entre les séquences d'ADN échantillonnées dans une population. Nous étudions l'application potentielle de généalogies génétiques à la découverte de variantes génétiques prédisposant à des maladies rares. Les séquences provenant d'individus affectés qui portent la même mutation prédisposant à la maladie sont groupées ensemble dans l'arbre ancestral correspondant à l'emplacement de la mutation. Dans cet exposé, nous décrivons des statistiques basées sur des arbres qui quantifient le regroupement d'individus affectés dans la généalogie génétique. Nous présentons des simulations qui montrent comment les statistiques proposées peuvent être utilisées pour trouver des régions présentant des variantes génétiques prédisposant à des maladies rares. Comme les véritables arbres ancestraux sont inconnus, nous discutons aussi brièvement de la gestion de l'incertitude liée aux arbres.

JEMILA HAMID & SAYANTEE JANA (McMaster University)

[Wednesday May 28 / mercredi 28 mai, 10:50]

A Test for the Growth Curve Model and its High-Dimensional Extension: Applications in Genomics

Test pour le modèle de courbe de croissance et sa généralisation de grande dimension: applications en génomique

The growth curve model (GCM) is a generalized multivariate analysis of variance (GMANOVA) model useful in the analysis of longitudinal data. Unlike MANOVA models, the GCM allows a structured mean where the change in mean is modeled through within individual design matrix though a bi-

Le modèle de courbe de croissance (MCC) est un modèle d'analyse de la variance multivariée généralisée (GMANOVA) utile en analyse de données longitudinales. Contrairement aux modèles d'analyse de la variance multivariée (MANOVA), le MCC confère une moyenne structurée où le changement de la moyenne est modélisé dans la configuration bilinéaire

3B-I1: Bioinformatics and Copula Models Bioinformatique et modèles de copules

linear setup. We consider a trace test for testing general linear hypothesis for the mean parameter in the GCM and discuss its properties through an extensive simulation. We provide high-dimensional extensions using the Moore-Penrose generalized inverse as well as through a Bayesian framework, where the covariance matrix is assumed to follow a Wishart distribution. We will explore properties of the test under high-dimensional scenarios and provide real data application for gene-filtering in time course genetic experiments.

d'une matrice de plan d'expérience individuel. Nous considérons un test de la trace pour évaluer l'hypothèse linéaire générale du paramètre de la moyenne du MCC et discutons ses propriétés par le biais d'une étude de simulation exhaustive. Nous proposons des généralisations de grande dimension à l'aide de l'inverse généralisée de Moore-Penrose, ainsi que par le biais d'un cadre bayésien, dans lequel on suppose que la matrice de covariance suit une distribution de Wishart. Nous examinons les propriétés du test selon différents scénarios de grande dimension et proposons des applications à des données réelles en matière de filtrage de gènes dans des expériences de chronologie génétique.

MICHELLE ZHOU (Simon Fraser University), **SHULIN ZHANG** (Southwestern University of Finance and Economics, China), **OSTAP OKHRIN** (Humboldt University of Berlin, Germany), **PETER SONG** (University of Michigan, USA)
[Wednesday May 28 / mercredi 28 mai, 11:20]

Goodness-of-fit Test for Specification of Copula-Based Semi-parametric Models

Test de validité de l'ajustement pour la spécification de modèles semi-paramétriques fondés sur les copules

Recently, copulas have become increasingly popular in modelling dependency among multiple variables since they allow one to model and estimate the marginal distribution and dependence structure separately. Semi-parametric copula models are characterized by unspecified marginal distributions and parametric copula functions. Misspecification on parametric structures of copula functions may cause misleading statistical estimation and inference. In this talk, we propose a new test constructed via the contrast between "in-sample" and "out-of-sample" pseudo-likelihoods. Under the null hypothesis that the copula model is correctly specified, we show that the proposed test statistic converges in probability to a constant equal to the dimension of the parameter space and establish the asymptotic normality for the test statistic.

Récemment, les copules connaissent une popularité grandissante pour la modélisation de la dépendance en présence de multiples variables, puisqu'elles permettent de modéliser et d'estimer la distribution marginale et la structure de la dépendance de façon séparée. Les modèles semi-paramétriques de copule se caractérisent par des distributions marginales non spécifiées et des copules paramétriques. Une mauvaise spécification des structures paramétriques de ces copules peut engendrer une erreur dans l'estimation statistique et l'inférence. Dans cet exposé, nous proposons un nouveau test construit en exploitant le contraste entre les pseudo-vraisemblances «intra-échantillon» et «extra-échantillon». Sous l'hypothèse nulle selon laquelle le modèle de copule est correctement spécifié, nous montrons que la statistique de test proposée converge en probabilité vers une constante égale à la dimension de l'espace paramétrique, puis nous établissons la normalité asymptotique de la statistique de test.

3B-I2: Modern Statistical Methods for Complex Health Data Méthodes statistiques modernes pour données sanitaires complexes

Chair/Président: Kevin He (University of Michigan)

Organizers/Responsables: Kevin He and/et Yi Li (University of Michigan)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: Hart House - Debates Rm

Abstracts/Résumés

RICHARD COOK (University of Waterloo)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Statistical Challenges Arising in Transfusion Research

Défis statistiques en recherche transfusionnelle

There are many important scientific questions in transfusion medicine that remain to be addressed. There is currently considerable interest in i) exploring the relation between the duration blood has been in storage prior to transfusion and adverse health outcomes in recipients, ii) the effect of transfusing blood to patients with a different blood type, and iii) the effect of transfusing gender mismatched blood (i.e. blood from female donors to male recipients). The statistical challenges and approaches to analysing data from a large registry of transfused patients are discussed with a view to exploring these issues.

Plusieurs questions scientifiques importantes en médecine transfusionnelle demeurent sans réponse. Un grand intérêt est actuellement porté à i) la relation entre la période d'entreposage du sang avant la transfusion et les effets néfastes sur la santé des patients transfusés, ii) l'effet de la transfusion de sang à des patients présentant un groupe sanguin différent, et iii) l'effet d'une transfusion de sang entre deux personnes de sexe différent (p. ex. le sang d'une femme transfusé à un homme). Nous discutons les défis et les approches statistiques pour l'analyse de données provenant d'un vaste registre de patients transfusés en vue d'examiner ces problématiques.

JOHN KALBFLEISCH, MATHIEU BRAY, PETER SONG, WEN WANG, ALAN LEICHTMAN & MICHAEL REES (University of Michigan)

[Wednesday May 28 / mercredi 28 mai, 10:42]

Optimization in a Kidney Paired Donation Program

Optimisation d'un programme de transplantation rénale utilisant donneurs vivants jumelés par échange de bénéficiaires

Kidney transplant is the best available treatment for patients with end stage kidney disease. Often, however, patients have a willing living donor, but that donor is incompatible in blood type and/or histology. A kidney paired donation (KPD) program consists of transplant candidates and their incompatible living donors as well as non-directed or altruistic donors (ADs). Exchanges of donors among candidates and chains created by ADs overcome

La transplantation rénale est le meilleur traitement disponible pour les patients atteints de maladie rénale au stade terminal. Souvent, les patients ont un donneur vivant prêt à donner, mais qui est refusé, soit à cause de l'incompatibilité entre les groupes sanguins ou entre les histologies du donneur et du patient. Un programme de transplantation rénale utilisant donneurs vivants jumelés par échange de bénéficiaires (KPD) est constitué de candidats à la transplantation rénale et leurs donneurs vivants incompatibles, ainsi

3B-I2: Modern Statistical Methods for Complex Health Data Méthodes statistiques modernes pour données sanitaires complexes

the incompatibilities. A problem of importance is how best to arrange exchanges and chains in order to achieve as many transplants as possible. We develop allocation schemes that account for uncertainties. These methods are compared through simulations with methods currently in use and illustrate substantial gains in transplants achieved.

que certains donneurs non-associés ou altruistes (AD). Les échanges entre donneurs et candidats, ainsi que les chaînes commencées par AD, surmontent ces incompatibilités. Nous développons des systèmes d'allocation qui intègrent certaines incertitudes.

YI LI, JI ZHU & KEVIN HE (University of Michigan)

[Wednesday May 28 / mercredi 28 mai, 11:04]

Gateaux Differential-Based Boosting for Time-Varying Effects in Survival Analysis

Boosting de Gâteaux basé sur le différentiel pour les effets variant dans le temps en analyse de survie

Survival models with time-varying effects provide a flexible framework for modeling the effects of covariates on event times. However, the difficulty of model construction increases dramatically as the number of variable grows. Existing constrained optimization and boosting methods suffer from computational complexity. We propose a new Gateaux differential-based boosting procedure for simultaneously selecting and automatically determining the functional form of covariates. The proposed method is flexible in that it extends the gradient boosting to functional differentials in general parameter space. In each boosting learning step of this procedure, only the best-fitting base-learner (and therefore the most informative covariate) is added to the predictor, which consequently encourages sparsity. In addition, the method controls smoothness, which is crucial for improving predictive performance.

Les modèles de survie présentant des effets variant dans le temps offrent un cadre flexible pour la modélisation des effets des covariables aux temps d'événements. Cependant, la modélisation se complique rapidement avec l'augmentation du nombre de variables. Les méthodes existantes d'optimisation sous contraintes et de boosting souffrent de la complexité des calculs. Nous proposons une nouvelle méthode de boosting de Gâteaux basée sur le différentiel qui permet la sélection simultanée et la détermination automatique des formes fonctionnelles de covariables. La méthode proposée est flexible, puisqu'elle élargit le boosting de gradients à des différentiels fonctionnels dans l'espace général des paramètres. Dans chacune des étapes d'apprentissage de cette procédure, seul l'apprenti de base le mieux ajusté (et par conséquent la covariable la plus informative) est ajouté à l'estimateur, ce qui favorise l'éparpillement. De plus, la méthode contrôle la régularité, capitale pour l'amélioration de la performance prédictive.

ROSS PRENTICE (Fred Hutchison Cancer Research Center)

[Wednesday May 28 / mercredi 28 mai, 11:26]

Nonparametric and Semiparametric Analysis of Bivariate Failure Time Data

Analyse non paramétrique et semi-paramétrique de données de temps de défaillance bivariées

The nonparametric maximum likelihood estimation problem is re-defined to entail maximization over marginal hazard rates, and over double failure hazard rates at all grid points formed by uncensored times where there is a possibility of double failure given preceding data. The resulting unique NPMLE can be calculated using a two-step procedure. The first involves setting aside all doubly censored observations that are interior to the risk region, lead-

Nous redéfinissons le problème de l'estimation du maximum de vraisemblance non paramétrique pour substituer la maximisation des taux de risque marginaux et des taux de risque de double défaillance à tous les points de la grille formés par les temps non censurés où une possibilité de double défaillance existe, selon les données précédentes. L'estimateur non paramétrique du maximum de vraisemblance (ENPMV) unique qui en résulte peut être calculé à l'aide d'une méthode en deux étapes. La première étape consiste

3B-I2: Modern Statistical Methods for Complex Health Data

Méthodes statistiques modernes pour données sanitaires complexes

ing to a Dabrowska-type NPMLE from the remaining data. A corresponding NPMLE arises also for the censoring 'survivor' function. The second brings in the omitted data using self-consistency, yielding non-iterative NPMLEs for both the failure and censoring distribution functions. Simulation studies demonstrate modest efficiency improvements. Regression generalizations will be briefly mentioned.

à mettre de côté toutes les données censurées en double à l'intérieur de la région de risque. Les données restantes mènent à un ENPMV de type Dabrowska, et un ENPMV est aussi obtenu pour la fonction de censure des survivants. La seconde étape ramène les données omises par une équation d'autocohérence, produisant des ENPMV non itératifs à la fois pour les fonctions de répartition de défaillance et de censure. Des études de simulation montrent de modestes améliorations de l'efficacité. Nous présentons aussi sommairement des généralisations de régression.

3B-13: New Challenges and Developments in Survival Analysis and its Applications Nouveaux problèmes et progrès en analyse de survie et ses applications

Organizer and Chair / Responsable et présidente: Jiajia Zhang (University of South Carolina)

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

ZHEZHEN JIN (Columbia University)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Nonparametric Threshold Selection with Censored Survival Data

Sélection d'un seuil non paramétrique pour des données de survie censurées

In biomedical research and practice, quantitative biomarkers are often used for diagnostic or prognostic purposes, with a threshold established on the measurement to aid binary classification. When prognosis is on survival time, single threshold may not be informative. It is also challenging to select threshold when the survival time is subject to random censoring. Using survival time dependent sensitivity and specificity, we extend classification accuracy based objective function to allow for survival dependent threshold. To estimate optimal threshold for a range of survival rate we adopt a non-parametric procedure, which produces satisfactory result in a simulation study. The method will be illustrated with a real example. This talk is based on joint work with Dr. Liu Xinhua at Columbia University.

Dans la recherche et la pratique biomédicales, les biomarqueurs quantitatifs sont souvent utilisés à des fins de diagnostics ou pronostics, où un seuil est établi sur la mesure afin d'aider à la classification binaire. Lorsqu'il s'agit d'un pronostic de survie, il est possible qu'un seuil unique ne soit pas informatif. Il est également difficile d'établir un seuil lorsque le temps de survie est sujet à la censure aléatoire. À l'aide de la sensibilité et de la spécificité dépendantes du temps de survie, nous généralisons la fonction objective de la classification basée sur la précision pour permettre d'établir un seuil dépendant de la survie. Nous utilisons une procédure non paramétrique pour estimer le seuil optimal d'un éventail de taux de survie, ce qui produit des résultats satisfaisants dans une étude de simulation. La méthode est illustrée à l'aide d'un exemple. Ces travaux sont effectués conjointement avec Liu Xinhua de la Columbia University.

YINGWEI PENG (Queen's University), **YI NIU** (Dalian University of Sciences and Technology)

[Wednesday May 28 / mercredi 28 mai, 10:50]

Mixture Cure Models for Clustered Survival Data

Modèles de mélange avec fraction non à risque pour des données de survie en grappes

Mixture cure models have long been used to model survival data with a cured fraction. Even though other types of cure models have been proposed as alternatives, the mixture cure model still received a great deal of attention in recent years. In this talk, I will review existing methods to extend mixture cure models for clustered survival data with a cured fraction. A newly proposed estimating equation approach will be discussed and compared with the existing methods. The performance of the approach will be demonstrated with simulation stud-

Les modèles de mélange sont utilisés depuis longtemps pour modéliser les données de survie avec une fraction non à risque. Bien que d'autres types de modèles avec fraction non à risque aient été proposés, le modèle de mélange avec fraction non à risque a suscité beaucoup d'intérêt au cours des dernières années. Dans cet exposé, je passe en revue les méthodes existantes pour généraliser les modèles de mélange avec fraction non à risque aux données de survie en grappes présentant une fraction non à risque. Je propose une nouvelle équation d'estimation, que je compare aux méthodes existantes. Des études de simulation illustre

3B-13: New Challenges and Developments in Survival Analysis and its Applications Nouveaux problèmes et progrès en analyse de survie et ses applications

ies. The approach will be employed to analyze a cancer survival data set.

la performance de l'approche proposée, qui est employée pour analyser un jeu de données de survie au cancer.

GRACE YI & YANG NING (University of Waterloo), **NANCY REID** (University of Toronto)

[Wednesday May 28 / mercredi 28 mai, 11:20]

Weighted Estimating Equations for Semiparametric Transformation Models With Missing Covariates

Équations d'estimation pondérées pour des modèles de transformation semi-paramétriques avec covariables manquantes

In survival analysis, covariate measurements often contain missing observations; ignoring this feature can lead to invalid inference. We propose a class of weighted estimating equations for right censored data with missing covariates under semiparametric transformation models. Time-specific and subject-specific weights are accommodated in the formulation of the weighted estimating equations. We establish unified results for estimating missingness probabilities that cover both parametric and non-parametric modeling schemes. To improve estimation efficiency, the weighted estimating equations are augmented by a new set of unbiased estimating equations. The resultant estimator has the so called "double robustness" property and enjoys the optimality property among a class of estimators. This is joint work with Ning Yang and Nancy Reid.

En analyse de survie, les mesures de covariables comportent souvent des observations manquantes. L'omission de ce fait peut mener à l'obtention d'une inférence erronée. Nous proposons une classe d'équations d'estimation pondérées pour des données censurées à droite présentant des covariables manquantes dans le cadre de modèles de transformation semi-paramétriques. Des poids spécifiques au temps et à l'individu sont utilisés dans l'élaboration des équations d'estimation pondérées. Pour estimer les probabilités que des données soient manquantes, nous établissons des résultats unifiés qui couvrent à la fois les modèles paramétrique et non paramétrique. Afin d'améliorer l'efficacité de l'estimation, les équations d'estimation pondérées sont bonifiées d'un nouvel ensemble d'équations non biaisées. L'estimateur qui en résulte est doublement robuste et possède l'exigence d'optimalité parmi une classe d'estimateurs. Ces travaux sont effectués conjointement avec Ning Yang et Nancy Reid.

3B-I4: New Developments in Statistical Neuroscience Récents progrès en neuroscience statistique

Organizer and Chair / Responsable et président: Shoja'eddin Chenouri (University of Waterloo)

Room/Salle: Hart House - South Dining

Abstracts/Résumés

SAM BEHSETA (California State University)

[Wednesday May 28 / mercredi 28 mai, 10:20]

A Semiparametric Bayesian Model for Detecting Multiway Synchrony Among Neurons

Modèle bayésien semi-paramétrique pour la détection de la synchronie multivoies entre neurones

We propose a semiparametric Bayesian model to capture dependencies among multiple neurons by detecting their co-firing (possibly with some lag time) patterns over time. After discretizing time so there is at most one spike at each interval, the resulting sequence of 1's and 0's for each neuron is modeled using the logistic function of a continuous latent variable with a Gaussian process prior. For multiple neurons, the corresponding marginal distributions is coupled to their joint probability distribution using a parametric copula model. This way, while the nonparametric component (i.e., the Gaussian process model) provides a flexible framework for modeling the underlying firing rates, the parametric component (i.e., the copula model) allows us to make inference regarding the relationships among neurons.

Nous proposons un modèle bayésien semi-paramétrique pour repérer la dépendance entre des neurones en détectant les motifs de décharge simultanée (ou comportant un léger décalage) dans le temps. En discrétisant le temps afin de limiter chaque intervalle à une impulsion, la séquence de 0 et de 1 obtenue pour chaque neurone est modélisée à l'aide de la fonction logistique d'une variable latente continue selon une loi a priori de processus gaussien. Pour les neurones multiples, les distributions marginales correspondantes sont couplées à leur distribution conjointe des probabilités à l'aide d'un modèle paramétrique de copule. De cette façon, alors que la composante non paramétrique (i.e. le modèle du processus gaussien) fournit un cadre flexible pour la modélisation des taux de décharge sous-jacents, la composante paramétrique (i.e. le modèle de copule) nous permet d'inférer concernant les relations entre les neurones.

MATTHEW HARRISON (Brown University)

[Wednesday May 28 / mercredi 28 mai, 10:50]

Robust Inference for Nonstationary Spike Trains

Inférence robuste pour les trains d'impulsion neuronaux non stationnaires

The coordinated spiking activity of simultaneously recorded neurons can reveal clues about the dynamics of neural information processing, about the mechanisms of brain disorders, and about the underlying anatomical microcircuitry. Statistical models and methods play an important role in these investigations. In cases where the scientific questions require disambiguating dependencies across multiple spatial and temporal scales, conditional inference can be used to create procedures that are strikingly robust to nonstationarity, model mis-

Les impulsions coordonnées neuronales enregistrées simultanément peuvent révéler des indices sur la dynamique du traitement de l'information neuronale, sur les mécanismes des troubles cérébraux et sur les microcircuits anatomiques sous-jacents. Les modèles et méthodes statistiques jouent un rôle important dans ces recherches. Dans les cas où les questions scientifiques requièrent l'élimination de l'ambiguïté liée aux dépendances sur de multiples échelles spatiales et temporelles, l'inférence conditionnelle permet d'élaborer des procédures remarquablement robustes à la non stationnarité, à la mauvaise spécification du modèle et aux

3B-I4: New Developments in Statistical Neuroscience Récents progrès en neuroscience statistique

specification, and incidental parameters problems, which are common neurostatistical challenges. Examples include testing for cell assembly dynamics in human epilepsy data and learning putative anatomical networks from spike train data in behaving rodents.

paramètres incidents, des problèmes courants en neurostatistique. Les exemples présentés comprennent des tests sur la dynamique de l'assemblage cellulaire avec des données sur l'épilepsie humaine, ainsi que l'apprentissage de réseaux anatomiques putatifs provenant de données de trains d'impulsion sur des rongeurs.

PAUL MARRIOTT (University of Waterloo)

[Wednesday May 28 / mercredi 28 mai, 11:20]

The Analysis of Multiple Spike Train Data

Analyse de données sur les trains d'impulsion multiple

Spike train data - the firing times of the neurons - based on measurements of multiple cells provides interesting new challenges for the analyst. These can be thought of as multivariate point process data with information on different time scales and with complex dependence structures. The talk looks at ways that ideas of dimension reduction, data visualization and modelling, which are familiar in a continuous response context, can be extended to this novel setting.

Les données sur les trains d'impulsion, c'est-à-dire le nombre de décharge des neurones, fondées sur des mesures de cellules multiples présente de nouveaux défis intéressants pour l'analyste. Elles peuvent être considérées comme des données de processus ponctuels multivariés comportant de l'information sur différentes échelles temporelles et des structures de dépendance complexes. Cet exposé porte sur des façons d'étendre à cette nouvelle perspective les concepts de réduction de la dimension, de visualisation de données et de modélisation, qui sont courants dans le contexte d'une réponse continue.

3B-I5: Statistical Education in Light of Philosophy of Science (A Panel Discussion) Philosophie de la science et éducation statistique

Chair/Président: Kevin Thorpe (University of Toronto)

Organizer/Responsable: Jeff Picka (University of New Brunswick)

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: UC - Croft House

Abstracts/Résumés

MAYA GOLDENBERG (University of Guelph)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Philosophy of Science Meets Evidence-Based Medicine

La philosophie de la science à la rencontre de la médecine factuelle

Because few would object to evidence-based medicine's (EBM) principal task of basing medical decision-making on the best evidence, it surprises many to learn of disagreement over the merits of this prolific movement. The controversy speaks to fundamental issues regarding how physicians reason in the clinical context. The complex practice of integrating multiple evidentiary warrants -- ranging from clinical trial data to patient preferences--is strongly argued by a vocal minority to be inadequately captured in EBM's ranking of research methods (the hierarchy of evidence) and virtual silence on the challenge of knowledge translation from population-based studies to individual patient care. Themes and concepts from the philosophy of science will be used to mediate this dispute and offer some promising correctives.

Puisque peu s'objecteraient à la tâche principale de la médecine factuelle de baser des processus décisionnels médicaux sur la meilleure preuve, il est surprenant d'apprendre qu'il existe un désaccord sur les mérites de ce mouvement prolifique. La controverse s'adresse aux questions fondamentales concernant la manière donc les médecins raisonnent dans un contexte clinique. La pratique complexe de l'intégration de multiples mandats de preuve - allant des données d'études cliniques aux préférences des patients - est fortement soutenue par une minorité vocale comme étant inadéquatement capturée dans la classification des méthodes de la médecine factuelle (la hiérarchie de preuve) et pratiquement le silence sur le défi de l'application des connaissances allant des études de populations que les soins individuels aux patients. Les thèmes et concepts de la philosophie de la science seront utilisés pour la médiation de ce différent et afin d'offrir certains correctifs prometteurs.

JEFF PICKA (University of New Brunswick)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Discussion of Using Science Studies Resources in Statistical Education

Discussion sur l'utilisation des ressources scientifiques pour l'éducation statistique

I will lead a discussion of how ideas and resources from the history, philosophy, and sociology of science can be used to improve the teaching of both statistical methods and statistical consulting technique.

J'animerai une discussion sur la façon dont les idées et les ressources tirées de l'histoire, de la philosophie et de la sociologie de la science peuvent servir à améliorer l'enseignement des méthodes statistiques et des techniques de consultation statistique.

3B-I5: Statistical Education in Light of Philosophy of Science (A Panel Discussion) Philosophie de la science et éducation statistique

NATHAN YOUNG (University of Ottawa)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Statistics and the Politics of Expert Knowledge: Understanding the 'Public Life' of Statistical Knowledge, Insights from the Social Sciences

La statistique et les politiques en matière d'expertise: comprendre la 'vie publique' de la connaissance statistique; perspectives des sciences sociales

If ours is indeed a 'knowledge society', it is one founded on a paradox. On the one hand, the widespread availability of science, data, and expertise enhances democracy. On the other hand, the increasing volume and variety of knowledge allows activists, politicians, and everyday citizens to selectively craft authoritative-sounding arguments based on prior ideological or political positions. This paper will explore the consequences of these developments for the statistics profession. Statisticians have long thought of themselves as 'honest brokers' who allow the data to speak for themselves. Is such a position still tenable in an age of knowledge politics, or do statisticians have to change how they produce and communicate their findings to combat potential abuses of their knowledge?

Si notre société en est une basée sur la connaissance, elle repose sur un paradoxe. D'une part, la grande disponibilité de la science, des données et de l'expertise favorise la démocratie. D'autre part, l'augmentation du volume et de la diversité de la connaissance permet aux activistes, aux politiciens et aux citoyens d'avancer des arguments en apparence savants fondés sur des avis idéologiques ou politiques préconçus. Dans cet article, nous examinons les conséquences de ces développements sur la profession statistique. Les statisticiens se perçoivent depuis longtemps comme des «intermédiaires honnêtes» qui permettent aux données de parler d'elles-mêmes. Cette vision est-elle toujours soutenable à l'ère des politiques fondées sur le savoir, ou les statisticiens doivent-ils modifier leur façon de produire et de communiquer leurs conclusions afin de lutter contre de potentiels abus de leur savoir ?

3B-I6: Stochastic Modelling in Industry Modélisation stochastique en industrie

Organizer and Chair / Responsable et président: Neal Madras (York University)

Session sponsored by the Probability Section and the Actuarial Science Section

Séance parrainée par le Groupe de probabilité et le Groupe de science actuarielle

Room/Salle: UC - 179

Abstracts/Résumés

GUANGYU FU (Scotiabank)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Managing Volatility Smile in the Option Market: SABR, Local Volatility and Other Models

Gestion du «sourire de volatilité» sur le marché des options: modèle SABR, de volatilité locale et autres modèles

Implied volatility smile patterns are commonly seen in option products. In this talk, we will review the evolution of mathematical models to manage these smiles. SABR model on interest rate products and local volatility model on equity products will be specifically introduced. Practical issues and some rule-of-thumb solutions adopted by the industry will be discussed as well.

Des motifs de volatilité en forme de sourire sont courants dans les produits d'options. Dans cet exposé, nous examinons l'évolution des modèles mathématiques de gestion de ces «sourires». Nous présentons spécifiquement le modèle SABR pour les produits de taux d'intérêt et le modèle de volatilité locale pour les opérations sur actions. Nous discutons également quelques questions pratiques et des solutions empiriques adoptées par l'industrie.

DAN WILSON (Invidi Technologies Corporation)

[Wednesday May 28 / mercredi 28 mai, 10:50]

Challenges of Centralized and Distributed Control for Addressable Advertising

Défis de la commande centralisée et répartie en matière de publicité adressable

The increased flexibility and efficiency of addressable television advertising incurs a similar increase in the complexity and difficulty of management and measurement. Many of the problems that arise are highly amenable to stochastic modelling. Three problems are presented. The first, a centralized problem of shared resource allocation, deals with both long and short-term prediction of collisions between commercial breaks across sets of networks. The second, prediction of audience sizes given complex targeting criteria, involves centralized control mixed together with distributed decision making. A third problem highlights the difficulty of correctly pacing the delivery of advertisements when final delivery decisions are made probabilistically by end-user devices in the face of lossy and slow two-way communication with those devices.

La flexibilité et l'efficacité accrues de la publicité télévisée adressable engendre une hausse similaire de la complexité et de la difficulté en matière de gestion et de mesure. Bon nombre des problèmes qui surviennent se prêtent très bien à la modélisation stochastique. Trois problèmes sont présentés. Le premier est un problème centralisé d'allocation partagée des ressources qui concerne les prédictions à long et à court terme des simultanités entre les pauses publicitaires des différents ensembles de réseaux. Le deuxième problème, relatif à la prédiction de la taille des auditoires étant donné un critère complexe de ciblage, concerne un contrôle centralisé combiné à un processus décisionnel réparti. Le troisième problème met en lumière la difficulté d'établir correctement le rythme de diffusion de publicités lorsque les décisions finales en matière de diffusion sont prises de façon probabiliste par les dispositifs d'utilisateur final en présence d'une communication bilatérale lente avec perte avec ces dispositifs.

3B-I6: Stochastic Modelling in Industry Modélisation stochastique en industrie

PING WU (Bank of Montreal)

[Wednesday May 28 / mercredi 28 mai, 11:20]

Pricing Interest Rate Exotic Under LIBOR Market Model

Tarification de taux d'intérêt exotiques à l'aide du modèle de marché LIBOR

LIBOR market model (LMM) is a complicated interest rate model and it is widely used in industry. Because of the non-Markov property of the LMM, a naively implemented tree will not recombine. Thus the size of this naïve tree will grow explosively and the tree cannot be efficiently evaluated by computer simulations. This talk presents a Least Square Monte Carlo method, proposed by Longstaff-Schwartz, to price high dimensional American-style derivatives. We first rewrite the discrete version of the LMM and then carry out the calibration for the LMM. Finally, we use the Least Square Monte Carlo method for pricing interest rate exotics. Numerical results suggest that this method can produce convergent and accurate pricing results for interest rate derivatives.

Le modèle de marché du taux interbancaire pratiqué à Londres (LIBOR) est un modèle de taux d'intérêt complexe utilisé couramment en économie. En raison de la propriété non markovienne du modèle du LIBOR, un arbre implémenté de façon naïve n'est pas recombinaison. Ainsi, la taille de cet arbre naïf connaîtra une forte croissance et ne pourra pas être évalué de façon efficace par des simulations informatiques. Dans cet exposé, il est question d'une méthode de Monte Carlo des moindres carrés, proposée par Longstaff-Schwartz, qui permet de tarifier des instruments dérivés de grande dimension de style américain. Nous réécrivons d'abord la version discrète du modèle du LIBOR, puis nous procédons à sa calibration. Enfin, nous utilisons la méthode de Monte Carlo des moindres carrés pour tarifier des taux d'intérêt exotiques. Les résultats numériques suggèrent que cette méthode peut produire une tarification convergente et précise d'instruments dérivés de taux d'intérêt.

3B-C1: Actuarial Science and Finance 3 Actuariat et finance 3

Chair/Président: Chengguo Weng (University of Waterloo)

Room/Salle: UC - 161

Abstracts/Résumés

MÉLINA MAILHOT & AMANDA SPIROULIAS (Concordia University)

[Wednesday May 28 / mercredi 28 mai, 10:20]

[CANCELED] Retention Level for an Optimal Reinsurance Treaty

[ANNULÉ] Niveau de rétention pour un traité de réassurance optimal

Reinsurance treaties rely on constraints set by the cedant and the reinsurer. Conflicts of interest are inevitable between both parties. It is desirable to establish an optimal reinsurance treaty for the insurer and reinsurer, based on covered risks. In this presentation, we will illustrate results of optimal reciprocal reinsurance treaty, based on the joint survival distribution of the risks, as presented in Cai et al. (2013). We study the special case when losses are modeled by a Poisson common shock model. Specific results using this model will be shown. New problems and research areas inspired by the results will also be presented.

Les traités de réassurance s'appuient sur des contraintes définies par la cédante et le réassureur. Il est souhaitable d'établir un traité de réassurance optimal pour l'assureur ainsi que pour le réassureur, en fonction des risques encourus. Dans cette présentation, nous allons illustrer les résultats du niveau de rétention optimal, basé sur la loi de survie conjointe des risques, tel que présenté dans Cai et al. (2013). Nous étudions le cas particulier lorsque les pertes sont modélisées par un modèle à choc commun Poisson. Des résultats spécifiques à l'aide de ce modèle seront présentés. De nouveaux problèmes et domaines de recherche inspirés par les résultats seront également exposés.

PETER ADAMIC (Laurentian University)

[Wednesday May 28 / mercredi 28 mai, 10:35]

Cause-Deleted Life Expectancy Improvement with a Cure Distribution

Amélioration de l'évaluation de l'espérance de vie en traitant les causes de mortalité éliminées avec une distribution de guérison

In many circumstances the increase in life expectancy when a certain cause of death is eliminated is sought. Traditionally, these calculations have been based on the assumption that the cause in question is simply omitted; which is equivalent to the cause being taken out of consideration, from the outset, with certainty. We will propose a general model whereby a probability distribution for the cure of a specific cause of death over time is incorporated so as to more accurately predict the increase in life expectancy. The actuarial and demographic applications of the theory will be emphasized and a data set involving HIV related deaths from the State of Colorado will be considered.

Dans bon nombre de circonstances, une amélioration de l'espérance de vie est recherchée lorsqu'une certaine cause de mortalité est éliminée. Traditionnellement, ces calculs étaient basés sur l'hypothèse que la cause en question est simplement omise, ce qui équivaut avec certitude à éliminer dès le début la cause de l'ensemble des facteurs pris en considération. Nous proposons un modèle général qui incorpore une distribution de probabilité de guérison dans le temps pour une cause spécifique de mortalité afin d'arriver à prédire plus précisément l'augmentation de l'espérance de vie. Nous attirons l'attention sur des applications actuarielles et démographiques de la théorie et nous examinons un ensemble de données ayant trait aux décès en rapport avec le VIH provenant de l'État du Colorado.

3B-C1: Actuarial Science and Finance 3 Actuariat et finance 3

TAEHAN BAE & ANDREI VOLODIN (University of Regina)

[Wednesday May 28 / mercredi 28 mai, 10:50]

Generalization of Three Parameter Crack Distribution for Loss Severity Modeling

Généralisation d'une distribution de fissures à trois paramètres pour la modélisation de la sévérité des pertes

The three parameter crack distribution family which includes the inverse Gaussian, length biased inverse Gaussian and Birnbaum-Saunders distribution as its members, has received great attentions not only for failure time modeling but also for actuarial and financial applications. However, the thin-tailedness of the Gaussian crack distribution restricts the use of the model for loss severity fitting. In this talk, we will discuss the construction of a more general class of models using various stress densities including heavy tailed ones. Some key distributional properties, methods for simulation and parameter estimation, and applications to real loss data will be presented.

La famille des distributions de fissures à trois paramètres, comprenant la loi gaussienne inverse, la loi gaussienne inverse pour données biaisées en longueur et la loi de Birnbaum-Saunders, a fait l'objet d'une grande attention, à la fois pour la modélisation du temps de défaillance et pour les applications actuarielles et financières. Toutefois, en raison de sa queue légère, la distribution gaussienne des fissures restreint l'utilisation du modèle aux fins d'ajustement de sévérité des pertes. Dans cet exposé, nous discutons l'établissement d'une classe de modèles plus générale utilisant diverses densités de stress, y compris celles à queue lourde. Nous présentons quelques éléments clés, notamment des propriétés de distribution, des méthodes de simulation et d'estimation de paramètres, et des applications à des données de pertes réelles.

JOSE GARRIDO (Concordia University), **CHRISTIAN GENEST & JULIANA SCHULZ** (McGill University)

[Wednesday May 28 / mercredi 28 mai, 11:05]

Generalized Linear Models for Dependent Claim Frequency and Severity

Modèles linéaires généralisés pour fréquence et sévérité de pertes dépendantes

An alternative approach to modelling the expected loss cost of an insurance portfolio is developed, allowing for dependence between the claim frequency and severity. A Generalized Linear Model (GLM) framework is developed for the aggregate claims model in the dependent setting using a conditional severity model and marginal frequency model. We find that the pure premium in the dependent aggregate claims model is the product of a marginal mean frequency, a modified marginal mean severity and a correction term for dependence. This dependence model is then compared with the independent GLM aggregate claims model. It is shown that the expected total loss amount derived in the independent model is in fact a special case of the dependent model.

Nous proposons une approche alternative à la modélisation des pertes espérées d'un portefeuille d'assurance, permettant la dépendance entre fréquence et sévérité. Un modèle linéaire généralisé (GLM) est défini pour les pertes agrégées dans le cas de dépendance, en utilisant un modèle de sévérité conditionnelle et de fréquence marginale. Nous montrons que la prime pure dans le modèle de dépendance pour les pertes agrégées est égale au produit d'un terme de fréquence moyenne marginale, d'une sévérité moyenne marginale modifiée et d'un terme de correction due à la dépendance. Ce modèle de dépendance est comparé au modèle usuel GLM indépendant. Cela permet de voir que la moyenne des pertes agrégées dans le modèle d'indépendance est un cas spécial du modèle de dépendance.

3B-C1: Actuarial Science and Finance 3 Actuariat et finance 3

JIANDONG REN (Western University)

[Wednesday May 28 / mercredi 28 mai, 11:20]

Analysis of IBNR with Markovian Arrival Processes

Analyse des réclamations END par processus d'arrivée de Markov

This paper analyzes the claims payment process of insurance companies. It is assumed that the insurance claims occur according to a Markovian arrival process. An incurred claim is categorized into various states that belong to a finite state space. The state space is divided into several sub state spaces, which represents the different stages of claim reporting and settlement, such as Incurred but not reported (IBNR), Reported but not settled (RBNS) and Settled (S). A claim changes states while the claim reporting and settling process evolves. The transitions among states are assumed to follow a Markovian law. We derive formulas for the joint distribution and the joint moments for the amount of INBR, RBNS and Settled claims.

Cet article analyse le processus de paiement des réclamations des compagnies d'assurance. Nous supposons que les réclamations d'assurance suivent un processus d'arrivée de Markov. Un sinistre encouru est catégorisé dans l'un des divers états d'un espace fini d'états. Ce dernier est divisé en plusieurs sous-espaces d'états, qui représentent les différentes étapes de déclaration et de traitement des sinistres, comme les sinistres encourus mais non déclarés (END), déclarés mais non encore réglés (DNR) et réglés (R). Une réclamation passe d'un état au suivant à mesure que le processus de déclaration et de traitement évolue. Nous assumons que les transitions entre états suivent une loi de Markov. Nous dérivons des formules pour la distribution jointe et les moments joints du nombre de réclamations END, DNR et R.

ALAN XINGHUA ZHOU & R. MARK REESOR (Western University)

[Wednesday May 28 / mercredi 28 mai, 11:35]

Misrepresentation and Capital Structure: Quantifying the Impact on Corporate Security Value in a Modified Merton Framework

Fausses déclarations et structure du capital : comment quantifier l'impact sur la valeur des titres de société dans un modèle de Merton modifié

Securities class actions typically involve misrepresentation by a firm that overstates its true value. The share price of the firm drops when the misrepresentation is revealed. The share price drop is used to assess potential damages to shareholders. Here we use a modified Merton framework to measure the impact of misrepresentation on the value of other components (e.g., debt, warrants) of a firm's capital structure. Using a relationship between equity and firm value, we show how observable equity information can be used to determine firm value and hence the value of other securities in the capital structure. Thus the effect of misrepresentation on firm value and the capital structure constituents can be measured from the observable drop in share price.

Les recours collectifs en matière de valeurs mobilières sont souvent déclenchés par de fausses déclarations de la part d'une entreprise qui exagère sa vraie valeur. Le cours de l'action de l'entreprise chute lorsque ces fausses déclarations sont révélées. Cette chute est utilisée pour évaluer les dommages éventuels pour les actionnaires. Ici, nous utilisons un modèle de Merton modifié pour mesurer l'impact des fausses déclarations sur la valeur d'autres éléments (p. ex., créances, bons de souscription) de la structure de capital d'une entreprise. En nous fondant sur la relation entre les fonds propres et la valeur de l'entreprise, nous montrons comment utiliser des informations observables sur les fonds propres pour déterminer la valeur de l'entreprise et la valeur des autres titres dans la structure de capital. Nous pouvons ainsi mesurer l'effet des fausses déclarations sur la valeur de l'entreprise et les éléments de la structure du capital à partir de la chute observable du cours de l'action.

3B-C2: Biostatistics - Survival, Longitudinal and Event History Analysis Biostatistique - analyses de survie, longitudinales et d'historique d'événements

Chair/Président: Joel Dubin (University of Waterloo)

Room/Salle: UC - 140

Abstracts/Résumés

GUN HO JANG (Ontario Institute for Cancer Research)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Convergence Rate of Scan Statistics for Rare Copy Number Variation Detection

Taux de convergence des statistiques de scan pour la détection de variations du nombre de copies rares

Tests with interval alternative hypothesis is common in biostatistics. Scan methodology is developed for detecting maximal variant statistic. The tail probability of scan statistics can be approximated using large deviation probability which shrinks a lot faster than the actual convergence rate. Accurate convergence rate is derived. Some tail probability approximations are compared in several simulation schemes.

En biostatistique, il est courant d'utiliser des tests comportant une hypothèse alternative d'intervalle. Une méthodologie basée sur le scan est élaborée pour repérer les statistiques de variante maximale. Les probabilités codales des statistiques de scan peuvent être approximée en utilisant le principe des grandes déviations dont la compression est plus rapide que le taux de convergence. Un taux de convergence précis est dérivé. Des approximations de probabilité dans les queues sont comparées selon divers schémas de simulation.

ELHAM KHODAYARI MOEZ & ELHAM KHODAYARI MOEZ (University of Alberta), **FARID ZAYERI** (Shahid Beheshti University of Medical Sciences), **IRINA DINU** (University of Alberta)

[Wednesday May 28 / mercredi 28 mai, 10:35]

The Accuracy of Joint Analysis of Survival and Longitudinal Data: A Simulation Study

Précision de l'analyse conjointe de données de survie et de données longitudinales : étude de simulation

Joint modelling of longitudinal and survival data is becoming increasingly popular in epidemiological studies.

We used the simulation method to assess the effect of sample size, number of repeated measures, censoring rate and association parameter on the accuracy of the joint analysis. In different conditions, the joint model was compared with the Cox models with/without time-dependent variable.

In the estimation of the survival function, the joint model performed superior to other models, except when there was a small number of repeated measures and a high censoring rate. For prediction of the individual survival probabilities, the joint model performed as accurately as the Cox model without time-dependent variable if the hazard ratio corresponding to longitudinal outcome was close to one.

La modélisation conjointe de données longitudinales et de données de survie gagne en popularité dans les études épidémiologiques.

Nous avons utilisé la méthode de simulation pour évaluer l'effet de la taille d'échantillon, du nombre de mesures répétées, du degré de censure et des paramètres d'association de la précision de l'analyse combinée. Le modèle a été comparé dans des conditions différentes avec les modèles Cox, avec ou sans variables chronologiques.

Dans l'estimation de la fonction de survie, le modèle combiné a surpassé les autres modèles, sauf en présence d'un petit nombre de mesures répétées et d'un haut degré de censure. Pour la prévision des probabilités individuelles de survie, la précision du modèle combiné a été comparable à celle du modèle de Cox sans variables chronologiques lorsque le rapport de risques correspondant aux résultats longitudinaux était proche de un.

3B-C2: Biostatistics - Survival, Longitudinal and Event History Analysis

Biostatistique - analyses de survie, longitudinales et d'historique d'événements

ELEANOR PULLENAYEGUM (Hospital for Sick Children)

[Wednesday May 28 / mercredi 28 mai, 10:50]

Methods for Analysing Irregular Longitudinal Data: How do the Assumptions Match Up with Visit Processes Encountered in Practice

Méthodes d'analyse de données longitudinales irrégulières : les hypothèses correspondent-elles aux processus de visite rencontrés dans la pratique?

Longitudinal data often features variable follow-up times, with no two patients sharing the same set of visit times. Moreover, the timing of the measurements is associated with disease course; for example, patients may visit more often when unwell. Statistical methods can help overcome the resulting bias, however these rely on assumptions about the dependence between visit times and outcome processes, and these assumptions differ across methods. We propose a classification system for visit processes encountered in practice, discuss how common modelling assumptions line up with the visit classification, and make some recommendations for the design and analysis of such studies.

Les données longitudinales incluent souvent des délais de consultation de suivi variables, sans que deux patients ne partagent le même ensemble de temps de visite. De plus, le moment des mesures est associé à l'évolution de la maladie ; ainsi, les patients consultent plus souvent lorsqu'ils ne se sentent pas bien. Il existe des méthodes statistiques permettant de surmonter le biais résultant, toutefois celles-ci reposent sur des hypothèses sur la dépendance entre les temps de visite et les processus résultants, hypothèses qui varient d'une méthode à l'autre. Nous proposons un système de classification pour les processus de visite rencontrés dans la pratique, discutons de l'adéquation des hypothèses de modélisation et de la classification des visites et émettons des recommandations concernant la conception et l'analyse de ces études.

DENIS TALBOT, GENEVIÈVE LEFEBVRE & JULI ATHERTON (Université du Québec à Montréal)

[Wednesday May 28 / mercredi 28 mai, 11:05]

A Cautionary Note on the Use of Stabilized Weights in Marginal Structural Models

Un avertissement concernant l'utilisation des poids stabilisés pour les modèles structureaux marginaux

Marginal structural models (MSMs) are commonly used to estimate the effect of a time-varying exposure in the presence of time-dependent confounding. When fitting a MSM to data, an analyst must specify both the inverse probability of treatment weights model and the marginal structural model. When implementing these models, the use of stabilized weights is recommended since they are usually less variable than standard weights. We are concerned with the use of stabilized weights in a situation where the structural model is specified to consider only partial treatment history information. We present various examples of settings where stabilized weights yield biased inferences while standard weights do not.

Les modèles structureaux marginaux (MSM) sont fréquemment utilisés pour estimer l'effet d'un traitement variant dans le temps en présence de confusion dépendante du temps. Afin d'ajuster un MSM aux données, l'analyste doit spécifier à la fois le modèle de pondération par l'inverse des probabilités de traitement et le MSM. Lors de l'implantation de ces modèles, l'utilisation des poids stabilisés est préconisée puisque ceux-ci sont généralement moins variables que les poids non stabilisés. Nous nous intéressons à l'utilisation des poids stabilisés lorsque le modèle structurel n'inclut qu'une partie de l'information sur l'historique de traitement. Nous présentons différentes situations où les poids stabilisés donnent des inférences biaisées, contrairement aux poids non stabilisés.

3B-C2: Biostatistics - Survival, Longitudinal and Event History Analysis Biostatistique - analyses de survie, longitudinales et d'historique d'événements

XIN XIN, JULIE HORROCKS & GERARDA DARLINGTON (University of Guelph)

[Wednesday May 28 / mercredi 28 mai, 11:20]

A Likelihood Approach to Accommodate Ties between Events and Covariate Change Times in Cox Proportional Hazards Models

Approche de vraisemblance pour tenir compte des liens entre temps d'événements et de changement de covariables dans les modèles des risques proportionnels de Cox

Previous work has demonstrated that ties between event times and the change times of a binary time-varying covariate can introduce bias to Cox proportional hazards model estimation. In this presentation, we present a novel partial likelihood expression based on the Cox partial likelihood and the Efron approximation, that is able to accommodate both ties among event times as well as ties between event times and covariate change times. Compared to the previously proposed Random Jitter (RJ) and Equally Weighted (EW) methods, the proposed partial likelihood enables a one-step model fitting process using the original data.

De précédents travaux ont montré que les liens entre les temps d'événements et de changement d'une covariable binaire variant dans le temps peuvent introduire un biais dans l'estimation du modèle des risques proportionnels de Cox. Dans cette présentation, nous présentons une expression de vraisemblance partielle nouvelle fondée sur la vraisemblance partielle de Cox et l'approximation d'Efron, qui est capable de tenir compte des liens entre temps d'événements et les liens entre temps d'événements et temps de changement de covariable. Par rapport aux méthodes précédentes de gigue aléatoire et de pondération égale, la vraisemblance partielle proposée permet un processus d'ajustement du modèle en une étape avec les données originales.

YAYUAN ZHU, JERRY LAWLESS & CECILIA COTTON (University of Waterloo)

[Wednesday May 28 / mercredi 28 mai, 11:35]

Event History Analysis in Observational Studies with Intermittent Observation

Analyse de l'historique des événements dans les études d'observation intermittente

In clinical experiments, variables are usually measured at regular and pre-specified time points. Then the observation process is marginally independent of the outcome process. However, in most longitudinal cohort studies some subjects may miss scheduled visits or choose to visit a clinic at irregular time points. Moreover, the frequency and timing of visits may also be associated with outcomes or outcome-related variables, including the previous observation history. We will discuss event history analysis in this setting and show how an informative observation scheme can produce bias in standard methods, and we aim to study how to handle the incomplete data and how to appropriately adjust for the outcome-dependent follow-up. The Toronto Psoriatic Arthritic Clinic Cohort Study is used for illustration.

Dans les expériences cliniques, les variables sont généralement mesurées à des moments réguliers préspecifiés. Le processus d'observation est alors marginalement indépendant du processus de résultat. Cependant, dans la plupart des études de cohorte longitudinales, certains sujets ratent des visites programmées ou choisissent de visiter une clinique de manière irrégulière. De plus, la fréquence et le moment des visites peuvent également être associés aux résultats ou à des variables liées aux résultats, dont l'historique d'observation antérieur. Nous discutons de l'analyse de l'historique des événements dans ce cadre et montrons comment un plan d'observation informatif peut produire un biais dans les méthodes standard. Nous cherchons à étudier comment traiter les données incomplètes et comment ajuster le modèle de manière appropriée en fonction du suivi dépendant des résultats. Nous illustrons notre approche par la Toronto Psoriatic Arthritic Clinic Cohort Study.

3B-C3: Some Specific Issues in Modelling and Estimation Quelques problématiques spécifiques en modélisation et estimation

Chair/Président: Fang Yao (University of Toronto)

Room/Salle: UC - A101

Abstracts/Résumés

ELIF ACAR (University of Manitoba)

[Wednesday May 28 / mercredi 28 mai, 10:20]

Conditional Copula Models with Multiple Covariates

Modèles à copules conditionnelles avec covariables multiples

This talk will present a nonparametric modelling strategy that can accommodate multiple covariates in conditional copula models. We consider a semiparametric conditional copula model where the copula function belongs to a parametric copula family and the copula parameter varies smoothly with the values of covariates. To alleviate the curse of dimensionality, we use an additive formulation of the copula parameter and estimate smooth component functions associated with each covariate via a local likelihood backfitting algorithm. The finite sample performance of the proposed approach will be demonstrated using simulated and real data. The talk will also address general identifiability restrictions and computational challenges.

Dans cet exposé, nous décrivons une stratégie de modélisation non paramétrique qui peut tenir compte des covariables multiples dans des modèles à copules conditionnelles. Nous examinons un modèle à copules conditionnelles semiparamétrique, où la fonction de copule appartient à une famille de copules paramétriques, et où le paramètre de copules varie uniformément en fonction des valeurs des covariables. Pour atténuer le fléau de la dimensionalité, nous utilisons une formulation additive du paramètre de copule et nous estimons des fonctions lisses associées à chaque covariable au moyen d'un algorithme de «backfitting» de la vraisemblance local. L'utilisation de données réelles et simulées démontrent la performance de l'échantillon fini dans l'approche proposée. Il est aussi question des restrictions d'identifiabilité et des défis au niveau des calculs.

OTHMANE KORTBI & ERIC MARCHAND (Université de Sherbrooke)

[Wednesday May 28 / mercredi 28 mai, 10:35]

Estimating a Multivariate Normal Mean with a Bounded Signal to Noise Ratio

Estimation d'une moyenne normale multivariée avec un rapport signal-bruit borné

For normal models with $X \sim N_p(\theta, \sigma^2 I_p)$, $S^2 \sim \sigma^2 \chi_k^2$, independent, we consider the problem of estimating θ under scale invariant squared error loss $\frac{\|d-\theta\|^2}{\sigma^2}$, when it is known that the signal-to-noise ratio $\frac{\|\theta\|}{\sigma}$ is bounded above by m .

Risk analysis is achieved by making use of a conditional risk decomposition and we obtain in particular sufficient conditions for an estimator to dominate either the unbiased estimator $\delta_{UB}(X) = X$, or the maximum likelihood estimator $\delta_{ML}(X, S^2)$, or both of these benchmark procedures.

Among our results, we show that, whenever $m \leq \sqrt{p}$ and $p \geq 2$, the boundary Bayes estimator $\delta_{BU,0}$

Pour des modèles normaux avec $X \sim N_p(\theta, \sigma^2 I_p)$, $S^2 \sim \sigma^2 \chi_k^2$, indépendant, nous analysons le problème de l'estimation de θ sous une perte quadratique invariante à l'échelle $\frac{\|d-\theta\|^2}{\sigma^2}$, lorsqu'il est établi que le rapport signal-bruit $\frac{\|\theta\|}{\sigma}$ a une borne supérieure m .

L'analyse de risque est réalisée en utilisant une décomposition conditionnelle du risque, et nous obtenons notamment des conditions suffisantes pour qu'un estimateur domine soit l'estimateur non biaisé $\delta_{UB}(X) = X$, ou l'estimateur au maximum de vraisemblance $\delta_{ML}(X, S^2)$, ou l'une et l'autre de ces deux procédures de référence.

Nos résultats montrent que lorsque $m \leq \sqrt{p}$ et $p \geq 2$, l'estimateur bayésien sur la frontière $\delta_{BU,0}$ domine à la fois δ_{UB}

3B-C3: Some Specific Issues in Modelling and Estimation

Quelques problématiques spécifiques en modélisation et estimation

dominates both δ_{UB} and δ_{ML} . Various other dominance results are obtained, illustrations are provided and commented upon.

et δ_{ML} . Divers autres résultats de dominance sont obtenus, et des illustrations sont fournies et commentées.

ZHENHUA LIN (University of Toronto), **JIGUO CAO & LIANGLIANG WANG** (Simon Fraser University), **HAONAN WANG** (Colorado State University)

[Wednesday May 28 / mercredi 28 mai, 10:50]

A Smooth and Locally Sparse Estimator for Functional Linear Regression via Functional SCAD Penalty

Estimateur lisse et localement parcimonieux pour une régression linéaire fonctionnelle au moyen d'une pénalité fonctionnelle pour l'écart absolu avec coupure lisse (SCAD).

We propose a new regularization technique called "functional SCAD", and combine it with the smoothing spline method to develop a smooth and locally sparse (i.e., zero on some sub-regions) estimator for the coefficient function in functional linear regression. The functional SCAD has a nice shrinkage property that enables us to identify the null sub-regions of the coefficient function without over shrinking the non-zero sub-regions. Compared to other methods, our method is theoretically sounder and computationally simpler. Asymptotic analysis shows that our estimator enjoys the oracle property, which means asymptotically it performs as well as if we knew the true null sub-regions. The method has superior numerical performance shown by simulation studies. Its practical merit is demonstrated on Canadian weather data.

Nous proposons une nouvelle technique de régularisation appelée «SCAD fonctionnel» (écart absolu avec coupure lisse fonctionnel), et nous la combinons à une méthode d'estimation lisse par spline afin d'élaborer un estimateur lisse et localement parcimonieux (c.-à-d. zéro dans certaines sous-régions) pour la fonction de coefficient d'une régression linéaire fonctionnelle. Le SCAD fonctionnel est doté d'une propriété de régularisation intéressante qui nous permet d'identifier les sous-régions nulles de la fonction de coefficient sans surcompression des sous-régions non nulles. En comparaison avec les autres méthodes, la nôtre est plus robuste sur le plan théorique, et plus simple en matière de calcul. L'analyse asymptotique montre que notre estimateur profite de la propriété oracle, ce qui signifie que sur ce plan, sa performance est aussi bonne que si nous connaissions les vraies sous-régions nulles. Les études de simulation révèlent que cette méthode réalise une performance numérique supérieure. Sur le plan pratique, ses qualités sont démontrées avec des données sur le climat canadien.

IRAJ YADEGARI & ÉRIC MARCHAND (Université de Sherbrooke), **FRANÇOIS PERRON** (Université de Montréal)

[Wednesday May 28 / mercredi 28 mai, 11:05]

Parameter Estimation under Bounded Loss Functions

Estimation de paramètres sous fonctions de pertes limitées

For the model $X \sim N_p(\theta, \sigma^2 I_p)$ and restriction $\|\theta\| \leq m$, for some known $m > 0$, we consider estimating θ under reflected-normal loss function $1 - \exp\{-\|\delta - \theta\|^2/2c^2\}$. Our goal is to obtain improvement on the maximum likelihood estimator $\delta_{mle}(X) = (m \wedge \|X\|)X/\|X\|$. For instance, we show that the boundary uniform prior Bayes estimator dominates $\delta_{mle}(X)$ for $m \leq \sqrt{c^2/(c^2+1)}$, with $c \rightarrow +\infty$ recovering results of [Marchand and Perron, Ann. Statist. (2001), 29, 1078-1093] for squared error loss function. This problem is also

Pour le modèle $X \sim N_p(\theta, \sigma^2 I_p)$ et la limite $\|\theta\| \leq m$, pour un $m > 0$ connu, nous cherchons à estimer θ sous la fonction de perte normale reflétée $1 - \exp\{-\|\delta - \theta\|^2/2c^2\}$. Notre objectif est d'obtenir une amélioration de l'estimateur du maximum de vraisemblance $\delta_{mle}(X) = (m \wedge \|X\|)X/\|X\|$. Ainsi, nous montrons que l'estimateur bayésien a priori uniforme de limite domine $\delta_{mle}(X)$ pour $m \leq \sqrt{c^2/(c^2+1)}$, avec $c \rightarrow +\infty$ recouvrant les résultats de [Marchand et Perron, Ann. Statist. (2001), 29, 1078-1093] pour la fonction de perte d'erreur quadratique. Ce problème est également motivé par le problème d'estimation de la densité prédictive sous des

3B-C3: Some Specific Issues in Modelling and Estimation

Quelques problématiques spécifiques en modélisation et estimation

motivated by predictive density estimation problem under Integrated L_2 and Hellinger distances which we will describe.

distances L_2 et de Hellinger intégrées, que nous décrivons.

TINGTING YU & LANG WU (University of British Columbia)

[Wednesday May 28 / mercredi 28 mai, 11:20]

Jointly Modeling Longitudinal Process with Measurement Errors, Missing Data, and Outliers

Modélisation conjointe de processus longitudinaux avec erreurs de mesure, données manquantes et données aberrantes

In many longitudinal studies, several longitudinal processes may be associated. In such cases, a joint inference is appealing since it can incorporate information of all processes simultaneously. In addition, longitudinal data often contain outliers, so robust methods for the joint models are necessary. In the presentation, we will discuss joint models for two correlated longitudinal processes with measurement errors, missing data, and outliers. We will also introduce our proposed robust method based on M-estimators, which addresses possible outliers for joint models. Simulation studies are conducted to evaluate the performances of the proposed method, and a real AIDS dataset is analyzed based on the method as well.

Dans de nombreuses études longitudinales, plusieurs processus longitudinaux sont associés. Dans ces cas, il est tentant de procéder à une inférence conjointe qui inclurait des informations sur tous les processus en même temps. De plus, les données longitudinales incluent souvent des données aberrantes, si bien qu'il est nécessaire d'établir des méthodes robustes pour les modèles conjoints. Dans cette présentation, nous discutons de modèles conjoints pour deux processus longitudinaux corrélés avec erreurs de mesure, données manquantes et données aberrantes. Nous introduisons également notre méthode robuste fondée sur les M-estimateurs qui répond au problème de données aberrantes pour les modèles conjoints. Nous effectuons des études par simulation pour évaluer la performance de la méthode proposée et l'utilisons pour analyser un ensemble de données réelles sur le SIDA.

DAVID RIEGERT, AARON SPRINGFORD & DAVID THOMSON (Queen's University at Kingston)

[Wednesday May 28 / mercredi 28 mai, 11:35]

Forecasting the Likelihood of Solar Flares Using an Inferred Solar Stress Index

Comment prévoir la vraisemblance d'une éruption solaire à l'aide d'un indice de stress solaire

Solar activity can have damaging effects on Earth-based systems. Power companies in particular are interested in predicting when solar storms will occur in order to mitigate damage to transmission systems. I will present a forecasting approach based on the hypothesis that subsurface stresses initiate solar flare events, which can result in coronal mass ejections and solar storms at Earth. The approach uses multitaper spectral estimates of solar gravity modes to forecast a cumulative solar stress index using a state-space model. Using this index, a six month qualitative forecast of the likelihood of large flare events was generated. I conclude by discussing the performance of the forecast and its utility for management of power transmission sys-

L'activité solaire peut avoir des effets dommageables pour les systèmes terrestres. Les compagnies d'électricité, en particulier, s'intéressent à la prédiction des tempêtes solaires afin de limiter les dommages à l'équipement du réseau de transport. Je présente une approche de prévision fondée sur l'hypothèse selon laquelle des stress dans la subsurface du soleil initient des éruptions solaires qui peuvent donner lieu à des éjections de masse coronale et des tempêtes solaires sur Terre. Cette méthode utilise des estimations spectrales « multitaper » des modes de gravité solaire pour prévoir un indice de stress solaire cumulé à l'aide d'un modèle d'espace d'états. Cet indice permet de générer une prévision qualitative à six mois de la vraisemblance de grosses éruptions. Je conclus en discutant de la performance de cette prévision et de son utilité pour la gestion à long terme

3B-C3: Some Specific Issues in Modelling and Estimation

Quelques problématiques spécifiques en modélisation et estimation

tems over an extended time horizon.

des équipements des réseaux de transport.

3C-A1: CJS Award Address
Allocution du récipiendaire du Prix de la RCS

Chair/Président: Jiahua Chen (University of British Columbia)
Organizer/Responsable: Denis Larocque (HEC Montreal)

Room/Salle: Hart House - Debates Rm

Abstract/Résumé

ART OWEN (Stanford University)

[Wednesday May 28 / mercredi 28 mai, 13:30]

Self-concordance for Empirical Likelihood

Auto-concordance pour la vraisemblance empirique

Empirical likelihood is a method of likelihood inference that does not require one to specify a parametric family for the data. Confidence regions for the mean of a vector reduce to a convex optimization. More general statistics are then handled via estimating equations. This work exhibits an approach to empirical likelihood computation via a self-concordant convex function. Newton's method with backtracking is guaranteed to converge to the global optimum for self-concordant strictly convex functions. Self-concordance is crucial since it keeps the Hessian matrix from behaving too unpredictably. A Bartlett correctable polynomial approximation for the empirical log likelihood, due to Corcoran, is also self-concordant.

La vraisemblance empirique est une méthode d'inférence qui ne nécessite pas que l'on spécifie une famille paramétrique pour les données. Les régions de confiance pour la moyenne d'un vecteur sont réduites à une optimisation convexe. Des statistiques plus générales sont ensuite traitées par des équations d'estimation. Ce travail présente une approche de calcul de vraisemblance empirique à l'aide d'une fonction convexe auto-concordante. La méthode de Newton avec retour en arrière est assurée de converger vers un optimum global pour les fonctions strictement convexes auto-concordantes. L'auto-concordance est cruciale puisqu'elle empêche la matrice hessienne de se comporter de manière trop imprévisible. Une approximation polynomiale corrigée de Bartlett pour le log de vraisemblance empirique, qui en raison de Corcoran, est aussi auto-concordante.

3C-A2: Isobel Loutit Invited Address Allocution invitée Isobel Loutit

Organizer and Chair / Responsable et président: Thomas Loughin (Simon Fraser University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: Hart House - Music Rm

Abstract/Résumé

BRADLEY JONES (JMP)

[Wednesday May 28 / mercredi 28 mai, 13:30]

What You Learn Depends on Where You Look

Ce que vous apprenez est fonction de ce que vous observez

Establishing causal relationships may be impossible without experimental studies. So, it is important to plan experimental studies to extract the most precise and unambiguous information about a system or a process. Understanding the ways variance and bias may be reduced through the choice of statistical model and experimental design is key to efficient generation of knowledge. This talk provides examples to demonstrate how to increase precision and reduce bias in experimental studies. My purpose is to empower the broader use of designed experiments in any field where it is possible to vary inputs and measure outputs of interest.

En l'absence d'études expérimentales, il peut être impossible d'établir des relations de causalité. Il est donc important de planifier les études expérimentales ayant trait à un système ou à un procédé afin d'en extraire l'information la plus précise et la moins ambiguë qui soit. La compréhension des façons dont la variance et le biais peuvent être réduits par le choix d'un modèle statistique et d'un plan d'expérience sont des facteurs clés de l'acquisition efficace de connaissances. Le présent exposé comprend des exemples qui démontrent des moyens pour améliorer la précision et réduire les biais des études expérimentales. Mon objectif est de favoriser une plus grande utilisation des expériences planifiées dans divers domaines dans lesquels il est possible de varier les données d'entrée et de mesurer les résultats qui présentent un intérêt.

3C-A3: Pierre Robillard Award Address
Allocution du récipiendaire du Prix Pierre-Robillard

Organizer and Chair / Responsable et président: Wenqing He (Western University)

Room/Salle: Hart House - East Common

Abstract/Résumé

LIQUN DIAO (University of Rochester)

[Wednesday May 28 / mercredi 28 mai, 13:30]

Copula Models for Multi-type Life History Processes

Modèles de copules pour les processus d'historiques multiples

Innovative models are developed for the analysis of multifaceted life history processes with the features that joint models facilitate direct inferences about marginal processes and dependence structures are modeled separately. These features are obtained by careful decomposition of joint densities, explicit conditional independence assumptions, and the use of copula functions. Likelihood and composite likelihood approaches are applied under simultaneous and two-stage estimation, and the approaches to inference are compared in terms of efficiency and robustness to model misspecification. Particular attention is given to marked point processes and multiple multistate processes. Motivating studies in transfusion medicine and in arthritis provide an opportunity to illustrate the application of these methods.

Des modèles innovateurs sont développés pour l'analyse de processus d'historiques à facettes multiples avec les caractéristiques que les modèles communs facilitent les inférences directes sur les processus marginaux et les structures de dépendances sont modélisées séparément. Ces caractéristiques sont obtenues par une décomposition minutieuse des densités conjointes, des hypothèses d'indépendance conditionnelles explicites, et de l'utilisation de fonctions de copules. Les approches de vraisemblance et de vraisemblance composite sont appliquées sous des estimations simultanées et en deux étapes, et les approches d'inférence sont comparées en terme d'efficacité et de robustesse aux erreurs de spécification du modèle. Une attention particulière est accordée aux processus ponctuels marqués et aux nombreux processus multi-états. Des études prometteuses en médecine transfusionnelle et en arthrite offrent une opportunité de démontrer l'application de ces méthodes.

3C-A4: Survey Methods Section Presidential Invited Address
Allocution de l'invité du président du Groupe des méthodes d'enquête

Organizer and Chair / Responsable et présidente: Susana Rubin-Bleuer (Statistics Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: UC - 140

Abstract/Résumé

CLAIRE DURAND (Université de Montréal)

[Wednesday May 28 / mercredi 28 mai, 13:30]

Surveys and Society

Sondages et société

What influence does society and its evolution have on survey research? What influence does survey research and the results it produces have on society? We rarely have time to think about these topics. What is the impact of recent changes in modes of communication on survey research? What is or will be the impact of population diversity on the categories that we use? Do statistics that show that some social groups -- like Aborigines and visible minorities -- are disadvantaged compared to other Canadians have an impact on the situation of these groups? Those are not technical questions, but social and sociological ones, but they have much to do with the work of survey statisticians.

Quelle est l'influence de la société et de son évolution sur la recherche par sondage? Quelle est l'influence de la recherche par sondage et de ses résultats sur la société? Nous avons rarement le temps de nous arrêter à ces questions. Quel est l'impact des changements récents dans les modes de communication sur la recherche? Quel est, quel sera l'impact de la diversification de la population sur les catégories utilisées? Les statistiques qui montrent que certains groupes sociaux -- Autochtones, minorités visibles -- sont désavantagés comparés aux autres Canadiens influencent-elles la situation de ces groupes? Ce ne sont pas des questions techniques mais des questions sociales, sociologiques, qui ont toutefois une relation importante avec le travail des statisticiens d'enquête.

3D-I1: Bridging the Gap between Theory and Practice in Statistics Comblent l'écart entre la théorie et la pratique en statistique

Organizer and Chair / Responsable et président: Fernando Camacho (University of Waterloo)

Session sponsored by the Accreditation Committee

Séance parrainée par le Comité d'accréditation

Room/Salle: Hart House - South Dining

Abstracts/Résumés

MICHAL ABRAHAMOWICZ (McGill University)

[Wednesday May 28 / mercredi 28 mai, 15:00]

Flexible Modelling of Cumulative Effects of Medications Use: Fighting with 2 Dragons (Analytical Challenges and Clinical Reviewers)

Modélisation flexible des effets cumulatifs de la consommation de médicaments: combat entre deux dragons (défis analytiques et évaluateurs cliniques)

Population-based studies of adverse or therapeutic effects of medications in real-life clinical practice are of paramount importance for both public health and society at large. Analysis of such observational, longitudinal studies has to account for variation over time in the duration, timing and dose of the drugs under study, and their cumulative effects. We propose the weighted cumulative exposure (WCE) model: $WCE(\tau|x(t), t < \tau) = \sum w(\tau t) * x(t)$, where the weight function $w(\tau t)$ assigns weights to past doses $X(t)$ and is modelled with cubic regression splines. We then apply the method to assess effects of 3 medications, focusing on (i) innovative methods to make the results interpretable to medical researchers, and (ii) (challenging) discussions with reviewers of clinical journals.

Les études basées sur la population portant sur les effets indésirables ou thérapeutiques des médicaments en pratique clinique réelle sont d'une importance capitale pour la santé publique et l'ensemble de la société. L'analyse de telles études observationnelles longitudinales doit tenir compte de la variation dans le temps de la durée, du calendrier et de la dose des médicaments à l'étude, ainsi que de leurs effets cumulatifs. Nous proposons le modèle pondéré de l'exposition cumulative (WCE) : $WCE(\tau|x(t), t < \tau) = \sum w(\tau t) * x(t)$, où la fonction de pondération $w(\tau t)$ attribue des poids aux doses antérieures $X(t)$ et est modélisée par des splines de régression cubiques. Nous appliquons ensuite la méthode pour évaluer les effets de trois médicaments, en insistant sur (i) des méthodes novatrices permettant l'interprétation des résultats par des chercheurs du domaine médical et (ii) des discussions (animées) avec des évaluateurs de revues cliniques.

PINA D'ANGELO (Novum Pharmaceutical)

[Wednesday May 28 / mercredi 28 mai, 15:30]

Data Analysis Concepts and Issues in Skin Adhesion and Irritation/Sensitization Studies

Concepts et problématiques d'analyse de données tirées d'études portant sur l'adhérence sur la peau, l'irritation et la sensibilisation

Skin adhesion and irritation/sensitization studies are required by regulatory authorities to demonstrate that the adhesion, irritation and sensitization properties of the generic test product are not inferior to those of the reference product. To demon-

Les études portant sur l'adhérence sur la peau, l'irritation et la sensibilisation sont exigées par les autorités réglementaires afin de démontrer que ces propriétés du produit générique testé ne sont pas inférieures à celles du produit de référence. Pour démontrer la non-infériorité du test par

3D-I1: Bridging the Gap between Theory and Practice in Statistics Comblent l'écart entre la théorie et la pratique en statistique

strate non-inferiority of the test compared to the reference product with regard to the cumulative irritation and adhesion scores, the FDA recommends that the upper bound of the one-sided 95% confidence interval of the mean test product score minus 1.25 times the mean RLD score must be less than or equal to 0. This requirement is stringent for well-adhering and low-irritating products. Various statistical procedures for this type of data analysis, including parametric and non-parametric approaches, will be presented.

rapport au produit de référence en matière de scores d'irritation cumulative et d'adhérence, la FDA recommande que la différence entre la borne supérieure de l'intervalle de confiance unilatéral à 95 % du score moyen du produit testé moins 1,25 fois le score RLD moyen soit inférieure ou égale à zéro. Cette exigence est rigoureuse pour les produits bien adhérents et peu irritants. Nous présentons diverses procédures statistiques pour ce type d'analyse de données, y compris des méthodes paramétriques et non paramétriques.

JANET E.A. MCDUGALL, HUANG QIANYI, AMRHEIN JOHN & CHEN HONG (McDougall Scientific Ltd.)

[Wednesday May 28 / mercredi 28 mai, 16:00]

Adaptive Randomization in a Seamless Phase II/Phase III Dose-Finding and Confirmatory Trial Design

Randomisation adaptative dans le cadre d'un essai clinique ininterrompu de phases II et III visant l'établissement et la confirmation de la posologie

Randomizing patients in a Phase II based on accumulating evidence identifying which dose(s) provide the most desirable profile, as per a matrix of efficacy and toxicity combinations rated by expert oncologists (0-100) from low efficacy/high toxicity to high efficacy/low toxicity; experts also define the region of 'acceptable' utility scores. Each patient is rated for efficacy and toxicity at a specified time-point. Simulation work defines the priors; combined with ratings these form the posteriors using the joint distribution. Resulting joint probabilities are estimated and those in the 'acceptable region' used to calculate the mean utility for each dose; the optimal dose is one with the maximum utility score. Randomization proportional to the resulting probabilities. Dose appropriate data used for Phase III.

La randomisation de patients en phase II est basée sur l'accumulation d'information permettant d'identifier les doses qui présentent les meilleurs profils selon une matrice combinant efficacité et toxicité, d'après l'évaluation d'oncologues experts (sur une échelle de 0 à 100), de faible efficacité et haute toxicité à haute efficacité et faible toxicité. Les experts déterminent aussi la région des scores d'utilité considérés comme acceptables. L'efficacité et la toxicité est évaluée à un instant précis pour chaque patient. Définies par des études de simulation, les lois a priori, combinées aux évaluations, forment les lois a posteriori à l'aide de la distribution conjointe. Les probabilités conjointes résultantes sont estimées et celles qui se retrouvent dans la «région acceptable» sont utilisées pour calculer l'utilité moyenne de chaque dose. La dose optimale est la dose qui présente le score d'utilité le plus élevé. La randomisation est proportionnelle aux probabilités obtenues. Les valeurs des doses appropriées sont utilisées pour la phase III.

3D-I2: Business Analytics: Where does the Statistician Fit? Méthodes analytiques de gestion : le rôle du statisticien

Organizer and Chair / Responsable et président: Georges Monette (York University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: Hart House - East Common

Abstracts/Résumés

KRZYSZTOF DZIECIOLOWSKI (Rogers)

[Wednesday May 28 / mercredi 28 mai, 15:00]

From Small Samples to Big Data

Partir de petits échantillons de données volumineuses

In the competitive business world, smart decisions are critical for success. Statistical methods have traditionally played an important role in helping companies achieving their goals. However in the modern, connected society and an era of Big Data, emerging disciplines of Analytics and Data Science have become principal references when searching for business insights and solutions. Shortage of qualified candidates for many open analytical positions reflects another challenge facing statistical education. I argue that in order to address key questions of modern business, statisticians need to recognize an ongoing transformation of interest and opportunities towards Big Data applications. I share learnings from the novel Rogers' Data Science Program and newly re-designed Data Mining curriculum in the John Molson School of Business.

Dans le monde concurrentiel des affaires, prendre rapidement des décisions intelligentes est essentiel au succès. Dans la société moderne et à l'heure des données volumineuses, plusieurs disciplines telles que science des données sont devenues les domaines de références pour la recherche d'opportunités et de perspectives d'affaires. Pour de nombreux postes en analytique, le manque de candidats qualifiés représente un défi pour l'enseignement statistique. Ainsi, afin de répondre aux questions et besoins cruciaux du monde des affaires, il faut que les statisticiens reconnaissent l'intérêt et les opportunités des données volumineuses. Cette tendance et ces changements se reflètent dans le nouveau programme interne de formation de Rogers ainsi que dans le nouveau programme d'enseignement en exploration de données de l'école de gestion John-Molson.

KAISER FUNG (Vimeo)

[Wednesday May 28 / mercredi 28 mai, 15:30]

The Nature of Big Data and the Importance of Numbersense

Nature des données volumineuses et importance du sens des nombres

Two emerging trends are raising the status of business statisticians: Big Data and "Moneyball" culture.

In Michael Lewis's Moneyball, a young general manager used data to outwit old-school baseball scouts who rely on gut feelings. In my experience, good data analysts let intuition guide them through messy data. Such instincts constitute

Deux nouvelles tendances rehaussent le prestige des statisticiens d'entreprise : les données volumineuses et la culture analytique à la Moneyball.

Dans l'oeuvre Moneyball de Michael Lewis, un jeune directeur général a utilisé des données pour déjouer les dépisteurs de baseball de la vieille école qui se fient à leur instinct. Selon mon expérience, les analystes de données efficaces se laissent guider par leur intuition à travers des données

3D-I2: Business Analytics: Where does the Statistician Fit? Méthodes analytiques de gestion : le rôle du statisticien

'numbersense', representing the third pillar of analytical success, in addition to technical and business knowledge.

In the Big Data age, more people are peddling more theories based on more data analyses. The

average citizen must apply numbersense to judge whether a published analysis is trustworthy or not. I will discuss the nature of Big Data, and suggest areas in which statisticians can push the envelope.

JEFFREY MCGILL (Queen's University)

[Wednesday May 28 / mercredi 28 mai, 16:00]

Business Analytics, Statistics, and Communication
Analytique d'affaires, statistique et communication

Modern statistical methods are at the center of the rapidly growing area of business analytics. There are many interesting technical problems, but the key challenges are in alignment of methods with business issues; integration of data collection, screening and analysis; and clear communication with decision-makers.

désordonnées. De tels flairs constituent «le sens des nombres», qui constitue le troisième pilier du succès analytique, en plus des connaissances techniques et de l'expertise des

affaires.

À l'ère des données volumineuses, plus de gens colportent un plus grand nombre de théories basées sur des analyses de données toujours plus nombreuses. Le commun des mor-

tels doit faire appel à son sens des nombres pour évaluer si une analyse publiée est digne de confiance.

Dans cet exposé, je discute de la nature des données volumineuses et je propose des domaines desquels les statisticiens peuvent repousser les limites.

Les méthodes statistiques modernes sont au coeur de l'analytique d'affaires, un domaine en plein essor. En plus d'un nombre de problèmes techniques intéressants, les principaux défis résident dans l'alignement des méthodes avec les questions d'affaires, notamment l'intégration de la collecte de données, le filtrage et l'analyse, ainsi qu'une communication claire avec les décideurs.

3D-I3: Evaluating the Effectiveness of our Introductory Courses Comment évaluer l'efficacité de nos cours d'introduction?

Organizer and Chair / Responsable et présidente: Bethany White (Western University)

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: UC - 163

Abstracts/Résumés

BOB DELMAS (University of Minnesota)

[Wednesday May 28 / mercredi 28 mai, 15:00]

Trends in Statistical Understanding: Results From CAOS

Tendances en matière de compréhension statistique : résultats du CAOS

The ARTIST project has been collecting data since 2005 on students' understanding of statistics through the administration of the Comprehensive Assessment of Outcomes in Statistics (CAOS) instrument. CAOS is a 40-item multiple-choice test that covers six topics: data collection and design, graphical representations, variability, sampling variability, tests of significance, and bivariate data. Over 30,000 secondary and tertiary level students in the United States who were enrolled in a college-level first course in statistics completed the CAOS test. The presentation will present results from a confirmatory factor analysis (CFA), explore trends in students' understanding of the six statistical topics across the 8-year period, and discuss implications for program evaluation and research.

Le projet ARTIST recueille des données depuis 2005 sur la compréhension des étudiants en matière de statistique grâce à l'instrument d'évaluation détaillée des résultats en statistique (CAOS). Cet instrument est un test de 40 questions à choix multiples qui évalue six sujets : la collecte de données et le plan d'expérience, les représentations graphiques, la dispersion, la variabilité échantillonnale, les tests statistiques et les données bivariées. Plus de 30 000 étudiants de niveau secondaire et tertiaire aux États-Unis qui se sont inscrits à un premier cours d'initiation à la statistique universitaire ont subi le test CAOS. La présentation décrit les résultats d'une analyse factorielle confirmatoire, explore les tendances des étudiants en matière de compréhension des six sujets statistiques au cours d'une période de huit ans, et présente les répercussions sur l'évaluation des programmes et sur la recherche.

BRUCE DUNHAM & GAITRI YAPA (University of British Columbia)

[Wednesday May 28 / mercredi 28 mai, 15:30]

Calibrating the Difficulty Level of Assessment Tools: "Blooming" a Statistics Exam

Étalonnage du niveau de difficulté des outils d'évaluation en appliquant la taxonomie de Bloom à un examen de statistique

When changes are made to the pedagogy used within a Statistics course, it invariably has impact on the style of assessment. For example, examination questions may move toward probing conceptual understanding and knowledge transfer. This can make it difficult to evaluate the effectiveness of the changes in pedagogy. An approach is described here that attempts to calibrate the difficulty

Lorsque des changements sont apportés à la pédagogie employée dans un cours de statistique, il en résulte invariablement un impact sur le style d'examen. Par exemple, les questions d'examen peuvent être formulées pour mesurer la compréhension conceptuelle et le transfert des connaissances. L'approche décrite ici est conçue pour étalonner le niveau de difficulté d'un examen et elle est axée sur la comparaison des questions avec les catégories de la tax-

3D-I3: Evaluating the Effectiveness of our Introductory Courses Comment évaluer l'efficacité de nos cours d'introduction?

of an examination, based on comparing the questions with levels on Bloom's taxonomy. A proposed instrument for "Blooming" Statistics examinations is discussed, and illustrated via examples of examination questions both before and after course transformations. Results are presented regarding the apparent impact of course changes on student learning.

onomie de Bloom. L'exposé propose un instrument pour appliquer la taxonomie de Bloom à des examens de statistique et présente des exemples comparés de questions d'examen avant et après avoir apporté des transformations au cours. Les résultats présentent l'impact apparent des changements de cours sur l'apprentissage de l'étudiant.

MICHELE MILLAR (Mount Saint Vincent University), **BETHANY WHITE** (Western University), **CANDACE SCHAU** (CS Consultants)

[Wednesday May 28 / mercredi 28 mai, 16:00]

Do Our Courses Improve Students' Attitudes?

Nos cours améliorent-ils l'attitude des étudiants?

We typically focus on student academic results and course evaluations to assess the success of our courses. Another important outcome to consider is students' attitudes towards statistics. Students with positive attitudes are more likely to engage in behaviors such as pursuing further studies in statistics, using statistics in their future activities and consulting statistical experts, when necessary. The Survey of Attitudes Toward Statistics (SATS), developed by Dr. Candace Schau, assesses six attitude components: Affect, Cognitive Competence, Difficulty, Value, Interest and Effort. In this talk, I will describe the SATS and share results of our assessment of the improvement in attitudes over a two semester introductory course based on four surveys administered at the beginning and end of each semester.

Pour évaluer la réussite de nos cours, nous nous concentrons généralement sur les résultats obtenus et sur les évaluations de cours. Un autre facteur important à considérer est l'attitude des étudiants à l'égard de la statistique. En effet, les étudiants dont l'attitude à cet égard est positive sont plus susceptibles d'exercer certaines activités qui ont trait à la statistique, notamment poursuivre des études avancées dans ce domaine, utiliser la statistique dans leurs futures activités, et consulter au besoin des spécialistes en statistique. L'étude des comportements à l'égard de la statistique (SATS) élaborée par Dr Candace Schau, évalue six composantes du comportement : l'affect, l'intellect, la compétence, la difficulté, la valeur, l'intérêt et l'effort. Dans le présent exposé, je décris le SATS et je partage les résultats de notre évaluation de l'amélioration en matière d'attitude lors d'un cours d'initiation de deux semestres basé sur quatre sondages réalisés au début et à la fin de chaque semestre.

3D-I4: Levy Processes and Applications Processus de Lévy et leurs applications

Organizer and Chair / Responsable et président: Xiaowen Zhou (Concordia University)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Room/Salle: UC - 179

Abstracts/Résumés

ALEXEY KUZNETSOV & DANIEL HACKMANN (York University)

[Wednesday May 28 / mercredi 28 mai, 15:00]

Numerical methods for Levy processes

Méthodes numériques pour des processus de Lévy

Levy processes are very popular building blocks used in various models in Physics, Biology, Mathematical Finance and Actuarial Mathematics. While the theory of Levy processes is very well-developed, there is still a general lack of good numerical methods which would allow to apply this theory in practice. In this talk I will describe our recent work on approximating an important family of Levy processes (the ones which have completely monotone density of jumps) by a class of simpler processes (with hyper-exponential jumps), for which the numerical computations can be done very easily and efficiently. I will discuss some applications of these results to Mathematical Finance and will provide numerical evidence which illustrates the simplicity and efficiency of our approach.

L'usage des processus de Lévy est très répandu dans divers modèles de physique, de biologie et de mathématiques financières et actuarielles. Même si la théorie des processus de Lévy est bien développée, les méthodes numériques permettant de la mettre en pratique sont encore rares. Dans cet exposé, il sera question de l'approximation d'une importante famille de processus de Lévy (ceux qui présentent une densité de sauts complètement monotone) par une classe de processus plus simples (dont les sauts sont hyper-exponentiels) pour lesquels les calculs numériques peuvent être obtenus très facilement et efficacement. Je présenterai quelques applications de ces résultats aux mathématiques financières, ainsi qu'une démonstration numérique illustrant la simplicité et l'efficacité de la méthode proposée.

ANDREAS KYPRIANOU (University of Bath, Great Britain), **JUAN CARLOS PARDO** (CIMAT, Mexico), **ALEX WATSON** (University of Zurich)

[Wednesday May 28 / mercredi 28 mai, 15:30]

Censored Stable Processes

Processus censurés stables

We look at a general two-sided jumping strictly α -stable process where $\alpha \in (0, 2)$. By censoring its path each time it enters the negative half line we show that the resulting process is a positive self-similar Markov Process. Using Lamperti's transformation we uncover an underlying driving Lévy process and, moreover, we are able to describe in surprisingly explicit detail the Wiener-Hopf factor-

Nous examinons un processus bilatéral général de sauts strictement α -stable, où $\alpha \in (0, 2)$. En censurant sa trajectoire chaque fois qu'il pénètre dans la moitié négative, nous obtenons un processus de Markov auto-similaire positif. À l'aide de la transformation de Lamperti, nous exposons un processus de Lévy dynamique sous-jacent, duquel nous présentons la factorisation Wiener-Hopf de façon étonnamment explicite. En combinant cette factorisation Wiener-Hopf

3D-I4: Levy Processes and Applications Processus de Lévy et leurs applications

ization of the latter. Using this Wiener-Hopf factorization together with a series of spatial path transformations, it is now possible to produce an explicit formula for the law of the original stable processes as it first *enters* a finite interval, thereby generalizing a result of Blumenthal, Gettoor and Ray for symmetric stable processes from 1961.

avec une série de transformations à trajectoire spatiale, nous obtenons une formule explicite décrivant la loi des processus stables originaux alors qu'elle *entre*, pour la première fois, dans un intervalle fini. La méthode proposée nous permet de généraliser un résultat de Blumenthal, Gettoor et Ray sur les processus stables symétriques datant de 1961.

JEAN-FRANCOIS RENAUD (Université du Québec à Montréal), **XIAOWEN ZHOU** (Concordia University), **RONNIE LOEFFEN** (University of Manchester), **HÉLÈNE GUÉRIN** (Université de Rennes 1)

[Wednesday May 28 / mercredi 28 mai, 16:00]

Occupation Times for Spectrally Negative Lévy Processes with Applications

Temps d'occupation pour des processus de Lévy spectralement négatifs et applications

We will present various results on occupation times for spectrally negative Lévy processes, having in mind applications to option pricing and the analysis of actuarial ruin models. The results are expressed in terms of the q -scale functions of the process, using recently obtained analytical identities for this family of functions.

Nous présenterons quelques résultats sur les temps d'occupation pour des processus de Lévy spectralement négatifs, l'objectif étant d'utiliser ceux-ci pour la tarification d'options et l'analyse de modèles de théorie de la ruine. Les résultats seront énoncés en utilisant les q -fonctions d'échelle du processus, grâce à de récentes identités analytiques.

3D-I5: Predictive Density Estimation Estimation de densité prédictive

Organizer and Chair / Responsable et président: Eric Marchand (Université de Sherbrooke)

Room/Salle: UC - 161

Abstracts/Résumés

FUMIYASU KOMAKI (University of Tokyo)

[Wednesday May 28 / mercredi 28 mai, 15:00]

Prior Construction Based on the Predictive Metric

Construction de lois a priori basées sur des mesures prédictives

Bayesian predictive densities when the observed data x and the target variable y to be predicted have different distributions are investigated. The performance of predictive densities is evaluated by the Kullback-Leibler divergence. It is shown that the volume element prior based a Riemannian metric, which we call the predictive metric, corresponds to the Jeffreys prior in the conventional setting in which x and y have the identical distribution. As an example, simultaneous predictive densities for independent Poisson processes with different durations are investigated

Nous étudions la densité de lois prédictives bayésiennes, où une donnée observée x et une variable cible y à prédire ont des distributions différentes. La performance des densités prédictives est évaluée par la divergence de Kullback-Leibler. L'étude démontre que la loi a priori basée sur l'élément de volume d'une mesure riemannienne, appelée mesure prévisionnelle, correspond à la loi a priori de Jeffreys dans un contexte traditionnel où x et y ont une distribution identique. À titre d'exemple, nous examinons des densités prévisionnelles simultanées pour des processus de Poisson indépendants de durée différente.

WILLIAM STRAWDERMAN (Rutgers University), **TATSUYA KUBOKAWA** (University of Tokyo), **ERIC MARCHAND** (Université de Sherbrooke)

[Wednesday May 28 / mercredi 28 mai, 15:30]

Predictive Density Estimation Under Integrated L1 and L2 Loss

Estimation de densité prédictive sous des pertes intégrées L1 et L2

We consider predictive density estimation under integrated L1 and L2 loss for p -dimensional location parameter problems. Among the topics considered are the form of Bayes estimators, and minimum risk equivariant estimators (MRE's) for integrated L2 loss. In particular, Bayes estimators for integrated squared error loss agree with those for Kullback-Leibler loss as do MRE's. Comparison of risks for plug-in estimators is facilitated in certain problems (particularly normal ones) by exhibiting a duality with a corresponding point estimation problem wherein the loss is a concave function of squared error loss. Using this duality we find improvements over the MRE in high dimensions. Similar duality results are developed for integrated L1 loss in the

Nous examinons l'estimation d'une densité prédictive pour les pertes intégrées L1 et L2 pour des paramètres de localisation p -dimensionnels. Parmi les sujets abordés, soulignons la forme des estimateurs bayésiens et les estimateurs équivariants à risque minimal (MRE) pour la perte intégrée L2. Signalons également les estimateurs bayésiens dont la fonction de perte est l'erreur quadratique intégrée, ceux avec une fonction de perte de Kullback-Leibler, ainsi que les MRE. Dans certains problèmes (particulièrement les problèmes normaux), la comparaison des risques en remplaçant les paramètres par leur estimateur est facilitée grâce à une dualité avec un problème correspondant d'estimation ponctuelle, où la perte est une fonction concave des pertes au carré. Grâce à cette dualité, nous constatons des améliorations des MRE de grande dimension. Des résultats simi-

3D-I5: Predictive Density Estimation Estimation de densité prédictive

spherically symmetric case.

laire de dualité sont atteints pour la perte intégrée L1 dans un cas de symétrie sphérique.

XINYIN XU (Ohio State University), **EDWARD GEORGE** (University of Pennsylvania), **FENG LIANG** (University of Illinois at Urbana-Champaign)

[Wednesday May 28 / mercredi 28 mai, 16:00]

From Minimax Shrinkage Estimation to Minimax Shrinkage Prediction

De l'estimation minimax par rétrécissement à la prévision minimax par retrécissement

In a remarkable series of papers beginning in 1956, Charles Stein set the stage for the future development of minimax shrinkage estimators of a multivariate normal mean under quadratic loss. More recently, parallel developments have seen the emergence of minimax shrinkage estimators of multivariate normal predictive densities under Kullback-Leibler risk. In this talk, I describe these parallels emphasizing the focus on Bayes procedures and the derivation of the superharmonic conditions for minimaxity as well as further developments of new minimax shrinkage predictive density estimators including multiple shrinkage estimators, empirical Bayes estimators, normal linear model regression estimators and nonparametric regression estimators.

Dans une série remarquable d'articles amorcée en 1956, Charles Stein a posé les jalons de l'élaboration future des estimateurs de rétrécissement d'une moyenne normale multivariée sous perte quadratique. Plus récemment, des développements parallèles ont vu l'émergence d'estimateurs minimax de rétrécissement de densités prédictives normales multivariées sous le risque Kullback-Leibler. Dans cet exposé, l'auteur décrit ces parallèles en mettant l'accent sur les procédures bayésiennes et la dérivation des conditions superharmoniques pour une minimaxité, ainsi que sur des développements de nouveaux estimateurs de densité prédictive de rétrécissement, y compris des estimateurs de rétrécissement multiples, des estimateurs bayésiens, des estimateurs de modèle de régression linéaires normaux, ainsi que des estimateurs de régression non paramétrique.

3D-16: Statistical Analysis for Infectious Diseases Analyse statistique des maladies infectieuses

Organizer and Chair / Responsable et présidente: Juxin Liu (University of Saskatchewan)

Room/Salle: Hart House - Music Rm

Abstracts/Résumés

ROB DEARDON, SCOTT HUNT, NADIA BIFOLCHI & LONGYAO CAI (University of Guelph)

[Wednesday May 28 / mercredi 28 mai, 15:00]

Optimal Experimental and Study Design for Infectious Disease Systems of Animals

Plan d'étude et d'expérience optimal pour des systèmes de pathologies infectieuses animales

Design for non-linear models is complicated by the fact that the optimal design depends upon the parameters. A Bayesian, and often simulation-based, framework is often a natural setting for such design problems. Of interest here are optimal design problems for both animal experiments, and observational studies, where the underlying goal is to identify some characteristics of the disease dynamics (e.g. a vaccine effect, or the infectious period). Generic results are hard to formulate for such systems, being highly dependent upon the system being observed, the models to be fitted, the characteristic of that system to be identified, and the parameters. This means that the application under study is key.

La définition d'un modèle non linéaire se complique en raison du fait qu'une conception optimale dépend des paramètres. Un contexte bayésien, souvent basé sur une simulation, constitue souvent le contexte pour de tels problèmes de configuration. Nous examinons ici des problèmes de conception optimale, à la fois pour des expériences animales et des études d'observation, où l'objectif sous-jacent consiste à identifier des caractéristiques de la dynamique pathologique (i.e. l'effet d'un vaccin ou de la période infectieuse). Les résultats génériques sont difficiles à formuler pour de tels systèmes, étant donné qu'ils dépendent fortement du système à l'étude, des modèles à ajuster et de la caractéristique du système à identifier, ainsi que des paramètres. L'application à l'étude est donc un élément clé.

NATHANIEL OSGOOD, JUXIN LIU & SETH DUECK (University of Saskatchewan), **DAVID VICKERS** (Alberta Health Services)

[Wednesday May 28 / mercredi 28 mai, 15:30]

Combining MCMC and Compartmental Modeling to Enhance Understanding of Chlamydia Control in Saskatchewan

Combinaison de MCMC et de modélisation compartimentale en vue d'améliorer la compréhension du contrôle de la chlamydia en Saskatchewan

Like many regions of the industrial world, Saskatchewan experienced falling notifications of intracellular bacterial infection *Chlamydia trachomatis* throughout the 1990s, followed by rapidly rising notifications. Spurred by historically unprecedented reported case counts hailed as a public health crisis, researchers have advanced a wide variety of explanations for the rising notification rates. To investigate drivers for this trend in light of reporting variability, we combined a previously published system dynamics simulation model with MCMC methods.

Comme pour bon nombre de régions du monde industriel, la Saskatchewan a connu une chute du taux de signalement d'infections bactériennes intracellulaires pendant les années 1990, avant d'enregistrer par la suite une hausse rapide des cas déclarés. Un nombre sans précédent de cas déclarés, reconnu comme une crise en santé publique, a incité des chercheurs à élaborer une grande diversité d'hypothèses pour expliquer les taux de signalement en hausse. Pour investiguer les facteurs favorisant cette tendance à la lumière de la variabilité du signalement, nous avons combiné à des méthodes MCMC un modèle de simulation de dynamique

3D-16: Statistical Analysis for Infectious Diseases Analyse statistique des maladies infectieuses

We leveraged province-wide test volumes and outcomes data via a previously contributed Chlamydia transmission model and a probabilistic model for the test data conditional on the solution of the transmission model. We discuss here convergence challenges encountered and insights secured.

de système déjà publié. Nous avons misé sur des données de tests de volumes et d'effets provenant de l'ensemble de la province et tirées d'un modèle déjà publié sur la transmission de la chlamydia et un modèle probabiliste pour les données du test conditionnellement à la solution du modèle de transmission. Nous discutons ici les problèmes de convergence rencontrés et présentons nos conclusions à la lumière des résultats.

ERIN SCHLIEP (Duke University), **DANIEL GREAR** (United States Department of Agriculture), **COLLEEN WEBB** (Colorado State University)

[Wednesday May 28 / mercredi 28 mai, 16:00]

Modeling Prevalence of Bovine Tuberculosis Using Multiple-Phase Testing

Modélisation de la prévalence de tuberculose bovine à l'aide de tests à phases multiples

Bovine tuberculosis (bTB) is an ongoing management problem for livestock and wildlife in Michigan. Transmission can exist for long periods of time before obvious symptoms are apparent. Therefore, we must rely on imperfect diagnostic tests to detect bTB at early stages to prevent transmission. We model herd-level prevalence of bovine tuberculosis within a Bayesian framework incorporating uncertainty in test sensitivity and specificity. The model allows for multiple-phase testing where animals are administered a series of tests in the absence of a gold standard. We estimate herd-level prevalence with uncertainty and compare different multiple-phase testing scenarios that are currently being administered in Michigan.

La tuberculose bovine (bTB) est un problème continu de gestion du bétail et d'espèces sauvages au Michigan. La transmission peut se produire sur de longues périodes de temps avant qu'apparaissent des symptômes évidents. Par conséquent, nous devons nous fier à des tests diagnostiques imparfaits pour détecter bTB à des stades précoces afin d'en empêcher la transmission. Nous modélisons la prévalence de tuberculose bovine au niveau du troupeau dans un cadre bayésien, en intégrant l'incertitude à la sensibilité et à la spécificité du test. Le modèle prévoit des phases de test multiples, où les animaux subissent une série de tests en l'absence d'un étalon de référence. Nous évaluons la prévalence au niveau du troupeau et son incertitude, et nous comparons des scénarios de tests à phases multiples différents qui ont cours actuellement au Michigan.

3D-C1: Actuarial Science and Finance 4 Actuariat et finance 4

Chair/Président: Adam Metzler (Wilfrid Laurier University)

Room/Salle: UC - A101

Abstracts/Résumés

RUODU WANG (University of Waterloo), **VALERIA BIGNOZZI** (ETH Zurich), **ANDREAS TSANAKAS** (City University London)

[Wednesday May 28 / mercredi 28 mai, 15:00]

How Superadditive Can a Risk Measure Be?

À quel point une mesure de risque peut-elle être suradditive?

There have been extensive debates on the comparison of risk measures in both academia and industry. In this talk, we address the fundamental question: "how superadditive can a risk measure be?" We introduce the extreme-scenario measures induced by risk measures to quantify the worst-case aggregate capital requirement in the presence of model uncertainty. Our results suggest that Expected Shortfall and the property of coherence indeed have some unique features never discovered in the literature and desired in a good risk measure. Our results are particularly relevant to the recent consultative documents by the Basel Committee on Banking Supervision.

Il y a eu de nombreux débats sur la comparaison des mesures du risque dans le monde universitaire et l'industrie. Dans cette présentation, nous abordons la question fondamentale de savoir « à quel point une mesure de risque peut être suradditive ». Nous introduisons les mesures de scénarios extrêmes induites par les mesures de risque pour quantifier le montant des capitaux exigés dans le pire scénario en la présence d'incertitude dans le modèle. Nos résultats suggèrent que le déficit prévu et la propriété de cohérence présentent effectivement des caractéristiques uniques, jamais découvertes dans la littérature, requises pour une bonne mesure du risque. Les résultats obtenus sont particulièrement pertinents dans le cadre des derniers documents de consultation publiés par le comité de Bâle sur le contrôle bancaire.

SUNGSOO KIM & STEVE DREKIC (University of Waterloo)

[Wednesday May 28 / mercredi 28 mai, 15:15]

Ruin Analysis in a Discrete-time Sparre Andersen Model with External Financial Activities and Random Dividends

Analyse de ruine dans un modèle à temps discret de Sparre Andersen comprenant des activités financières et des dividendes aléatoires

In this work, we consider a risk model which incorporates multiple threshold levels characterizing an insurer's minimal capital requirement, dividend paying situations, and external financial activities. Our model is based on discrete monetary and time units, and the main quantities of interest are the finite-time ruin probabilities and the expected total discounted dividends paid prior to ruin. We mainly focus on the development of computational methods to attain these quantities of interest. Note that we consider only non-stochastic interest rates for

Dans cette étude, nous analysons un modèle de risque qui intègre des seuils multiples caractérisant les exigences de capital minimum requis d'un assureur, les cas de paiements de dividendes et les activités financières externes. Notre modèle est fondé sur des unités monétaires et temporelles discrètes, et les principales quantités à l'étude sont les probabilités de ruine à temps fini et la valeur actualisée des dividendes totaux attendus préalablement à la ruine. Nous concentrons l'analyse sur l'élaboration de méthodes de calcul permettant d'obtenir ces quantités à l'étude. Il est à noter qu'à des fins d'économie de temps de calcul, nous étudions

3D-C1: Actuarial Science and Finance 4

Actuariat et finance 4

both lending and borrowing activities for computational time. After constructing appropriate recursive formulae for the quantities of interest, we investigate various numerical examples and make some observations concerning the impact our threshold levels have on the quantities of interest.

seulement les taux d'intérêt non stochastiques tant pour les activités de prêts que d'emprunt. Après avoir établi les formules récursives appropriées pour les quantités à l'étude, nous analysons divers exemples quantitatifs et nous formulons des observations sur l'impact qu'ont les seuils sur les quantités à l'étude.

JEAN-PHILIPPE BOUCHER & ROFICK INOUSSA (Université du Québec à Montréal)

[Wednesday May 28 / mercredi 28 mai, 15:30]

A Posteriori Ratemaking with Panel Data

Tarifcation a posteriori avec données de panel

Usually, the ratemaking process is done in two steps. In the first step, a priori ratemaking, an a priori premium is computed based on the characteristics of the insureds. In the second step (the a posteriori ratemaking), the past claims experience of each insured is considered to the a priori premium and set the final net premium. In practice, for automobile insurance, this correction is usually done with bonus-malus systems, which offer many advantages. Recently, insurers have accumulated longitudinal information on their policyholders. For this kind of data, we propose an alternative to the two-step ratemaking approach and argue this old approach should no longer be used. Using car insurance data, a numerical illustration is included to support this discussion.

La tarification en assurance non-vie se fait en deux étapes distinctes. Dans la tarification a priori, la prime est calculée sur la base des caractéristiques des assurés. Dans la tarification a posteriori, l'historique de sinistre de chaque assuré est considéré afin de corriger la prime a priori. En pratique, cette correction a posteriori est généralement faite à l'aide des systèmes bonus-malus qui offrent de nombreux avantages. Toutefois, depuis quelques années, les assureurs ont accumulé des données longitudinales sur leurs assurés et nous estimons que la théorie des systèmes bonus-malus se doit ainsi d'être modernisée. Nous affirmons que l'ancienne procédure de tarification ne doit plus être utilisée. À l'aide de données d'assurance automobile, une application numérique sera présentée.

NICOLAS ESSIS-BRETON & PATRICE GAILLARDETZ (Concordia University)

[Wednesday May 28 / mercredi 28 mai, 15:45]

Parallel Computing Methods for Variable Annuities

Méthodes de calcul parallèle pour les rentes variables

A flexible premium variable annuity gives the investor the freedom to contribute each year into a variable annuity, that is an investment product embedded with financial options. In this talk, we consider the valuation of such product by dynamic programming, with an emphasis on the numerical aspect. We discuss the so-called three curses of dimensionality, and how to break them with current technology: GPU, PiCloud and Fortran.

Une rente à prime variable donne à l'investisseur la liberté de contribuer à une rente variable à chaque année, c'est-à-dire un produit de placement doté d'options financières. Dans cette présentation, nous considérons l'évaluation d'un tel produit par la programmation dynamique en mettant l'accent sur l'aspect numérique. Nous discutons des soi-disant trois malédictions de la dimensionnalité ainsi que comment les briser avec la technologie actuelle : GPU, PiCloud et Fortran.

3D-C1: Actuarial Science and Finance 4

Actuariat et finance 4

GENEVIÈVE GAUTHIER, DEBBIE DUPUIS & FRÉDÉRIC GODIN (HEC Montréal)

[Wednesday May 28 / mercredi 28 mai, 16:00]

Short-term Hedging for an Electricity Retailer

Stratégie de couverture pour un revendeur d'électricité

A dynamic global hedging procedure making use of futures contracts is developed for a retailer of the electricity market facing price, load and basis risk. Statistical models reproducing stylized facts are developed for the electricity load, the day-ahead spot price and futures prices in the Nord Pool market. These models serve as input to the hedging algorithm, which also accounts for transaction fees. Backtests with market data from 2007 to 2012 show that the global hedging procedure provides considerable risk reduction when compared to hedging benchmarks found in the literature.

Une stratégie de couverture dynamique basée sur les contrats à terme est développée pour un revendeur d'électricité faisant face à l'incertitude à propos des prix, des quantités et au fait que les contrats à terme ne sont pas parfaitement corrélés aux flux monétaires. Des modèles statistiques reproduisant les faits stylisés des comportements de la demande, des prix spot et des contrats à terme sont développés pour le marché Nord Pool. Ces modèles servent d'intrant à l'algorithme de couverture qui prend aussi en compte les frais de transaction. Des tests hors échantillon démontrent que la stratégie de couverture proposée permet de réduire considérablement le risque lorsque comparée aux procédures de couverture existantes.

DAVID EDWARD WILSON (University of Waterloo)

[Wednesday May 28 / mercredi 28 mai, 16:15]

Volatility and Intensity

Volatilité et intensité

This lecture concerns models with stochastic jump intensity. It has become commonplace in such models to make the Semi-Static Arbitrage assumption in lieu of a martingale. However, it is possible to make a martingale assumption for the discounted stock price, provided that the jump intensity is correlated with it. The model presented improves on the Barndorff-Nielsen and Shephard (2001) model in that its two diffusions are easily correlated. However, only a conditional characteristic for the log price is available in closed form for the new model, and calibration to options prices needs to be done by conditional Quasi-Monte Carlo. Nonetheless, it will be shown that the new model is more stable than Semi-Static Arbitrage.

Cette présentation concerne les modèles avec intensité de saut stochastique. Ces modèles prennent souvent pour hypothèse l'arbitrage semi-statique plutôt que la martingale. Cependant, il est possible de faire une hypothèse de martingale pour le cours réduit des actions, pourvu que l'intensité du saut soit corrélée. Le modèle que nous présentons est une amélioration du modèle de Barndorff-Nielsen et Shephard (2001) car ses deux diffusions sont facilement corrélées. Toutefois, seule une caractéristique conditionnelle du prix en logarithme est disponible en forme close pour le nouveau modèle ; il faut calibrer le prix des options par un Quasi-Monte Carlo conditionnel. Néanmoins, nous montrons que le nouveau modèle est plus stable que l'arbitrage semi-statique.

3D-C2: Biostatistics - Variable Selection Biostatistique - sélection de variables

Chair/Président: Martin Lysy (University of Waterloo)

Room/Salle: Hart House - Debates Rm

Abstracts/Résumés

YEYING ZHU (University of Waterloo), **DONNA COFFMAN & DEBASHIS GHOSH** (Pennsylvania State University)
[Wednesday May 28 / mercredi 28 mai, 15:00]

A Boosting Algorithm for Estimating Generalized Propensity Scores

Algorithme d'augmentation pour l'estimation des scores de propension généralisés

We study causal inferences with a continuous treatment variable using propensity score-based methods. The generalized propensity score is defined as the conditional density of the treatment level given covariates. When the dimension of the covariates is large, the traditional nonparametric density estimation suffers from the curse of dimensionality. A two-step estimation procedure has been suggested by first modeling the mean function. We propose a boosting algorithm to estimate the mean function. A tuning parameter is the number of trees, which determines the trade-off between bias and variance of the causal estimator. We propose a criterion called average absolute correlation coefficient to determine the optimal number of trees. The proposed methodology is illustrated through the Early Dieting in Girls study.

Nous étudions les inférences causales avec une variable de traitement continue à l'aide de méthodes fondées sur le score de propension. Le score de propension généralisé se définit comme la densité conditionnelle du niveau de traitement de covariables données. Lorsque la dimension des covariables est grande, l'estimation de la densité non paramétrique traditionnelle souffre du problème de dimensionnalité. Une procédure d'estimation en deux étapes a été suggérée, qui commencerait par une modélisation de la fonction moyenne. Nous proposons un algorithme d'augmentation pour estimer la fonction moyenne. L'un des paramètres de réglage est le nombre d'arbres, qui détermine l'équilibre entre le biais et la variance de l'estimateur causal. Nous proposons un critère, le coefficient de corrélation absolu moyen, pour déterminer le nombre d'arbres optimal. Nous illustrons la méthode proposée par l'étude Early Dieting in Girls.

HALIMAH ASIRI & JIANAN PENG (Acadia University)

[Wednesday May 28 / mercredi 28 mai, 15:15]

Simultaneous Inference for Low Dose Risk Estimation with Abbott-adjusted Weibull Model in Benchmark Analysis

Inférence simultanée pour l'estimation du risque à faible dose avec modèle de Weibull ajusté selon la formule d'Abbott pour l'analyse d'un banc d'essais

Risk assessment studies where human, animal or ecological data are used to set safe low dose levels of a toxic agent is challenging as study information is limited to high dose levels of the agent. Buckley and Piegorsch (Statistical Methodology, 2008) developed the hyperbolic band for low dose inference for Abbott-adjusted Weibull quantal response models in quantitative risk analysis. However, the shape of the constant width band may be more desirable than that of the hyperbolic band in some

Les études d'évaluation du risque dans lesquelles des données sur les humains, les animaux ou l'écologie sont utilisées pour établir des niveaux sécuritaires de faibles doses d'un agent toxique présentent un défi réel étant donné que l'information des études est limitée à des niveaux de doses élevées de l'agent toxique. Buckley et Piegorsch (Méthodologie statistique, 2008) ont développé une bande hyperbolique pour l'inférence à faible dose pour des modèles quantitatifs de Weibull ajustés selon la formule d'Abbott pour l'analyse de risque d'une réponse quantale. Cepen-

3D-C2: Biostatistics - Variable Selection Biostatistique - sélection de variables

applications in risk analysis as a confidence band should be narrower especially near the lower end of low doses. Today, we will show that a constant width band should do better than the hyperbolic band considered by Buckley and Piegorsch.

dant, une forme de bande à largeur constante semble plus appropriée que la bande hyperbolique dans certaines applications d'analyse du risque, étant donné que la bande de confiance devrait être plus étroite, particulièrement pour les plus faibles doses. Dans cet exposé, nous montrons qu'une bande à largeur constante devrait être plus efficace que la bande hyperbolique proposée par Buckley et Piegorsch.

YANG NING & GRACE YI (University of Waterloo), **BAOJIANG CHEN** (University of Nebraska Medical Center), **NANCY REID** (University of Toronto)

[Wednesday May 28 / mercredi 28 mai, 15:30]

Variable Selection and Inference for Ultra-High Dimensional Survival Data With Missing Covariates Under Proportional Hazards Models

Sélection de variable et inférence pour des données de survie en très grande dimension avec des covariables manquantes dans des modèles à risques proportionnels

Proportional hazards models have been perhaps the most popular models used for survival data analysis. Such models, however, break down for settings with high dimensional covariates, which are subject to missingness. In this paper, we address this important problem and develop simultaneous inferential procedures that handle both model selection and parameter estimation for ultra-high dimensional survival data with missing covariates. Our methods are developed for a broad class of folded concave penalties, including the LASSO and SCAD penalties to conduct variable selection. The strong oracle property and the asymptotic distributions of our proposed estimators are rigorously established. The performance of the proposed methods is numerically assessed through simulation studies and analysis of a genetic data set.

Les modèles à risques proportionnels ont sans doute été les modèles les plus utilisés pour l'analyse des données de survie. Par contre, ces modèles échouent pour des paramètres comportant des covariables à grande dimension qui sont susceptibles de faire l'objet de données manquantes. Dans cet article, nous discutons ce problème important et nous élaborons des processus d'inférence simultanée adaptés à la fois à la sélection de modèle et à l'estimation des paramètres pour des données de survie en très grande dimension qui comportent des covariables manquantes. Nous avons élaboré nos méthodes sur une classe étendue de pénalités concaves repliées, y compris les pénalités LASSO et SCAD pour la sélection de variable. Une forte propriété oracle et des distributions asymptotiques des estimateurs proposés ont été rigoureusement établies. La performance des méthodes proposées fait l'objet d'une évaluation numérique à l'aide de simulations et d'une analyse d'un jeu de données génétiques.

SAADIA PINKY & GISEON HEO (University of Alberta)

[Wednesday May 28 / mercredi 28 mai, 15:45]

Hypothesis Testing with Power Calculation and Variable Selection with Sparse Dirichlet-Multinomial in Clostridium Difficile data

Tests d'hypothèse avec calcul de puissance et sélection de variable pour une régression multinomiale de Dirichlet parcimonieuse avec des données portant sur le Clostridium difficile

New developments in sequencing technology have enabled researchers to study microbiome composition using direct sequencing; a process that outputs bacterial taxa counts for each microbiome sample. In this work, a parametric approach is applied

De nouveaux développements en technologie séquentielle ont permis à des chercheurs d'étudier la composition du microbiome par séquençage direct; un processus qui produit un dénombrement des taxons bactériens pour chaque échantillon de microbiome. Dans cette étude, une approche

3D-C2: Biostatistics - Variable Selection Biostatistique - sélection de variables

to Clostridium Difficile (CDiff) data on 19 patients for pre Fecal Microbiota Transplantation(FMT) and post FMT samples. In particular, several tests of hypothesis and power/sample size calculations are applied to these samples in the purpose of illustrating the taxonomic abundance distribution and rank abundance distribution of the data. The Dirichlet-multinomial distribution allows the analyst to perform these in order to describe microbiome properties. We also propose to model the taxa counts using a Dirichlet-multinomial regression model to account for overdispersion of observed counts.

paramétrique est appliquée à des données sur le Clostridium difficile sur 19 patients devant subir une bactériothérapie fécale. Des échantillons ont été prélevés avant et après la greffe. Plusieurs tests d'hypothèse ainsi que des calculs de puissance et de taille d'échantillon ont été appliqués à ces échantillons afin d'illustrer les distributions d'abondance taxonomique et de rang. La distribution multinomiale de Dirichlet permet à l'analyste de produire ces distributions afin de décrire les propriétés du microbiome. Nous proposons aussi de modéliser les quantités de taxons en utilisant un modèle de régression multinomiale de Dirichlet pour tenir compte de la dispersion excédentaire des quantités observées.

FARHAD SHOKOOHI, ABBAS KHALILI & MASOUD ASGHARIAN (McGill University)

[Wednesday May 28 / mercredi 28 mai, 16:00]

Variable Selection in Mixture of Survival Models

Sélection de variables dans les modèles de mélange de survie

Variable selection in high dimensional spaces has been extensively studied in different settings over the past decades. In some applications, finite mixture of regression (FMR) models are often used to capture the heterogeneity in the population. Variable selection in FMR models has been recently studied. However, such methodologies are not applicable to follow-up studies on time-to-event data when data is incomplete. We have considered variable selection in FMR models when observations are subject to right censoring. We have proposed a penalized likelihood method for this problem. Large sample properties of the proposed method are studied. Simulation studies are carried out to evaluate the performance of the proposed method. A real data is analyzed for illustrative purposes.

La sélection de variables dans les espaces de grandes dimensions a fait l'objet d'études dans de nombreux contextes ces dernières années. Dans certaines applications, des modèles de mélange de régression finis (MRF) sont utilisés pour mesurer l'hétérogénéité de la population. La sélection de variables dans les modèles MRF a également fait l'objet de diverses études. Toutefois, ces méthodes ne sont pas applicables aux études de suivi pour les données de durée de vie, surtout lorsque les données sont incomplètes. Nous étudions la sélection de variables dans les modèles MRF lorsque les observations sont sujettes à une censure à droite. Nous proposons une méthode de vraisemblance pénalisée pour ce problème. Nous étudions les propriétés de cette méthode pour de grands échantillons. Des études par simulation sont effectuées pour évaluer la performance de la méthode proposée. Des données réelles sont analysées à des fins d'illustration.

Author List • Liste des auteurs

- Abarin, Taraneh, 47, 186 (2C-P2)
 Abdallah, Anas, 49, 199 (2D-I3)
 Abdelrazeq, Ibrahim, 25, 90 (1B-C2)
 Abrahamowicz, Michal, 32 (1D-C2), 66 (3D-I1), 114 (1D-C2), 280 (3D-I1)
 Acar, Elif, 64, 272 (3B-C3)
 Adamic, Peter, 63, 266 (3B-C1)
 Addario-Berry, Louigi, 31, 105 (1D-I5)
 Agard, Bruno, 46, 184 (2C-P2)
 Aghababazadeh, Farnoosh Abbas, 32, 112 (1D-C2)
 Aghdam, Rosa, 36, 137 (1E-C1)
 Ahmed, Ejaz, 38, 147 (1E-C4)
 Ait Aoudia, Djilali, 58, 246 (2E-C4)
 Ajavon, Ayi, 33, 121 (1D-C4)
 Alam, Shomoita, 43, 166 (2B-C1)
 Albert-Green, Alisha, 39, 150 (1E-C5)
 Allab, Imène, 58, 248 (2E-C4)
 ALmutairi, Malak, 26, 92 (1B-C3)
 Altman, Naomi, 37, 144 (1E-C3)
 Alvo, Mayer, 32, 112 (1D-C2)
 An, Hongyu, 50, 202 (2D-I4)
 Anand, Kanika, 52, 216 (2D-C3)
 Anastasia, Teterina, 36, 137 (1E-C1)
 Anderson, Marti, 26, 92 (1B-C3)
 Angers, Jean-François, 57, 243 (2E-C3)
 Angevaare, Justin, 46, 181 (2C-P2)
 Ansari, Ferhana, 22, 74 (1B-I1)
 Asgharian, Masoud, 29 (1D-I1), 44 (2B-C4), 69 (3D-C2), 176 (2B-C4), 297 (3D-C2)
 Ashta, Di Wang, Namdar Homayounfar, Peter Lee, Serena Lu, Shiva, 28 (1C-D2)
 Ashta, Shiva, 43, 170 (2B-C2)
 Asiri, Halimah, 69, 295 (3D-C2)
 Atherton, Juli, 64, 270 (3B-C2)
 Augusta, Carolyn, 46, 181 (2C-P2)
 Austin, Peter, 41, 158 (2B-I3)
 Azadbakhsh, Mahdis, 43, 171 (2B-C2)
- Babul, Arif, 50, 203 (2D-I4)
 Badescu, Andrei L., 49, 200 (2D-I3)
 Bae, Taehan, 32 (1D-C1), 63 (3B-C1), 110 (1D-C1), 266 (3B-C1)
 Balabdaoui, Fadoua, 42, 162 (2B-I5)
 Balakrishnan, N., 58, 246 (2E-C4)
 Balakrishnan, Narayanaswamy, 33, 116 (1D-C3)
 Balette, Dylan Moore, Dayeol Choi, Christopher, 28 (1C-D1)
 Bang, Heejung, 24, 81 (1B-I5)
 Bappee, Fateha Khanam, 53, 218 (2D-C3)
 Barnes, Michael, 24 (1B-I4), 33 (1D-C3), 80 (1B-I4), 117 (1D-C3)
 Barwick, Melanie, 54 (2E-I1)
 Beaumont, Jean-François, 59, 250 (2E-C5)
 Beg, Mirsa Faisal, 50, 203 (2D-I4)
 Behseta, Sam, 62, 260 (3B-I4)
 Belalia, Mohamed, 26, 93 (1B-C3)
 Béliveau, Audrey, 39, 152 (1E-C5)
 Benedetti, Andrea, 25 (1B-C1), 52 (2D-C2), 87 (1B-C1), 212 (2D-C2)
 Ben Hadj Slimene, Latifa, 58, 246 (2E-C4)
 Ben Salah, Zied, 56, 237 (2E-C1)
 Ben-Salah, Zied, 49, 200 (2D-I3)
 Berke, Olaf, 36, 136 (1E-C1)
 Bernard, Carole, 40, 154 (2B-I1)
 Beyene, Joseph, 23 (1B-I4), 43 (2B-C1), 79 (1B-I4), 167 (2B-C1)
 Bhatnagar, Maxime Turgeon, Kevin Mcgregor, Sahir, 28 (1C-D1)
 Bickel, David, 32, 112 (1D-C2)
 Bifulchi, Nadia, 68, 290 (3D-I6)
 Biggs, Michael, 23, 75 (1B-I2)
 Bignozzi, Valeria, 69, 292 (3D-C1)
 Bingham, Derek, 26 (1B-C2), 49 (2D-I2), 91 (1B-C2), 197 (2D-I2)
 Bodson-Clermont, Paule, 57, 243 (2E-C3)
 Boright, Andrew, 24, 79 (1B-I4)

- Bouchard-Côté, Alexandre, 26, 90 (1B-C2)
 Boucher, Jean-Philippe, 49 (2D-I3), 69 (3D-C1), 199 (2D-I3), 293 (3D-C1)
 Boudreau, Christian, 25, 85 (1B-C1)
 Boudreault, Mathieu, 31, 109 (1D-C1)
 Bouezmarni, Taoufik, 26 (1B-C3), 53 (2D-C4), 58 (2E-C4), 93 (1B-C3), 219 (2D-C4), 247 (2E-C4)
 Bower, Geoffrey, 56, 233 (2E-I5)
 Brault, Jean-Jules, 38, 144 (1E-C3)
 Bray, Mathieu, 61, 255 (3B-I2)
 Bretz, Frank, 54, 227 (2E-I2)
 Briollais, Laurent, 36, 134 (1E-I6)
 Brown, Patrick, 22 (1A-A1)
 Browne, Ryan, 38 (1E-C3), 52 (2D-C3), 58 (2E-C3), 145 (1E-C3), 217 (2D-C3), 244 (2E-C3)
 Browne, Ryan P., 38, 144 (1E-C3)
 Brunzell, John, 24, 79 (1B-I4)
 Bubenik, Peter, 47, 185 (2C-P2)
 Bull, Shelley, 24, 79 (1B-I4)
 Bull, Shelley B., 52, 214 (2D-C2)
 Burkett, Kelly, 60, 253 (3B-I1)
 Burr, W., 47, 187 (2C-P2)
- CADENA-BARAJAS, JOSE MARTIN, 46, 181 (2C-P2)
 Cai, Jun, 49, 199 (2D-I3)
 Cai, Lihu Liu, Moyan Mei, Maria Reyes, Michael Butler, Yun, 28 (1C-D2)
 Cai, Longyao, 68, 290 (3D-I6)
 Cai, Song, 43, 170 (2B-C2)
 Cai, Yanling, 46, 182 (2C-P2)
 Camirand, Felix, 58, 247 (2E-C4)
 Camirand Lemyre, Félix, 53, 219 (2D-C4)
 Campbell, David, 33, 120 (1D-C4)
 Canty, Angelo, 24, 79 (1B-I4)
 Cao, Jiguo, 64, 273 (3B-C3)
 Carriere Chough, Keumhee, 51, 209 (2D-C1)
 Cashore, Chunlin Wang, Peter Raboud, Lin Chen, Allison Devlin, Marco Sung, Nathan Wilson, Massey, 28 (1C-D1)
 Célini, Léonide, 32, 114 (1D-C2)
 Chandrasekaran, Venkat, 22, 72 (1A-A2)
 Chang, Hsing-Ming, 30, 101 (1D-I3)
 Charpentier, Arthur, 24, 83 (1B-I6)
 Chen, Baojiang, 38 (1E-C4), 69 (3D-C2), 147 (1E-C4), 296 (3D-C2)
 Chen, Bingshu, 57, 239 (2E-C2)
 Chen, Fuqi, 58, 249 (2E-C5)
 Chen, Hao, 46, 182 (2C-P2)
 Chen, Jiahua, 23 (1B-I3), 43 (2B-C2), 77 (1B-I3), 170 (2B-C2)
 Chen, Li, 46, 179 (2C-P1)
 Chen, Min, 43, 171 (2B-C2)
 Chen, Ray-Bing, 29, 100 (1D-I2)
 Chen, Senke, 35, 132 (1E-I5)
 Chen, Si, 52, 217 (2D-C3)
 Cheung, Siu Hung, 54, 227 (2E-I2)
 Chhetri, Bimal, 36, 136 (1E-C1)
 Chipman, Hugh, 20, 22 (1B-I1), 38 (1E-C3), 47 (2C-P2), 53 (2D-C3), 73 (1B-I1), 146 (1E-C3), 188 (2C-P2), 217 (2D-C3)
 Choi, Yun-Hee, 36, 134 (1E-I6)
 CHOULAKIAN, Vartan, 53, 222 (2D-C5)
 Chowdhury, Monsur, 29, 99 (1D-I2)
 Coffman, Donna, 69, 295 (3D-C2)
 Cook, Richard, 30 (1D-I4), 37 (1E-C1), 37 (1E-C2), 57 (2E-C2), 61 (3B-I2), 104 (1D-I4), 138 (1E-C1), 141, 142 (1E-C2), 239 (2E-C2), 255 (3B-I2)
 Cossette, Hélène, 31 (1D-C1), 49 (2D-I3), 110 (1D-C1), 199 (2D-I3)
 Côté, Marie-Pier, 31, 109 (1D-C1)
 Cotton, Cecilia, 37 (1E-C2), 64 (3B-C2), 140 (1E-C2), 271 (3B-C2)
 Coulombe, Janie, 57, 243 (2E-C3)
 Cribben, Ivor, 50, 202 (2D-I4)
 Cristina, Tortora, 38, 144 (1E-C3)
 Crommelin, Daan, 56, 236 (2E-I6)
 Csorgo, Miklos, 38 (1E-C4), 42 (2B-I6), 148 (1E-C4), 164 (2B-I6)
 Cunningham, Charles E., 37, 138 (1E-C1)
 Czado, Claudia, 55, 229 (2E-I3)
- Daly, Caitlin, 51, 208 (2D-C1)
 D'Angelo, Pina, 66, 280 (3D-I1)
 Danila, Oana, 37, 140 (1E-C2)
 Dankwa, Ernest, 38, 147 (1E-C4)
 Darlington, Gerarda, 46 (2C-P2), 64 (3B-C2), 181 (2C-P2), 270 (3B-C2)
 Datta, Gauri, 35, 132 (1E-I5)
 Davies, Katherine, 58, 246 (2E-C4)
 Davison, Matt, 34, 128 (1E-I2)
 Dawson, Don, 42, 164 (2B-I6)
 Dean, Charmaine, 47 (2C-P2), 53 (2D-C4), 187 (2C-P2), 220 (2D-C4)
 Deardon, Rob, 46, 48 (2C-P2), 54 (2D-C5), 68 (3D-I6), 181, 191 (2C-P2), 223 (2D-C5), 290 (3D-I6)
 Dehling, Herold, 35, 130 (1E-I3)

- delMas, Bob, 67, 284 (3D-I3)
Demnati, Abdellatif, 48, 195 (2D-I1)
Deng, Dianliang, 32, 112 (1D-C2)
Derkach, Andriy, 50, 204 (2D-I5)
de TIBEIRO, Jules Joseph Sunda, 53, 222 (2D-C5)
Diao, Liqun, 66, 278 (3C-A3)
Dinu, Irina, 51 (2D-C1), 63 (3B-C2), 210 (2D-C1), 269 (3B-C2)
Do, Kim-Anh, 51, 209 (2D-C1)
Doecke, James, 51, 209 (2D-C1)
Dong, Chi, 48, 193 (2C-P3)
Donner, Allan, 37, 140 (1E-C2)
Doss, Charles, 42, 162 (2B-I5)
Drekic, Steve, 69, 292 (3D-C1)
Du, Fang, 50, 204 (2D-I5)
Duanmu, Haosui, 48, 193 (2C-P3)
Dubin, Joel, 31, 107 (1D-I6)
Duchi, John, 22, 72 (1A-A2)
Dueck, Seth, 68, 290 (3D-I6)
Dunham, Bruce, 21, 67, 284 (3D-I3)
Dupuis, Debbie, 58 (2E-C5), 69 (3D-C1), 249 (2E-C5), 293 (3D-C1)
Duque, Shamika Thirimanna, Mi Shen, Pawel Morzywolek, Daniel Rodriguez, 28 (1C-D2)
Durand, Claire, 66, 279 (3C-A4)
Durocher, Stephane, 44, 176 (2B-C4)
Duschesne, Pierre, 44, 174 (2B-C3)
Dziociolowski, Krzysztof, 67, 282 (3D-I2)
- El Ghouch, Anouar, 53, 219 (2D-C4)
El Ktaibi, Farid, 25, 89 (1B-C2)
Elliott, Michael, 55, 232 (2E-I4)
Escarela, Gabriel, 53, 222 (2D-C5)
Espin-Garcia, Osvaldo, 46, 183 (2C-P2)
Essis-Breton, Nicolas, 69, 293 (3D-C1)
Evans, Michael, 25, 87 (1B-C1)
Evans, Mike, 22 (1A-A1)
- Fan, Jianqing, 41, 157 (2B-I2)
Fan, Yingying, 40, 156 (2B-I2)
Fan, Zhao Zhi, 53, 219 (2D-C4)
Fang, Ethan, 49, 198 (2D-I2)
Farrell, Patrick, 32 (1D-C3), 47 (2C-P2), 116 (1D-C3), 184 (2C-P2)
Favre Martinoz, Cyril, 59, 250 (2E-C5)
Feldman, Brian, 54 (2E-I1)
Feng, Cindy X., 42, 166 (2B-C1)
Feng, Feng (Frank), 54, 222 (2D-C5)
- Feng, Runhuan, 40, 154 (2B-I1)
Feng, Zeny, 48, 191 (2C-P2)
Fick, Gordon H., 57, 241 (2E-C2)
Firouzi, Hassan Omid, 31, 109 (1D-C1)
Flachaire, Emmanuel, 24, 83 (1B-I6)
Fode, Tounkara, 26, 95 (1B-C4)
Fontaine, Charles, 53, 219 (2D-C4)
Ford, L. Andria, 50, 202 (2D-I4)
Foster, Gary, 37, 138 (1E-C1)
Fraser, Don, 39, 150 (1E-C5)
Fraser, Robert, 44, 176 (2B-C4)
Frees, Edward W., 55, 229 (2E-I3)
Frei, Christoph, 51, 206 (2D-I6)
Fried, Roland, 35, 130 (1E-I3)
Fu, Guangyu, 62, 264 (3B-I6)
Fung, Kaiser, 67, 282 (3D-I2)
Funk, Benedikt, 53, 219 (2D-C4)
Furman, Edward, 24 (1B-I6), 56 (2E-C1), 83 (1B-I6), 238 (2E-C1)
- Gadag, Veeresh, 37, 141 (1E-C2)
Gagliano, Sarah, 24 (1B-I4), 33 (1D-C3), 80 (1B-I4), 117 (1D-C3)
Gaillardetz, Patrice, 69, 293 (3D-C1)
Gao, Huan, 46, 179 (2C-P1)
Gao, Xin, 43, 171 (2B-C2)
Garrido, Jose, 63, 267 (3B-C1)
Garrido, José, 49, 200 (2D-I3)
GARRO, ANDREW, 46, 183 (2C-P2)
Gauthier, Geneviève, 31 (1D-C1), 69 (3D-C1), 109 (1D-C1), 293 (3D-C1)
Gavanji, Parisa, 57, 239 (2E-C2)
Genest, Christian, 31 (1D-C1), 63 (3B-C1), 109 (1D-C1), 267 (3B-C1)
George, Edward, 22 (1B-I1), 68 (3D-I5), 73 (1B-I1), 289 (3D-I5)
Ghaemi, Mohammad Sajjad, 46, 184 (2C-P2)
Ghahramani, Melody, 38, 148 (1E-C4)
Ghodsi, Ali, 23, 75 (1B-I2)
Ghosh, Debashis, 69, 295 (3D-C2)
Gibbs, Alison, 31, 108 (1D-I6)
Gillis, Daniel, 46, 181 (2C-P2)
Girard, Claude, 34, 123 (1D-C5)
Godin, Frédéric, 69, 293 (3D-C1)
Golchi, Shirin, 33, 120 (1D-C4)
Goldenberg, Maya, 62, 262 (3B-I5)
Gong, Lan, 49, 200 (2D-I3)
Gouzi, Naima, 27, 95 (1B-C4)

- Gramacy, Robert, 49, 197 (2D-I2)
 Graubard, Barry, 55, 231 (2E-I4)
 Gravel, Christopher, 47, 184 (2C-P2)
 Gear, Daniel, 68, 291 (3D-I6)
 Greenwood, Celia, 52, 212, 214 (2D-C2)
 Guerin, Helene, 49, 200 (2D-I3)
 Guérin, Hélène, 68, 287 (3D-I4)
- Habeck, Christian, 50, 202 (2D-I4)
 Hackmann, Daniel, 67, 286 (3D-I4)
 Halder, Avik, 33, 120 (1D-C4)
 Halladay, Amanda, 27, 96 (1B-C4)
 Hamid, Jemila, 33 (1D-C3), 43 (2B-C1), 51 (2D-C1), 60 (3B-I1), 116 (1D-C3), 167 (2B-C1), 208 (2D-C1), 253 (3B-I1)
 Han, David, 54, 223 (2D-C5)
 Han, Dezhao, 56, 237 (2E-C1)
 Hannachi, Abdel, 56, 235 (2E-I6)
 Harrison, Matthew, 62, 260 (3B-I4)
 Hasler, Caren, 34, 123 (1D-C5)
 Haziza, David, 34 (1D-C5), 59 (2E-C5), 124 (1D-C5), 250 (2E-C5)
 He, Bing, 50, 204 (2D-I5)
 He, Kevin, 61, 256 (3B-I2)
 Heeringa, Steven, 20, 22, 73 (1B-I1)
 Heo, Giseon, 44 (2B-C4), 47 (2C-P2), 69 (3D-C2), 177 (2B-C4), 185 (2C-P2), 296 (3D-C2)
 Hesham, Tameem, 23, 75 (1B-I2)
 Higdon, Dave, 22, 73 (1B-I1)
 Hines, Gordon, 34, 126, 127 (1E-I1)
 Holzmann, Hajo, 23, 77 (1B-I3)
 Hong, Chen, 66, 281 (3D-I1)
 Horrocks, Julie, 64, 270 (3B-C2)
 Hossain, Shahadut, 43, 167 (2B-C1)
 Hossain, Shakhawat, 38, 147, 148 (1E-C4)
 Hu, Joan, 30, 103 (1D-I4)
 Hu, Yu Luo, Zhe Li, Ya-Ting Chang, Shuanglu Liu, Bo, 28 (1C-D2)
 Huang, Mei Ling, 26 (1B-C3), 53 (2D-C4), 94 (1B-C3), 220 (2D-C4)
 Huang, Minyi, 51, 206 (2D-I6)
 Hunt, Scott, 68, 290 (3D-I6)
 Hwang, Heungsun, 52, 213 (2D-C2)
 Hyslop, Theresa, 35 (1E-I4)
- Ingrassia, Salvatore, 41, 160 (2B-I4)
 Inoussa, Rofick, 69, 293 (3D-C1)
 Islam, Shofiqul, 43, 167 (2B-C1)
- Ivanoff, Gail, 25, 89, 90 (1B-C2)
- Jafari Jozani, Mohammad, 33 (1D-C3), 47 (2C-P2), 118 (1D-C3), 189 (2C-P2)
 Jahromi, Kian, 32, 112 (1D-C2)
 Jana, Sayantee, 33 (1D-C3), 60 (3B-I1), 116 (1D-C3), 253 (3B-I1)
 JANG, Gun Ho, 63, 269 (3B-C2)
 Jankowski, Hanna, 29 (1D-I1), 43, 171 (2B-C2)
 Jerry, Maples, 35, 132 (1E-I5)
 Jessiman, B., 47, 187 (2C-P2)
 Jewell, Sean, 26, 90 (1B-C2)
 Ji, Hongkai, 50, 204 (2D-I5)
 Jiang, Jiming, 35, 36, 132 (1E-I5)
 Jiang, Wenyu, 57, 239 (2E-C2)
 Jin, Zhezhen, 61, 258 (3B-I3)
 J.N.K., Rao, 48, 195 (2D-I1)
 Joe, Harry, 58, 244 (2E-C3)
 John, Amrhein, 66, 281 (3D-I1)
 Johnson, Brad, 47, 189 (2C-P2)
 Jones, Bradley, 65, 277 (3C-A2)
 Jones, Bruce L., 46, 179 (2C-P1)
 Jordan, Michael, 22, 72 (1A-A2)
 Jovic, B., 47, 187 (2C-P2)
 Juarez-Cerrillo, Sergio Francisco, 46, 181 (2C-P2)
 Juarez-Colunga, Elizabeth, 47, 187 (2C-P2)
 Julien, Pierre-Olivier, 27, 96 (1B-C4)
 Jun, Seong-Hwan, 26, 93 (1B-C3)
 Jun, Tingting Yu, Seong-Hwan, 28 (1C-D1)
- Kalbfleisch, John, 61, 255 (3B-I2)
 Kasahara, Hiroyuki, 23, 78 (1B-I3)
 Katsahian, Sandrine, 25, 85 (1B-C1)
 Khadraoui, Khader, 33, 117 (1D-C3)
 Khalili, Abbas, 69, 297 (3D-C2)
 Khan, Mohmmad Kaviul, 36, 136 (1E-C1)
 Khan, Shahedul, 51, 208 (2D-C1)
 Khodayari Moez, Elham, 63, 269 (3B-C2)
 Kiermer, Veronique, 35 (1E-I4)
 Kim, Sungsoo, 69, 292 (3D-C1)
 Klar, Neil, 37, 140 (1E-C2)
 Knight, Jo, 24 (1B-I4), 33 (1D-C3), 80 (1B-I4), 117 (1D-C3)
 Komaki, Fumiyasu, 68, 288 (3D-I5)
 Kondo, Yumi, 47, 185 (2C-P2)
 Kong, Linglong, 50 (2D-I4), 57 (2E-C2), 202 (2D-I4), 240 (2E-C2)
 Kortbi, Othmane, 64, 272 (3B-C3)

- Kouritzin, Michael, 51, 206 (2D-I6)
 Kovacevic, Milorad, 48, 195 (2D-I1)
 Kovacev-Nikolic, Violeta, 47, 185 (2C-P2)
 Krewski, Daniel, 32 (1D-C3), 47 (2C-P2), 116 (1D-C3),
 184 (2C-P2)
 Krupskiy, Pavel, 58, 244 (2E-C3)
 Kubokawa, Tatsuya, 68, 288 (3D-I5)
 Kulik, Rafal, 25, 90 (1B-C2)
 Kulperger, Reg, 35, 130 (1E-I3)
 Kuznetsov, Alexey, 67, 286 (3D-I4)
 Kwong, Grace P.S., 54, 223 (2D-C5)
 Kyprianou, Andreas, 68, 286 (3D-I4)
- Labadi, Luai Al, 44, 173 (2B-C3)
 Labbe, Aurelie, 52, 213 (2D-C2)
 Lafaye de Micheaux, Pierre, 44, 174 (2B-C3)
 Lafleur, François, 57, 243 (2E-C3)
 Lakhal-Chaieb, Lajmi, 36 (1E-I6), 52 (2D-C2), 134 (1E-
 I6), 212 (2D-C2)
 Landsman, Victoria, 55, 231 (2E-I4)
 Larrivée-Hardy, Etienne, 31, 110 (1D-C1)
 Lau, Shariah Shams, Tim To, Ruoyong Xu Changchang
 Xu, Lei Miao, Gabriel, 28 (1C-D2)
 Lau, Shariah Shams, Tim To, Ruoyong Xu, Changchang
 Xu, Lei Miao, Gabriel, 28 (1C-D1)
 Law, Casey, 56, 233 (2E-I5)
 Lawless, Jerald F, 32, 113 (1D-C2)
 Lawless, Jerry, 50 (2D-I5), 64 (3B-C2), 204 (2D-I5), 271
 (3B-C2)
 Lawrence, Earl, 56, 233 (2E-I5)
 Lebenbaum, Michael, 46, 183 (2C-P2)
 Leblanc, Alexandre, 26 (1B-C3), 44 (2B-C4), 93 (1B-C3),
 176 (2B-C4)
 Leclerc, Martin, 36, 134 (1E-I6)
 Lee, Sharon, 41, 160 (2B-I4)
 Lee, Shoo K., 47, 188 (2C-P2)
 Lee, Shun Fu, 37, 140 (1E-C2)
 Lee, Sunghee, 55, 232 (2E-I4)
 Lee, Wan-Chen, 33, 121 (1D-C4)
 Lefebvre, Geneviève, 64, 270 (3B-C2)
 Leichtman, Alan, 61, 255 (3B-I2)
 Lemieux, Christiane, 49, 199 (2D-I3)
 Leung, Andy, 21
 Li, Deli, 58, 246 (2E-C4)
 Li, Haijun, 55, 229 (2E-I3)
 Li, Jia, 47, 186 (2C-P2)
 Li, Longhai, 42, 166 (2B-C1)
 Li, Pengfei, 23 (1B-I3), 29 (1D-I1), 77 (1B-I3)
 Li, Ye, 46, 183 (2C-P2)
 Li, Yi, 61, 256 (3B-I2)
 Li, Zhong, 47, 186 (2C-P2)
 Liang, Feng, 68, 289 (3D-I5)
 Liang, Kun, 25, 85 (1B-C1)
 Lin, Luyao, 46, 179 (2C-P1)
 Lin, Shivon Sue-Chee, Lingling Fan, Wei, 28 (1C-D1)
 Lin, Wei, 39, 150 (1E-C5)
 Lin, X. Sheldon, 49, 200 (2D-I3)
 Lin, Zhenhua, 64, 273 (3B-C3)
 Lingling, Fan, 55, 231 (2E-I4)
 Liu, Fangda, 49, 199 (2D-I3)
 Liu, Fei Huang, Yuying Xie, Di Shu, Ken Xiao, Xin, 28
 (1C-D2)
 Liu, Juxin, 68, 290 (3D-I6)
 Liu, Kuan, 25, 85 (1B-C1)
 Liu, Simon Greaves, Xinyu Zhang, Yao Song, Yicen Guo,
 Xi, 28 (1C-D2)
 Liu, Xiaoming, 46, 179 (2C-P1)
 Liu, Yukun, 23, 77 (1B-I3)
 Liu, Zi Zhen, 35, 130 (1E-I3)
 Lix, Lisa M., 57, 241 (2E-C2)
 Lockhart, Richard, 39, 152 (1E-C5)
 Loeffen, Ronnie, 68, 287 (3D-I4)
 Loepky, Jason, 49 (2D-I2), 56 (2E-I5), 197 (2D-I2), 234
 (2E-I5)
 Lokku, Lauren Erdman, Jennifer Gillis, Naim Panjwani,
 Osvaldo Espin-Garcia, Armend, 29 (1C-D2)
 Loredó-Osti, J Concepción, 26 (1B-C3), 43 (2B-C2), 93
 (1B-C3), 172 (2B-C2)
 Loughin, Tom, 44, 176 (2B-C4)
 Louis-Paul, Rivest, 26, 95 (1B-C4)
 Lourenço, Vanda Milheiro, 32, 113 (1D-C2)
 Lu, Wilson, 34, 124 (1D-C5)
 Lu, Xuewen, 57, 240 (2E-C2)
 Lu, Yi, 26 (1B-C2), 46 (2C-P1), 91 (1B-C2), 179 (2C-P1)
 Lu, Zihang, 36, 137 (1E-C1)
 Lundy, Erin, 47, 187 (2C-P2)
 Luong, Andrew, 31, 109 (1D-C1)
 Lv, Jinch, 40, 156 (2B-I2)
 Lysy, Martin, 49, 198 (2D-I2)
- Ma, Lily, 22, 74 (1B-I1)
 Macdonald, Ahsan Bhatti, Chengwei Qin, Hedan Han,
 Keya Biswas, Regina Kampo, Sandip Barui, Pe-
 ter, 29 (1C-D2)
 Mackay, Anne, 40, 154 (2B-I1)

- MacKay, Jock, 37 (1E-C2), 52 (2D-C3), 140 (1E-C2), 217 (2D-C3)
- MacNab, Ying, 30, 101 (1D-I3)
- Mailhot, Mélina, 63, 266 (3B-C1)
- Mak, Simon, 26, 91 (1B-C2)
- Malik, Rajat, 54, 223 (2D-C5)
- Mallette, Hervé Mensah, Imad Yanni, Ghizlene Zerghini, Jean-Phillipe, 29 (1C-D2)
- Mamon, Rogemar, 46, 179 (2C-P1)
- Mamun, Mahmud, 47, 187 (2C-P2)
- Mandal, Saumen, 29, 99 (1D-I2)
- Marceau, Etienne, 31, 110 (1D-C1)
- Marchand, Eric, 64 (3B-C3), 68 (3D-I5), 272 (3B-C3), 288 (3D-I5)
- Marchand, Éric, 58 (2E-C4), 64 (3B-C3), 246 (2E-C4), 273 (3B-C3)
- Marriott, Paul, 62, 261 (3B-I4)
- Martell, David, 39, 150 (1E-C5)
- Masuadi, Emad, 44, 173 (2B-C3)
- Maurer, Willi, 54, 227 (2E-I2)
- Maximova, Katerina, 36, 136 (1E-C1)
- Maxwell, Christian, 34, 128 (1E-I2)
- McCulloch, Robert, 22, 73 (1B-I1)
- McDougall, Janet E.A., 66, 281 (3D-I1)
- McGill, Jeffrey, 67, 283 (3D-I2)
- McLachlan, Geoff, 41, 160 (2B-I4)
- McLeish, Don, 42, 164 (2B-I6)
- McLeod, A.Ian, 48, 192 (2C-P2)
- McLeod, Ian, 35, 130 (1E-I3)
- McNicholas, Paul, 38 (1E-C3), 41 (2B-I4), 58 (2E-C3), 145 (1E-C3), 160 (2B-I4), 244 (2E-C3)
- McNicholas, Paul D., 38 (1E-C3), 38 (1E-C4), 144 (1E-C3), 149 (1E-C4)
- Mellipilan, Roberto Edmundo, 22, 73 (1B-I1)
- Mendivil, Franklin, 52, 216 (2D-C3)
- Metzler, Adam, 35 (1E-I2), 58 (2E-C4), 128 (1E-I2), 247 (2E-C4)
- Meyer, Mary, 42, 162 (2B-I5)
- Mihram, G. Arthur, 54, 224 (2D-C5)
- Mihram, Ph. D., Danielle, 54, 224 (2D-C5)
- Milevsky, Moshe, 40, 155 (2B-I1)
- Millar, Michele, 67, 285 (3D-I3)
- Miller, Arden Ewan, 57, 243 (2E-C3)
- Mirea, Lucia, 47, 188 (2C-P2)
- Mirshahi, Mina, 44, 176 (2B-C4)
- Mitric, Ilie-Radu, 56, 238 (2E-C1)
- Mizera, Ivan, 39, 151 (1E-C5)
- Moez, Elham Khodayari, 63, 269 (3B-C2)
- Mohammad, Nagham, 35, 130 (1E-I3)
- Mohammadi, Farhood, 51, 209 (2D-C1)
- Moiseev, Alexander, 50, 203 (2D-I4)
- Monahan, Adam, 56, 235 (2E-I6)
- Monahan, Adam H., 56, 236 (2E-I6)
- Moodie, Erica, 57, 241 (2E-C2)
- Moon, Nathalie, 57, 239 (2E-C2)
- Morales, Manuel, 49, 200 (2D-I3)
- Morin, Amy, 39, 150 (1E-C5)
- Morrison, Jason, 44, 176 (2B-C4)
- Mouret, Geoffroy, 38, 144 (1E-C3)
- Muhe, Armend Lokku, Nadia, 28 (1C-D1)
- Mulayath Variyath, Asokan, 43, 172 (2B-C2)
- Mullah, Muhammad Abu Shadeque, 52, 212 (2D-C2)
- Murray, Paula, 58, 244 (2E-C3)
- Muthukumarana, Saman, 25 (1B-C1), 27 (1B-C4), 87 (1B-C1), 97 (1B-C4)
- Nadarajah, Tharshanna, 43, 172 (2B-C2)
- Nadeem, Khurram, 47, 188 (2C-P2)
- Naik, Shanoja, 32, 110 (1D-C1)
- Nasari, M. Masoud, 42, 164 (2B-I6)
- Nasari, Masoud M, 38, 148 (1E-C4)
- Nathoo, Farouk, 50, 203 (2D-I4)
- Nguyen, Thuan, 35, 36, 132 (1E-I5)
- Niccols, Alison, 37, 138 (1E-C1)
- Nikolic, Dragan, 47, 185 (2C-P2)
- Ning, Yang, 61 (3B-I3), 69 (3D-C2), 259 (3B-I3), 296 (3D-C2)
- Nirmalkanna, Kunasekaran, 47, 189 (2C-P2)
- Niu, Yi, 61, 258 (3B-I3)
- Nkurunziza, Sévérien, 58, 249 (2E-C5)
- Nolde, Natalia, 24, 84 (1B-I6)
- Nourmohammadi, Mohammad, 47, 189 (2C-P2)
- Nudds, Tom, 34, 127 (1E-I1)
- Offringa, Martin, 54 (2E-I1)
- O'Hara Hines, Jeanette, 34, 126 (1E-I1)
- O'Hara Hines, R. Jeanette, 34, 126 (1E-I1)
- Okhrin, Ostap, 60, 254 (3B-I1)
- Omidi, Hassan, 49, 200 (2D-I3)
- On behalf of CONNECTIONS, (CONNECTIONS), 37, 138 (1E-C1)
- Orbanz, Peter, 23, 75 (1B-I2)
- Ortmann, Janosch, 31, 106 (1D-I5)
- Osgood, Nathaniel, 68, 290 (3D-I6)
- Oualkacha, Karim, 52, 212, 214 (2D-C2)
- Owen, Art, 65, 276 (3C-A1)

- Oyet, Alwell, 37, 141 (1E-C2)
- PADILLA, ALBERTO, 47, 190 (2C-P2)
- Papageorgiou, Nicolas, 58, 249 (2E-C5)
- Pardo, Juan Carlos, 68, 286 (3D-I4)
- Park, Chul Gyu, 52, 213 (2D-C2)
- Parker, Gary, 46, 179 (2C-P1)
- Partovi Nia, Vahid, 38 (1E-C3), 44 (2B-C4), 144 (1E-C3), 176 (2B-C4)
- Partovi-Nia, Vahid, 46, 184 (2C-P2)
- Paterson, Andrew, 24, 79 (1B-I4)
- Pelletier, Eric, 27, 96 (1B-C4)
- Peng, Jianan, 26 (1B-C3), 55 (2E-I2), 69 (3D-C2), 92 (1B-C3), 227 (2E-I2), 295 (3D-C2)
- Peng, Yingwei, 61, 258 (3B-I3)
- Perreault, Luc, 57, 243 (2E-C3)
- Perron, Francois, 33, 121 (1D-C4)
- Perron, François, 58 (2E-C4), 64 (3B-C3), 246 (2E-C4), 273 (3B-C3)
- Petkau, John, 47, 185 (2C-P2)
- Petrov, Pavel, 44, 177 (2B-C4)
- Pezeshk, Hamid, 36, 137 (1E-C1)
- Picka, Jeff, 62, 262 (3B-I5)
- Pierre, Duchesne, 44, 173 (2B-C3)
- Pignal, Jean, 22, 74 (1B-I1)
- Pinky, Saadia, 69, 296 (3D-C2)
- Pires, Ana M., 32, 113 (1D-C2)
- Platt, Robert, 33 (1D-C3), 35 (1E-I4), 54 (2E-I1), 118 (1D-C3)
- Pohlkamp-Hartt, Joshua, 44, 175 (2B-C3)
- Poljak, Zvonimir, 36, 136 (1E-C1)
- Pradel, Roger, 39, 152 (1E-C5)
- Pratola, Matt, 22, 73 (1B-I1)
- Premarathna, Lasantha, 25, 86 (1B-C1)
- Prentice, Ross, 61, 256 (3B-I2)
- Priam, Rodolphe, 48, 196 (2D-I1)
- Provost, Serge, 33, 121 (1D-C4)
- Pullenayegum, Eleanor, 63, 269 (3B-C2)
- Punzo, Antonio, 41, 160 (2B-I4)
- Pyne, Saumyadipta, 51, 210 (2D-C1)
- Qian, Peter, 29, 99 (1D-I2)
- Qianyi, Huang, 66, 281 (3D-I1)
- Qiu, Jiaheng, 29, 100 (1D-I2)
- Qiu, Shi, 42, 166 (2B-C1)
- Quastel, Jeremy, 31, 106 (1D-I5)
- Quessy, Jean-François, 58, 247 (2E-C4)
- Quijano, Oscar, 49, 200 (2D-I3)
- Rad, Narges Nazeri, 32, 113 (1D-C2)
- Rahim, Karim, 45, 178 (2B-C4)
- Rahman, Adam, 29 (1C-D2)
- Ranjan, Pritam, 52, 53, 216, 217 (2D-C3)
- Rao, J. Sunil, 36, 132 (1E-I5)
- Rees, Michael, 61, 255 (3B-I2)
- Reese, Shane, 56, 233 (2E-I5)
- Reesor, R. Mark, 63, 268 (3B-C1)
- Reesor, R.M., 35, 128 (1E-I2)
- Reid, Nancy, 39 (1E-C5), 43 (2B-C2), 61 (3B-I3), 69 (3D-C2), 150 (1E-C5), 170 (2B-C2), 259 (3B-I3), 296 (3D-C2)
- Remenik, Daniel, 31, 106 (1D-I5)
- Rémillard, Bruno, 58, 249 (2E-C5)
- Ren, Jiandong, 63, 267 (3B-C1)
- Renaud, Jean-Francois, 68, 287 (3D-I4)
- Riegert, David, 64, 274 (3B-C3)
- Robinson, Beren, 34, 127 (1E-I1)
- Rodrigues, Paulo C., 32, 113 (1D-C2)
- Romdhani, Hela, 52, 213 (2D-C2)
- Rooch, Aeneas, 35, 130 (1E-I3)
- Rosella, Laura, 46, 183 (2C-P2)
- Rosenthal, Jeffrey, 40, 153 (2A-A)
- Rosolowsky, Erik, 49 (2D-I2), 56 (2E-I5), 197 (2D-I2), 234 (2E-I5)
- Rosychuk, Rhonda, 30, 101 (1D-I3)
- Rotondi, Michael, 27, 97 (1B-C4)
- Ruth, William, 44, 176 (2B-C4)
- Sadeghi, Mehdi, 36, 137 (1E-C1)
- Sadeghi, Samira, 25, 89 (1B-C2)
- Safari, Abdollah, 36, 137 (1E-C1)
- Safari, Elena Szefer, Abdollah, 28 (1C-D1)
- Saha, Krishna, 25, 86 (1B-C1)
- Saha, Krishna K., 44, 174 (2B-C3)
- Sajobi, Tolulope, 57, 241 (2E-C2)
- Salakhutdinov, Ruslan, 23, 76 (1B-I2)
- Salisbury, Tom, 40, 155 (2B-I1)
- Salvi, Giovanni, 35, 129 (1E-I2)
- Sanches, Marcos, 59, 250 (2E-C5)
- Sangalli, Laura M., 39, 151 (1E-C5)
- Sango, Joel, 44, 173 (2B-C3)
- Sattar, Abdus, 30, 103 (1D-I4)
- Saulnier, Justine Zehr, Steve Ferreira, Bertrand, 28 (1C-D1)
- Schau, Candace, 67, 285 (3D-I3)
- Schaub, Michael, 39, 152 (1E-C5)
- Schliep, Erin, 68, 291 (3D-I6)

- Schmirler, Matthew, 38, 148 (1E-C4)
 Schnitzer, Mireille, 41, 158 (2B-I3)
 Schulz, Juliana, 63, 267 (3B-C1)
 Schwaiger, Florian, 23, 77 (1B-I3)
 Schwarz, Carl, 39, 152 (1E-C5)
 Schwarz, Carl James, 25, 86 (1B-C1)
 Scott, Alexandre, 58, 247 (2E-C4)
 Scott, Steve, 22, 73 (1B-I1)
 Selvaratnam, Selvakkadunko, 37, 141 (1E-C2)
 Sen, Bodhi, 42, 162 (2B-I5)
 Sen, Debaraj, 44, 174 (2B-C3)
 Sendova, Kristina, 47 (2C-P2), 49 (2D-I3), 186 (2C-P2),
 201 (2D-I3)
 Severn, Daniel, 37, 140 (1E-C2)
 Sezer, Deniz, 31, 105 (1D-I5)
 Shaddick, Gavin, 20
 Shah, Prakesh S., 47, 188 (2C-P2)
 Shary, Osvaldo Espin-Garcia, Simon Greaves, Sepehr
 Jafar-Pour, Yitian Tao, Kate Tsiplova, Coralie Wong,
 28 (1C-D1)
 She, Xichen, 34, 124 (1D-C5)
 Shi, Qian, 57, 240 (2E-C2)
 Shi, Xiaozhou Wang, Sile Tao, Qian, 29 (1C-D2)
 Shimotsu, Katsumi, 23, 78 (1B-I3)
 Shin, H.S., 47, 187 (2C-P2)
 Shin, Ji-Hyung, 52, 214 (2D-C2)
 Shlomo, Natalie, 48, 196 (2D-I1)
 Shokoohi, Farhad, 69, 297 (3D-C2)
 Simard, Jacques, 36, 134 (1E-I6)
 Sinaei, Leila, 47, 190 (2C-P2)
 Singh, Gurbakhshash, 57, 241 (2E-C2)
 Sinha, Sanjoy, 30, 103 (1D-I4)
 Sirken, Monroe, 55, 231 (2E-I4)
 Sivathayalan, Ahalya, 32, 116 (1D-C3)
 Skala, Matthew, 44, 176 (2B-C4)
 Smith-Doiron, M., 47, 187 (2C-P2)
 Soave, David, 43, 168 (2B-C1)
 Sohrabi, Maryam, 39, 152 (1E-C5)
 Song, Peter, 60 (3B-I1), 61 (3B-I2), 254 (3B-I1), 255 (3B-
 I2)
 Spencer, Brittany, 56, 233 (2E-I5)
 Spencer, Neil, 26 (1B-C2), 52 (2D-C3), 90 (1B-C2), 216
 (2D-C3)
 Spiroulias, Amanda, 63, 266 (3B-C1)
 Springford, Aaron, 33 (1D-C4), 64 (3B-C3), 122 (1D-C4),
 274 (3B-C3)
 Srestha, Nabin, 24, 79 (1B-I4)
 Srivastava, Nitish, 23, 76 (1B-I2)
 Stafford, Jamie, 22 (1A-A1)
 Stanley, Anu, 48, 191 (2C-P2)
 Steiner, Stefan, 37 (1E-C2), 52 (2D-C3), 140 (1E-C2),
 217 (2D-C3)
 Stephens, Dave, 41, 159 (2B-I3)
 Stephens, David A., 46, 181 (2C-P2)
 Stevens, Nathaniel, 52, 217 (2D-C3)
 Stieb, D., 47, 187 (2C-P2)
 Stingo, Francesco, 51, 209 (2D-C1)
 Strawderman, William, 68, 288 (3D-I5)
 Su, Jianxi, 24 (1B-I6), 56 (2E-C1), 83 (1B-I6), 238 (2E-
 C1)
 Su, Wanhua, 45, 177 (2B-C4)
 Sujanthan, Andy Koh, Gabriela Gonzalez, Jordan Collins,
 John Jarvis, Sriskondarajah, 29 (1C-D2)
 Sun, Jianping, 52, 214 (2D-C2)
 Sun, Lei, 43 (2B-C1), 50 (2D-I5), 168 (2B-C1), 204 (2D-
 I5)
 Sun, Wei, 50, 205 (2D-I5)
 Sun, Zheng, 44, 175 (2B-C3)
 Sura, Philip, 56, 235 (2E-I6)
 Suzer Gurtekin, Tuba, 55, 232 (2E-I4)
 Swishchuk, Anatoliy, 35, 129 (1E-I2)
 Sylvestre, Marie-Pierre, 24, 79 (1B-I4)
 Takahara, Glen, 33, 120 (1D-C4)
 Talbot, Denis, 64, 270 (3B-C2)
 Tan, Ken Seng, 56, 237 (2E-C1)
 Tang, Boxin, 47, 189 (2C-P2)
 Tang, Dameng, 49, 200 (2D-I3)
 Tang, Yang, 38, 145 (1E-C3)
 Tao, Nadia Muhe, Coralie Wong, Yitian F., 29 (1C-D2)
 Tao, Sile, 39, 151 (1E-C5)
 Taqqu, Murad, 35, 130 (1E-I3)
 Tatsinkou, Joseph François Tagne, 44, 174 (2B-C3)
 Tekougang, Thierry Chekouo, 51, 209 (2D-C1)
 Thabane, Lehana, 37 (1E-C1), 43 (2B-C1), 54 (2E-I1),
 138 (1E-C1), 167 (2B-C1)
 Thomas, Doneal, 25, 87 (1B-C1)
 Thomassin, Tommy, 31, 109 (1D-C1)
 Thompson, Mary, 43, 171 (2B-C2)
 Thompson, William, 56, 236 (2E-I6)
 Thomson, David, 33 (1D-C4), 45 (2B-C4), 60 (3A-A2), 64
 (3B-C3), 122 (1D-C4), 178 (2B-C4), 252 (3A-
 A2), 274 (3B-C3)
 Thomson, Trevor, 38, 148 (1E-C4)
 Thorlund, Kristian, 24, 81 (1B-I5)
 Thorpe, Kevin, 31, 107 (1D-I6)

- Thorpe, Lucas, 26, 94 (1B-C3)
 Tibshirani, Rob, 20
 Tichon, Jenna, 27, 97 (1B-C4)
 Tillé, Yves, 34, 123 (1D-C5)
 Tong, Jinjun, 26, 93 (1B-C3)
 Torabi, Mahmoud, 30 (1D-I3), 33 (1D-C3), 35 (1E-I5),
 102 (1D-I3), 118 (1D-C3), 132 (1E-I5)
 Torkashvand, Elaheh, 33, 118 (1D-C3)
 Trejo, Maria Esther Perez, 33, 118 (1D-C3)
 Trépanier, Martin, 46, 184 (2C-P2)
 Troupé, Marylène, 32, 114 (1D-C2)
 Trufin, Julien, 56, 238 (2E-C1)
 Trufinh, Julien, 31, 110 (1D-C1)
 Tsanakas, Andreas, 69, 292 (3D-C1)
- Vaillant, Jean, 32, 114 (1D-C2)
 Valdez, Emiliano, 55, 229 (2E-I3)
 Vallée, Audrey-Anne, 34, 124 (1D-C5)
 van Bommel, Matthew, 53, 217 (2D-C3)
 Vander Wiel, Scott, 56, 233 (2E-I5)
 Vanniyasingam, Thuva, 37, 138 (1E-C1)
 Varet, Hugo, 25, 85 (1B-C1)
 Variyath, Asokan, 53, 219 (2D-C4)
 Vasudevan, Chithran Vadaverkkot, 53, 219 (2D-C4)
 Vatanpour, Shabnam, 51, 210 (2D-C1)
 Vavasis, Stephen, 23, 75 (1B-I2)
 Veugelers, Paul, 36, 136 (1E-C1)
 Vickers, David, 68, 290 (3D-I6)
 Vidyashankar, Anand, 40, 156 (2B-I2)
 Vincent, Kyle, 27, 97 (1B-C4)
 Virji-Babul, Naznin, 50, 203 (2D-I4)
 Volkmer, Hans, 40, 154 (2B-I1)
 Volodin, Andrei, 63, 266 (3B-C1)
 Vrbik, Irene, 41, 160 (2B-I4)
- Wainwright, Martin, 22, 72 (1A-A2)
 Wallace, Michael, 57, 241 (2E-C2)
 Wang, Haonan, 64, 273 (3B-C3)
 Wang, Liangliang, 64, 273 (3B-C3)
 Wang, Ruodu, 69, 292 (3D-C1)
 Wang, Teresa Chan, Farhood Mohammadi, Pavel Petrov,
 Samira Sadeghi, Xiaozhou, 28 (1C-D1)
 Wang, Weichung, 29, 100 (1D-I2)
 Wang, Wen, 61, 255 (3B-I2)
 Wang, Xiaoming, 51, 210 (2D-C1)
 Wang, Xu (Sunny), 53, 218 (2D-C3)
 Wang, Yishu, 32, 114 (1D-C2)
 Wang, Zilin, 52, 217 (2D-C3)
- Watier, Francois, 58, 248 (2E-C4)
 Watson, Alex, 68, 286 (3D-I4)
 Weale, Michael, 24 (1B-I4), 33 (1D-C3), 80 (1B-I4), 117
 (1D-C3)
 Webb, Colleen, 68, 291 (3D-I6)
 Wei, Yuhong, 38, 145 (1E-C3)
 Welch, William, 46, 182 (2C-P2)
 Wellner, Jon, 42, 162 (2B-I5)
 Wen, Lan, 30 (1D-I4), 37 (1E-C2), 104 (1D-I4), 141 (1E-
 C2)
 Wendler, Martin, 35, 130 (1E-I3)
 Wendy, Lou, 55, 231 (2E-I4)
 Weng, Chengguo, 32 (1D-C1), 56 (2E-C1), 111 (1D-C1),
 237 (2E-C1)
 Weng, Jiaying (Vicky), 48, 191 (2C-P2)
 White, Bethany, 67, 285 (3D-I3)
 Whitehead, Hal, 43, 172 (2B-C2)
 Wilhelm, Matthieu, 39, 151 (1E-C5)
 Willan, Andy, 24, 82 (1B-I5)
 Wilson, Dan, 62, 264 (3B-I6)
 Wilson, David Edward, 69, 294 (3D-C1)
 Wolters, Mark, 53, 220 (2D-C4)
 Wong, Monica H.T., 38, 149 (1E-C4)
 Wong, Weng Kee, 29, 100 (1D-I2)
 Woolford, Douglas, 39, 150 (1E-C5)
 Wu, Changbao, 34 (1D-C5), 43 (2B-C2), 124 (1D-C5),
 171 (2B-C2)
 Wu, Lang, 64, 274 (3B-C3)
 Wu, Ping, 62, 264 (3B-I6)
 Wu, Ying, 37, 138 (1E-C1)
- Xiao, Jinkun, 48, 192 (2C-P2)
 Xin, Lu, 38, 146 (1E-C3)
 Xin, Xin, 64, 270 (3B-C2)
 Xu, Ximing, 43, 172 (2B-C2)
 Xu, Xinyin, 68, 289 (3D-I5)
 Xu, Yin, 53, 220 (2D-C4)
 Xue, Lingzhou, 41, 157 (2B-I2)
- Yadegari, Iraj, 64, 273 (3B-C3)
 Yan, Ying, 43, 168 (2B-C1)
 Yang, Chen, 49, 201 (2D-I3)
 Yang, Junmin, 47, 188 (2C-P2)
 Yanling, Yumi Kondo, Md Mahsin, Kyle Hambrook,, 29
 (1C-D2)
 Yao, Fang, 60, 251 (3A-A1)
 Yapa, Gaitri, 67, 284 (3D-I3)
 Yasmin, Farzana, 51, 210 (2D-C1)

Yi, Grace, 38 (1E-C4), 43 (2B-C1), 61 (3B-I3), 69 (3D-C2), 147 (1E-C4), 168 (2B-C1), 259 (3B-I3), 296 (3D-C2)

Yi, Yanqing, 37, 141, 142 (1E-C2)

Yilmaz, Yildiz, 36, 135 (1E-I6)

Yin, Guosheng, 51, 210 (2D-C1)

You, Jiashen, 35, 132 (1E-I5)

Young, Nathan, 62, 262 (3B-I5)

Yu, Hao, 20, 35, 130 (1E-I3)

Yu, Tingting, 64, 274 (3B-C3)

Yuan, Ming, 42, 163 (2B-I5)

Yuan, Yan, 43, 166 (2B-C1)

Yue, Karl Knopf, Yuanhao Lai, Jiaoying Li, Kexin Luo, Shengchao Yan, Wei, 28 (1C-D1)

Yuen, Wai Kong, 53, 220 (2D-C4)

Zarepour, Mahmoud, 39 (1E-C5), 44 (2B-C3), 152 (1E-C5), 173 (2B-C3)

Zayeri, Farid, 63, 269 (3B-C2)

Zeng, Leilei, 30 (1D-I4), 37 (1E-C2), 57 (2E-C2), 104 (1D-I4), 141 (1E-C2), 239 (2E-C2)

Zhang, Bei, 42, 166 (2B-C1)

Zhang, Peng, 45, 177 (2B-C4)

Zhang, Shulin, 60, 254 (3B-I1)

Zhang, Xingliang Huang, Honglei Xie, Ning, 29 (1C-D2)

Zhang, Xuemao, 53, 221 (2D-C4)

Zhang, Ying, 47, 188 (2C-P2)

Zhang, Yuchen, 22, 72 (1A-A2)

Zhang, Yukun, 57, 240 (2E-C2)

Zhang, Yuzhou, 46, 179 (2C-P1)

Zhao, Yinshan, 47, 185 (2C-P2)

Zheng, Zemin, 40, 156 (2B-I2)

Zhong, Yujie, 37, 142 (1E-C2)

Zhou, Alan Xinghua, 63, 268 (3B-C1)

Zhou, Michelle, 60, 254 (3B-I1)

Zhou, Xiaowen, 68, 287 (3D-I4)

Zhu, Hongtu, 50, 202 (2D-I4)

Zhu, Ji, 61, 256 (3B-I2)

Zhu, Mu, 38, 146 (1E-C3)

Zhu, Qing Xu, Pingchuan Cao, Mengnan, 28 (1C-D1)

Zhu, Yayuan, 64, 271 (3B-C2)

Zhu, Yeying, 69, 295 (3D-C2)

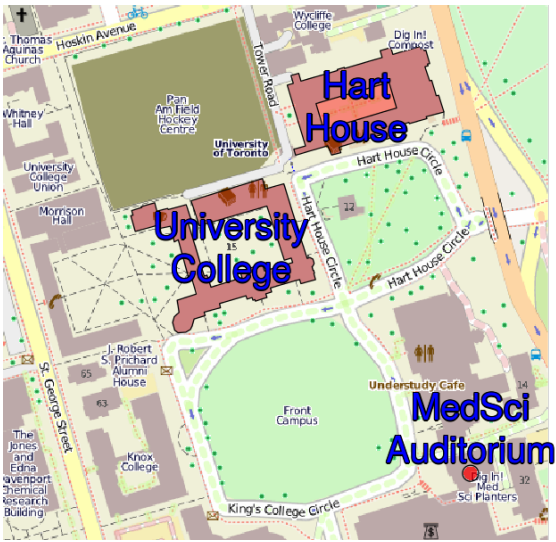
Zidek, James, 46, 182 (2C-P2)

Zidek, Jim, 20

Zitikis, Ricardas, 24, 83 (1B-I6)

Zou, Hui, 41, 157 (2B-I2)

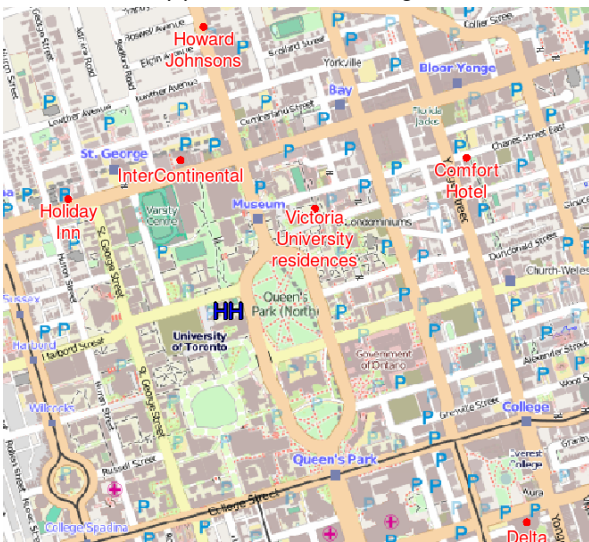
sessions	Room	Monday	Tuesday	Wednesday
am1	140 Debates Rm MedSci Auditorium	SSC Presidential Invited Address	Gold Medal Address	CRM-SSC Prize in Statistics Address SSC Impact Award Address
am2	140 161 163 179 244 A101 Croft House Debates Rm East Common Music Rm South Dining	Statistical Genetic and Genomic Mode Biostatistics - New Estimation Metho Statistical Issues in Health Economi Statistical Methods in Actuarial Sci Probability Theory and Applications Statistical Inference and Applicatio Big Data (SMS) Machine Learning Recent Developments in Finite Mixtur Survey Methodology 2	Shape-Constrained Maximum Likelihood Stochastic Processes and Their Appli Biostatistics - Measurement Errors a Nonregular Statistical Test Problems Empirical and Composite Likelihood a Statistical Inference and Applicatio New Directions in Causal Inference (New Developments in High Dimensional Semi and Unsupervised Learning (BISS Actuarial Finance (ActSci)	Biostatistics - Survival Longitudin Actuarial Science and Finance 3 Stochastic Modelling in Industry (Pr Some Specific Issues in Modelling an Statistical Education in Light of Ph Modern Statistical Methods for Compl Bioinformatics and Copula Models (Sp New Challenges and Developments in S New Developments in Statistical Neur
pm1	Great Hall	Case Study 1 Optimal Design of Experiments Statistical Inference and Applicatio Teaching Introductory Statistics wit Stochastic Spatial Processes (Probab	Actuarial Science Poster Session Statistical Challenges in High-Throu Statistical Inference and Applicatio Weakly Interacting Processes in Fina Biostatistics - Association Studies Biostatistics - Applications 3 Business and Industrial Statistics 1 Quantile, Spatial and Logistic Regre Graduate Student Presentations in Ac Gaussian Process Regression: Innovat New Methods for Analyzing Brain Imag David Binder Memorial Session (SMS)	Survey Method Section Presidential I
pm2	144 152 161 163 179 244 256 A101 Croft House Debates Rm East Common Music Rm South Dining	Clustering, Classification and Dimen Survival Methods in Genetics Biostatistics - Clinical Trials and Biostatistics - Applications 2 Statistical Inference and Applicatio Recent Developments in Small Area Es Probability Theory and Estimation Delivering a Successful Introductory Financial Mathematics New Developments in Nonlinear Time S Publishing and Reproducible Research	Robust and Shrinkage Estimation Actuarial Science and Finance 2 Statistics of Sea Surface Winds and Biostatistics - New Inferential Stra Business and Industrial Statistics 2 Statistical Challenges in Modern Ast Probability Theory and Finance Bridging the Gap between Methods Dev Recent Advances in the Analysis of C Challenges and New Developments in C Copula Theory and Methods	Predictive Density Estimation Evaluating the Effectiveness of our Levy Processes and Applications (Pro Actuarial Science and Finance 4 Biostatistics - Variable Selection Business Analytics: Where does the S Statistical Analysis for Infectious Bridging the Gap between Theory and



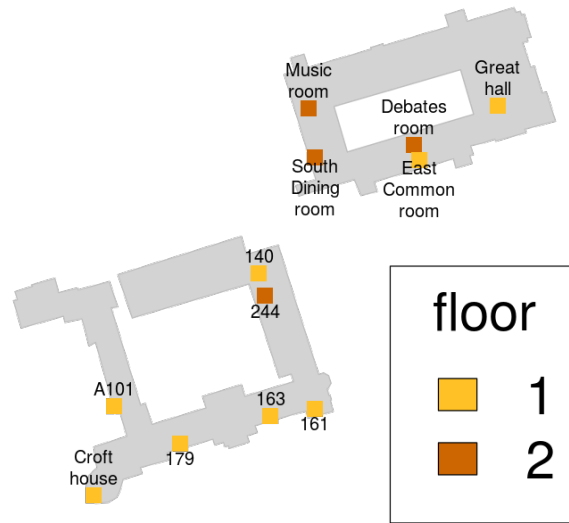
(a) Conference Buildings



(b) Food and Drink



(c) Hotels



(d) Rooms

Figure 1 – Maps by <http://openstreetmap.org> CC-BY-SA

	Sat 24/5	Sun 25/5	Mon 26/5	Tue 27/5	Wed 28/5
08:00					
09:00	09:00 – 17:00 Student Conference	09:00 – 16:30 Section Workshops	08:30 – 09:50 Presidential Invited Address	08:45 – 09:50 Gold medal address	08:45 – 09:50 Sessions 3A
10:00			10:20 – 11:50 Sessions 1B	10:20 – 11:50 Sessions 2B	10:20 – 11:50 Sessions 3B
11:00					
12:00					
13:00			13:30 – 15:00 Sessions 1C	13:30 – 15:00 Sessions 2C	13:30 – 14:30 Sessions 3C
14:00					
15:00			15:30 – 17:00 Sessions 1D	15:30 – 17:00 Sessions 2D	15:00 – 16:30 Sessions 3D
16:00					
17:00			17:00 – 18:00 AGM	17:00 – 18:00 Section AGM's	
18:00		18:30 – 21:00 Opening Reception	18:00 – 20:00 Student Barbecue	18:00 – 22:00 Banquet	
19:00					

Wifi

network: UofT

username: SSC2014

password: UofT



Statistical Society of Canada
Soci t  statistique du Canada

UNIVERSITY OF
TORONTO