



SSC Statistical Society of Canada
Société Statistique du Canada

41st Annual Meeting
of the
Statistical Society of Canada

41^e Congrès annuel
de la
Société statistique du Canada

May 26 - 29, 2013
26 au 29 mai 2013

University of Alberta

Table of Contents • Table des matières

Table of Contents • Table des matières	1
Welcome • Bienvenue	2
Sponsors • Commanditaires	3
Exhibitors • Exposants	3
Organizers • Organismes	4
Language of Presentations • Langue des présentations	5
General Information • Information générale	5
Committees and Meetings • Comités et réunions	15
Social and Information Events • Événements sociaux et informatifs	17
List of Sessions • Liste des sessions	19
Program • Programme	23
Abstracts • Résumés	65
Author List • Liste des auteurs	233

Welcome • Bienvenue

The University of Alberta welcomes you to the 41st Annual Meeting of the Statistical Society of Canada. The University of Alberta was opened in 1908 and has been an integral part of the history of Edmonton, the capital of the Province of Alberta. Since its founding with 37 students, the University of Alberta has grown to have five campuses, 18 faculties, 400 programs, and over 39,000 students. As a leading public research-intensive university, the University of Alberta has achieved excellence in many fields, including the sciences, business, engineering, and medicine. Sitting on the edge of the North Saskatchewan River valley, the North Campus (main campus) has over 150 buildings that cover about 50 blocks. Recent expansions have made substantive changes to the campus' skyline and have provided state-of-the art facilities for teaching and research. The Universiade Pavilion (Butterdome) provides athletic facilities in a landmark building. Visitors can also enjoy the main quad and the collections in the University of Alberta Museums.

Located in the heart of Alberta's capital region, Edmonton is a bustling city with over 800,000 residents. Edmonton - Canada's Festival City - is renowned for cultural, art, music, theatre and sports festivals. Visitors can always discover something new in Edmonton's shopping, sports, dining and recreation scenes as well as its world-class attractions and events. City attractions include Fort Edmonton Park, Alberta Legislature Building, Art Gallery of Alberta, and the Muttart Conservatory. Edmonton's river valley, the largest stretch of urban parkland in North America, features more than 200 kilometres of trails that will delight outdoor enthusiasts. Shoppers will enjoy Old Strathcona and the famous West Edmonton Mall. Further afield, visitors can plan to visit the spectacular Rocky Mountains or experience pioneer life at Ukrainian Cultural Heritage Village. For more information about Edmonton and the surrounding area please visit www.edmonton.com/for-visitors.aspx and www.travelalberta.com, respectively.

L'Université de l'Alberta vous accueille au 41^e congrès annuel de la Société statistique du Canada. Depuis sa fondation en 1908, l'Université de l'Alberta est une partie intégrante de l'histoire d'Edmonton, la capitale de la province de l'Alberta. N'ayant accueilli que 37 étudiants à son ouverture, l'Université compte maintenant à son actif 5 campus, 18 facultés, 400 programmes et plus de 39 000 étudiants. Elle est actuellement l'une des universités publiques les plus actives en recherche et excelle dans de nombreux domaines, dont les sciences, le commerce, le génie et la médecine. Perché au sommet de la vallée de la rivière Saskatchewan Nord, le campus Nord (campus principal) inclut plus de 150 édifices qui s'étendent sur plus d'une cinquantaine de pâtés de maisons. Des développements récents ont transformé la ligne d'horizon du campus et fourni des installations à la fine pointe de l'enseignement et de la recherche. Le pavillon Universiade (Butterdome) offre des infrastructures sportives de pointe dans un édifice emblématique. Les visiteurs pourront également se balader sur la place principale ou profiter des collections des musées de l'Université.

Située au cœur de la région de la capitale de l'Alberta, Edmonton est une ville énergique de plus de 800 000 habitants. Ville canadienne des festivals, elle est réputée pour ses festivals de culture, d'art, de musique, de théâtre et de sports. Les visiteurs y découvriront toujours du neuf : magasins, sports, arts, restaurants et divertissements, sans oublier ses attractions et événements de réputation mondiale. Citons notamment le parc Fort Edmonton, l'Édifice de l'Assemblée législative de l'Alberta, le Musée des beaux-arts de l'Alberta et le jardin botanique Muttart. La vallée fluviale d'Edmonton, qui constitue le plus grand parc urbain d'Amérique du nord, est équipée de plus de 200 kilomètres de sentiers qui feront la joie des amateurs de plein air. Si vous aimez magasiner, vous adorerez le quartier Old Strathcona et le célèbre West Edmonton Mall. Plus loin, vous pourrez explorer les paysages spectaculaires qu'offrent les Rocheuses ou goûter à la vie des pionniers au Village du patrimoine culturel ukrainien. Pour plus d'informations à propos d'Edmonton et ses environs, consultez www.edmonton.com/french.aspx et www.tourismealberta.ca, respectivement.

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 :

- Alberta Innovates - Health Solutions
- Atlantic Association for Research in the Mathematical Sciences
- Canadian Institutes of Health Research/Instituts de recherche en santé du Canada
- Centre de recherches mathématiques
- Department of Mathematical and Statistical Sciences, University of Alberta
- Edmonton Tourism
- Faculty of Arts, University of Alberta
- Faculty of Arts and Science, Grant MacEwan University
- Faculty of Medicine & Dentistry, University of Alberta
- Faculty of Science, University of Alberta
- Festival City
- Fields Institute for Research in Mathematical Sciences
- Pacific Institute for the Mathematical Sciences
- School of Public Health, University of Alberta
- University of Alberta
- Women & Children's Health Research Institute

Exhibitors • Exposants

The exhibitors, providing displays for examination and purchase, can be found in the Katz Group Centre Atrium. The exhibitors will be available Monday, May 27 and Tuesday, May 28 from 9:00 am to 4:30 pm.

Les kiosques des exposants seront situés dans l'Atrium du Centre Katz Group. Les exposants seront présents lundi 27 mai et mardi 28 mai de 9h00 à 16h30.

- CRC Press - Taylor and Francis Group LLC
- Nelson Education
- W. H. Freeman
- Wiley

Organizers • Organisateurs

Local Arrangements Committee • Comité des arrangements locaux

Rhonda Rosychuk (Chair • Présidente)

Nicola DiNicola, Giseon Heo, Lihui Liu, Ivan Mizera, N.G.N. Prasad, Barb Robinson, A. (Sentil) Senthilselvan, Wanhua Su, Yutaka Yasui, Yan Yuan

The local arrangements committee is indebted to a strong team of dedicated local volunteers across faculties and institutions. We especially thank the graduate student volunteers who have been active in many areas including AV support and registration desk activities. We have benefited from an experienced pair of local conference organizers, Barb Robinson and Cathy Hurst from University of Alberta Conference Services. At the SSC Office, Sarah Watson and Alan Kelm have provided crucial support for many aspects, including registration and exhibitors. The 2012 local arrangements chair, Brian Allen, and previous chairs have generously shared their experiences and have willingly answered 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. Our committee is sincerely grateful for all the direct and indirect efforts made by individuals and units to enable this event.

Le comité des arrangements locaux est très reconnaissant à la solide équipe de bénévoles locaux dévoués, de 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. Nous avons profité de l'expérience de deux organisatrices hors pair, Barb Robinson et Cathy Hurst des Services de conférence de l'Université de l'Alberta. Au bureau de la SSC, Sarah Watson et Alan Kelm ont fourni un appui essentiel pour divers aspects, dont l'inscription et les exposants. Brian Allen, président des arrangements locaux du congrès 2012, et d'autres anciens présidents 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. Notre comité est sincèrement reconnaissant envers tous ces individus et organismes pour tous leurs efforts, directs et indirects, qui ont permis l'organisation de cet événement.

Program Committee • Comité du programme

Chair • Présidente: Debbie J. Dupuis, *HEC Montréal*

Members • Membres:

Actuarial Science Section • Groupe de science actuarielle

Gordon Willmott, *University of Waterloo*

Biostatistics Section • Groupe de biostatistique

Russell Steele, *McGill University*

Business and Industrial Statistics Section • Groupe de statistique industrielle et de gestion





Mu Zhu, *University of Waterloo*





Statistical Education Section • Groupe d'éducation en statistiqueKevin Thorpe, *University of Toronto***Probability Section • Groupe de probabilité**Barbara Szyszkowicz, *Carleton University***Survey Methods Section • Groupe de méthodologie d'enquête**Abdel El Shaarawi, *National Water Research Institute***Accreditation Committee • Comité d'accréditation**John Amrhein, *McDougall Scientific*

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.

Un élément important de nos congrès est la présentation des résumés et des supports visuels des conférences plénières dans les deux langues officielles. La traduction a été exécutée de manière très compétente sous la supervision du comité du bilinguisme.

Language of Presentations • Langue des présentations

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 , , whereas ,  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 ,  indiquent le support visuel. La lettre à l'intérieur identifie la langue : F pour français et E pour anglais (English). Veuillez noter que le support visuel des conférences plénières sera présenté dans les deux langues.

General Information • Information générale

Registration • Inscription

Registration will be available in the Katz Group Centre Atrium according to the following schedule:

- Sunday, May 26: 8:00 am to 7:30 pm
- Monday, May 27: 7:30 am to 5:00 pm
- Tuesday, May 28: 8:15 am to 5:00 pm
- Wednesday, May 29: 8:15 am to 4:00 pm

Vous pourrez vous inscrire dans l'Atrium du Centre Katz Group, en respectant l'horaire suivant :

- Dimanche 26 mai : 8h00 à 19h30
- Lundi 27 mai : 7h30 à 17h00
- Mardi 28 mai : 8h15 à 17h00
- Mercredi 29 mai : 8h15 à 16h00

On-campus accommodation • Logement sur le campus

Those staying in campus residences may check in at the Lister Conference Centre. The desk will be open 24 hours per day and may be reached by phone at 780 492-6056 or by email at guest.services@ualberta.ca.

Les participants qui logent en résidence universitaire peuvent s'enregistrer au centre de conférences Lister. Ce bureau sera ouvert 24 heures sur 24 ; vous pouvez le joindre par téléphone au 780 492-6056 ou par courriel : guest.services@ualberta.ca.

Directions • Emplacement

Campus maps are available in the conference bag and in this program. Those staying at the Lister Conference Centre or the Campus Tower Suite Hotel will have a short walk to the session venues. Those staying at The Coast Edmonton Plaza Hotel and The Fairmont Hotel Macdonald will likely find Edmonton's Light Rapid Transit (LRT) to be an easy way to travel between the University of Alberta and downtown Edmonton.

Des cartes du campus sont disponibles dans le sac du congressiste et dans ce programme. Les délégués séjournant au centre de conférences Lister ou à l'hôtel Campus Tower Suite Hotel sont à quelques minutes de marche du lieu du congrès. Ceux qui séjournent à l'hôtel Coast Edmonton Plaza ou au Fairmont Macdonald pourront venir à l'université par le TLR (train léger sur rail) d'Edmonton.

Parking on Campus • Stationnement sur le campus

Parking permits are required to park on campus. Daily parking permits cost \$15 per day (\$6 on weekends) at Education Car Park and \$14 per day (\$6 on weekends) for Jubilee Car Park and Stadium Car Park. Parking permits can be purchased at the entrance either by cash or credit/debit cards and must be displayed on the dashboard. These parking places are the closest to buildings where most sessions will be held. For those staying in the Lister Conference Centre with a Hotel Room (guest room) booking, parking passes are included in the accommodation rates. Guests staying in dorm rooms have to pay \$14 per day for parking. Please see www.asinfo.ualberta.ca/ParkingServices.aspx for more information about parking on campus.

Un permis de stationnement est obligatoire pour stationner sur le campus. Un permis quotidien coûte 15 \$ par jour (6 \$ en fin de semaine) pour le stationnement Education et 14 \$ par jour (6 \$ en fin de semaine) pour les stationnements Jubilee et Stadium. Vous pouvez vous procurer un permis quotidien à l'entrée en espèces ou par carte de crédit/débit ; vous devez le placer bien en évidence sur votre tableau de bord. Ces stationnements sont situés près des édifices où se tiennent la plupart des séances. Pour les délégués logeant au centre de conférences Lister avec une réservation de type hôtel (chambre d'hôtes), le prix de la chambre inclut un permis de stationnement. Les délégués logeant en résidence universitaire doivent payer 14 \$ par jour pour stationner. Pour plus d'information sur le stationnement, veuillez consulter www.asinfo.ualberta.ca/ParkingServices.aspx.

Campus Security • Sécurité sur le campus

Fire/Police/Ambulance: 911

Protective Services: 780 492-5050 (complaints/information, 24 hrs)

Building Emergency (flood, outage, etc.): 780 492-5555

Edmonton Police Service: 780 423-4567

Pompiers/Police/Ambulance : 911

Services de protection : 780 492-5050 (plaintes/information, 24 heures sur 24)

Secours d'immeuble (inondation, panne de courant, etc.) : 780 492-5555

Service de police d'Edmonton : 780 423-4567

Rooms • Salles

Scientific sessions will occur in Education Centre South (EdSouth, ED), Edmonton Clinic Health Academy (ECHA), Katz Group Centre (KATZ), and Medical Sciences (MedSci, MS). Coffee breaks, exhibits and posters are found in the KATZ Atrium. The Welcome Reception is held in the Centennial Centre for Interdisciplinary Science (CCIS). Most noontime business meetings will occur in ECHA and KATZ, with lunch for business meeting participants delivered to the meeting room. The Monday evening barbeque will occur at Room at the Top (RATT) in the Students' Union Building (SUB). For the Tuesday evening banquet at Fort Edmonton Park, bus transportation will be available.

Les séances scientifiques se dérouleront au Education Centre South (EdSouth, ED), à l'Edmonton Clinic Health Academy (ECHA), au centre Katz Group (KATZ), et au Medical Sciences Centre (MedSci, MS). Vous trouverez les pauses-café, les exposants et les affiches dans l'atrium du centre KATZ. La réception de bienvenue se tiendra au Centennial Centre for Interdisciplinary Science (CCIS). La plupart des réunions d'affaires de midi se dérouleront à l'ECHA et au KATZ, avec un repas prévu pour les participants aux réunions d'affaires servi dans la salle de réunion. Lundi soir, le barbecue se tiendra à Room at the Top (RATT) du Students' Union Building (SUB). Pour le banquet du mardi soir au parc Fort Edmonton, des navettes seront disponibles.

Internet Access • Accès internet

Guest@UofA is the SSID of a guest wireless internet service across the campus (www.aict.ualberta.ca/cnc/wifi#guest). Access is free, but is only granted once you have read and agreed to the conditions of use. The connection allows for simple web surfing (http and https) over standard TCP/IP ports 80 and 443; common VPN clients will also work. The University of Alberta also participates in Eduroam, providing internet access for visiting students and staff from several other partner institutions. To authenticate, all usernames must be qualified by your home institution's domain name. Your username is then your username at your home institution, followed by your university's domain name: `user@domainname.ca`. Public workstations are available at the University of Alberta Libraries and a temporary computing ID may be required (check with a Service Desk). Printing and copying are available at SUBprint in the Students' Union Building.

Guest@UofA est le SSID du réseau Internet sans fil public sur le campus (www.aict.ualberta.ca/cnc/wifi#guest). L'accès est gratuit une fois que vous avez lu et accepté les conditions d'utilisation. La connexion permet une simple navigation (http et https) via les ports TCP/IP standard 80 et 443; les clients VPN habituels fonctionnent aussi. L'Université de l'Alberta fait également partie du réseau Eduroam, qui offre l'accès Internet aux étudiants et au personnel des institutions partenaires. Pour vous authentifier, vous devez posséder un nom d'utilisateur autorisé sur le domaine de votre institution. Votre nom d'utilisateur est alors votre nom d'utilisateur à votre institution, suivi du nom de domaine de votre université : `utilisateur@nomdedomaine.ca`. Vous trouverez aussi des postes de travail publics dans les bibliothèques de l'université, pour lesquels vous aurez besoin d'un ID temporaire (vérifiez avec le service d'assistance). Un service d'impression et de copie est disponible chez SUBprint dans le Students' Union Building.

Smoking Prohibited • Interdiction de fumer

Smoking is prohibited in all campus buildings and within five metres of building entrances and exits, fresh-air intakes and compressed-gas storage areas.

Il est interdit de fumer dans les édifices du campus ainsi que dans un rayon de cinq mètres des entrées et sorties des édifices, des prises d'air frais et des zones d'entreposage de gaz comprimé.

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

There are multiple food options at the University of Alberta and in nearby neighbourhoods within walking distance. Close to the sessions, the Katz Group Centre (Tim Hortons), Edmonton Clinic Health Academy (Starbucks, Engrained), Research Transition Facility (Stir Crazy), Education Centre (EuroMarket), Students' Union Building (various), University of Alberta Hospital (food court), and Lister Centre all have options to suit different tastes. Food trucks will likely be positioned at ECHA on Monday to Wednesday. HUB International provides many choices as well. There are multiple restaurants along 112 Street between 87 Avenue and Whyte (82) Avenue (e.g., Earls, Hudsons, Wendy's). These restaurants are the best options for food on weekends. Sobeys is a grocery located in College Plaza on 112 Street and 83 Avenue. Restaurants along the 109 Street and Whyte Avenue corridors provide additional options. Downtown Edmonton features many culinary delights, and a LRT ride from the North Campus makes these restaurants easily accessible. The campus map (www.campusmap.ualberta.ca), web restaurant list (www.ssc.ca/en/meetings/2013), and other guides (e.g., www.urbanspoon.com/c/131/Edmonton-restaurants.html) provide further information.

Vous avez le choix des restaurants à l'Université de l'Alberta et dans les quartiers avoisinants. Près des séances, le Centre Katz Group (Tim Hortons), l'Edmonton Clinic Health Academy (Starbucks, Engrained), le Research Transition Facility (Stir Crazy), l'Education Centre (EuroMarket), le Students' Union Building (divers), l'hôpital de l'Université de l'Alberta (aire de restauration), et le Lister Centre (divers) proposent de quoi satisfaire tous les goûts. Des camions de restauration rapide stationneront vraisemblablement devant l'ECHA du lundi au mercredi. HUB International propose de nombreux choix aussi. Vous trouverez de nombreux restaurants le long de la 112^e rue entre la 87^e Avenue et Whyte (82^e) Avenue (Earls, Hudsons, Wendy's, etc.). La fin de semaine, ces restaurants représentent la meilleure option. Un supermarché Sobeys est situé au centre commercial College au coin de la 112^e rue et de la 83^e Avenue. Il existe aussi divers restaurants le long de la 109^e rue et de Whyte Avenue. Le centre-ville d'Edmonton, qui regorge de restaurants, est facilement accessible par le TLR depuis le campus Nord. Pour plus de renseignements, consultez le plan du campus (www.campusmap.ualberta.ca), la liste des restaurants (www.ssc.ca/fr/congr/%C3%A8s/2013) et d'autres guides (p. ex., www.urbanspoon.com/c/131/Edmonton-restaurants.html).

Use of Athletics Facilities • Utilisation des installations sportives

The University of Alberta has several athletic facilities. The Van Vliet Centre and the Universaide Pavilion are located on the north campus and offer a variety of activities (e.g., pools, gymnasiums, courts, fitness centre). The drop-in passes (\$8.50, photo ID required) can be obtained at Equipment Room W-78, Van Vliet Centre lower level of West Wing (780 492-3276, www.physedandrec.ualberta.ca/Facilities/EquipmentRoom.aspx), as well as the information about pertinent schedules and accessibility (see also www.physedandrec.ualberta.ca/Facilities.aspx). The City of Edmonton also has multiple athletic facilities, including the Kinsmen Sports Centre (located in Kinsmen Park not far away from North Campus).

L'Université de l'Alberta propose plusieurs installations sportives. Le centre Van Vliet et le pavillon Universaide, qui sont situés sur le campus Nord, abritent diverses activités (piscine, gymnases, terrains, centre de conditionnement). Vous pouvez vous procurer une carte d'utilisateur (8,50 \$, pièce d'identité avec photo exigée) à la salle d'équipement W-78, centre Van Vliet, niveau inférieur du West Wing (780 492-3276, www.physedandrec.ualberta.ca/

Facilities/EquipmentRoom.aspx), où vous trouverez plus d'informations sur les horaires d'ouverture et l'accessibilité (voir aussi www.physedandrec.ualberta.ca/Facilities.aspx). La ville d'Edmonton a de nombreuses autres installations sportives, dont le centre sportif Kinsmen (situé dans le parc Kinsmen près du campus Nord).

Workshops • Ateliers

Sunday, May 26, 9:00 - 4:00 pm • Dimanche 26 mai, 9h00 - 16h00 Edmonton Clinic Health Academy

Workshops organized by the sections will be held in the Edmonton Clinic Health Academy (ECHA), Sunday 9:00 am to 4:00 pm. The Biostatistics Section workshop will be held in L1-150, and the Business and Industrial Section in L1-230 the Probability Section in L1-220. Workshop participants are asked to pick up their registration packages at the registration desk located in Katz Group Centre Atrium prior to attending the workshop.

Les ateliers organisés par les groupes se tiendront à l'Edmonton Clinic Health Academy (ECHA), dimanche de 9h à 16h. L'atelier du groupe de biostatistique se tiendra dans la salle L1-150, celui du groupe de statistique industrielle et de gestion dans la salle L1-230 et celui du groupe de probabilité dans la salle L1-220. Les participants sont priés de récupérer leur trousse d'inscription au bureau des inscriptions, situé dans l'atrium du centre Katz Group avant les ateliers.

Afternoon Excursions • Excursions en après-midi

Sunday, May 26, 12:30 pm • Dimanche 26 mai, 12h30 Katz Group Centre Atrium

Those who have purchased a ticket for transportation to Fort Edmonton Park, the Muttart Conservatory, the Art Gallery of Alberta, or a river valley walk will meet at the registration desk 12:15 pm for a 12:30 pm departure by shuttle for the various attractions. The bus will be parked on the west side of Edmonton Clinic Health Academy. The return times from the attractions to the University of Alberta will be posted at the registration desk. Please show your ticket to the shuttle driver and note the return pick up times from your attraction. If you wish to depart the attraction earlier or later, please make your own transportation plans. The ticket includes transportation only and participants will have to pay for their own entry fee and applicable taxes for the attraction.

Les délégués qui ont acheté un billet de transport pour les excursions au parc Fort Edmonton, au jardin botanique Muttart, au Musée des beaux-arts de l'Alberta ou à la promenade le long de la vallée fluviale sont priés de rejoindre le bureau des inscriptions à 12h15 pour un départ en navette à 12h30. Le bus stationnera à l'entrée ouest de l'Edmonton Clinic Health Academy. Les horaires de retour à l'Université de l'Alberta seront affichés au bureau des inscriptions. Veuillez présenter votre billet au chauffeur de la navette et noter l'horaire de retour depuis l'attraction de votre choix. Si vous souhaitez quitter l'attraction plus tôt ou plus tard, veuillez organiser votre propre transport. Le billet inclut uniquement le transport ; les participants devront régler les frais d'entrée (et taxes) aux attractions sur place.

Information on 2013 NSERC Discovery Grant Competition • Information sur le concours 2013 des subventions à la découverte du CRSNG

Sunday, May 26, 4:00 - 6:00 pm • Dimanche 26 mai, 16h00 - 18h00 Edmonton Clinic Health Academy, room/local L1-140

A representative from NSERC will make a presentation to inform researchers on the peer review process for NSERC Discovery Grants. Summary statistics on the 2013 competition will be provided. All researchers are welcome to attend.

Un représentant du CRSNG donnera une présentation afin d'informer les chercheurs sur le processus d'évaluation par les pairs des subventions à la découverte du CRSNG. Des statistiques résumant le concours de 2013 seront fournies. Tous les chercheurs sont les bienvenus.

Public Lecture • Conférence grand public

Sunday, May 26, 4:00 pm • Dimanche 26 mai, 16h00 Engineering Teaching and Learning Complex room/local E1003

To celebrate the International Year of Statistics, the SSC is proud to present a Public Lecture: "Solving Crimes Using Math: the Lottery Retailer Scandal", by Jeffrey S. Rosenthal, Professor, Department of Statistics, University of Toronto, and author of *Struck by Lightning: The Curious World of Probabilities*. Join us at the University of Alberta, Engineering Teaching and Learning Complex (ETLC), room E1003, on Sunday May 26th, 2013, at 4:30 pm. Juice and cookies will be served from 4:00 pm.

Pour célébrer l'Année mondiale de la statistique, la SSC est fière de présenter la conférence grand public : «Résoudre des crimes à l'aide des maths : le scandale des revendeurs de loterie», par Jeffrey S. Rosenthal, Professeur, Département de Statistique, University of Toronto, et auteur de *Struck by Lightning : The Curious World of Probabilities*. La conférence (en anglais) aura lieu à l'Université de l'Alberta, au Engineering Teaching and Learning Complex (ETLC), salle E1003, le dimanche 26 mai 2013 à 16h30. Du jus et des biscuits seront servis dès 16h00.

Welcome Reception • Réception de bienvenue

Sunday, May 26, 6:00 - 8:00 pm • Dimanche 26 mai, 18h00 - 20h00 .. Centennial Centre for Interdisciplinary Sciences, South Atrium

All conference participants are cordially invited to attend a welcome reception in the South Atrium of the Centennial Centre for Interdisciplinary Sciences (CCIS), from 6:00 pm to 8:00 pm, on Sunday, May 26. Tickets for two complimentary beverages will be found in your registration envelope. Please pick up your registration package at the Katz Group Centre Atrium before 7:15 pm on Sunday to obtain your tickets. Please note that to conform with Alberta liquor laws, liquor can only be consumed within the licensed area main floor of the CCIS South Atrium.

Les participants au congrès sont tous cordialement invités à une réception de bienvenue dans l'atrium du Centennial Centre for Interdisciplinary Sciences, dimanche 26 mai de 18 heures à 20 heures. Vous trouverez des coupons donnant droit à deux consommations gratuites dans votre enveloppe d'inscription. Veuillez récupérer votre trousse d'inscription dans l'atrium du centre Katz Group avant 19h15 dimanche pour obtenir vos coupons. Veuillez noter qu'en vertu des lois sur les boissons alcooliques en Alberta, celles-ci ne peuvent être consommées que dans l'aire sous licence, au rez-de-chaussée de l'atrium sud du CCIS.

Dinner of the Caucus for Women in Statistics and the Committee on Women in Statistics • Souper du Caucus des femmes en statistique et du comité de promotion de la femme en statistique

Sunday, May 26, 7:30 pm • Dimanche 26 mai, 19h30 Kyoto Japanese Cuisine

The SSC Committee on Women in Statistics and the Caucus for Women in Statistics have organized an informal supper at 7:30 pm, Sunday May 26 at Kyoto Japanese Cuisine (8709 109 Street, www.yelp.ca/biz/kyoto-japanese-cuisine-edmonton-2). Interested women and men are asked to email ying.zhang@acadiiau.ca by May 19. This event is not included in the registration fee. See you there!

Le comité de promotion de la femme en statistique de la SSC et le Caucus des femmes en statistique organisent un dîner informel dimanche 26 mai 2013 à 19h30 au Kyoto Japanese Cuisine (8709 109 Rue, www.yelp.ca/biz/kyoto-japanese-cuisine-edmonton-2). Toutes les intéressées et tous les intéressés sont priés d'envoyer un courriel à ying.zhang@acadiu.ca avant le 19 mai. Cet événement n'est pas inclus dans les frais d'inscription. En espérant vous y voir !

Poster Sessions • Séances d'affichage

All poster sessions will take place in the Katz Group Centre Atrium. Case Studies I and II, along with the Poster Session, will take place on Monday, May 27 from 1:30 pm to 3:30 pm. 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.

Les séances d'affichage se tiendront dans l'atrium du centre Katz Group. Les études de cas I et II, ainsi que la séance d'affiches, se tiendront lundi 27 mai de 13h30 à 15h30. Les auteurs seront présents jusqu'à la fin de la pause café à 15h30. Les présentateurs trouveront plus d'instructions sur le site Web du congrès.

uWaterloo Math Alumni Lunch • Dîner des anciens étudiants en maths de la uWaterloo

Monday, May 27, 12:00 - 1:30 pm • Lundi 27 mai, 12h00 - 13h30 Hudsons Canadian Tap House (Campus: 11113 87 Avenue)

Waterloo alumni attending the SSC or living in the Edmonton area are invited to a complimentary social lunch! Enjoy lunch and a couple beverages while reconnecting with fellow alumni. Professor Steve Brown will share exciting updates from the Faculty of Math and from the university campus. Guests welcome. Remember to bring your business card for a chance to win great door prizes! RSVP to Kate Jessop, Senior Alumni Officer Math by emailing kjessop@uwaterloo.ca. See you there.

Les anciens étudiants de Waterloo qui assistent au congrès annuel de la SSC ou qui vivent dans la région d'Edmonton sont invités à un déjeuner de confraternité offert! Profitez de cette occasion pour renouer avec vos anciens camarades. Le professeur Steve Brown vous présentera les dernières nouvelles de la Faculté de maths et du campus. Les invités sont les bienvenus. N'oubliez pas votre carte d'affaires pour participer à un tirage au sort avec des prix de présence spéciaux! Veuillez confirmer votre présence à Kate Jessop, chargée des relations avec les anciens étudiants en mathématiques : kjessop@uwaterloo.ca. Nous espérons vous voir nombreux.

Barbeque • Barbecue

Monday, May 27, 6:00 - 8:00 pm • Lundi 27 mai, 18h00 - 20h00 Room at the Top (RATT)

A barbeque will be held at RATT (located on the 7th floor of the Students' Union Building) from 6:00 pm until 8:00 pm, Monday, May 27. All graduate students who registered, and anyone else who has paid for this meal, will find a ticket in their registration envelope. This event is sold out so if you have a ticket that you do not plan to use, please return your ticket to the registration desk by 2:00 pm, Monday, May 27.

Un barbecue est organisé au RATT (situé au 7^e étage du Students' Union Building), lundi 27 mai de 18 heures à 20 heures. Les étudiants gradués ainsi que les autres participants qui ont acheté des billets pour cet événement les trouveront dans leur enveloppe d'inscription. L'événement affiche complet ; si vous avez un billet que vous ne pensez pas utiliser, veuillez le rendre au bureau des inscriptions avant lundi 27 mai à 14h.

Old Strathcona Restaurant Night • Soirée restaurant dans le quartier Old Strathcona

Monday, May 27, 6:15 - 9:00 pm • Lundi 27 mai, 18h15 - 21h00Katz Group Centre Atrium

Small groups will go to restaurants in the Old Strathcona area (2 to 2.8 km from North Campus). Six diverse restaurants have been chosen: The restaurants are: Murrieta's, Chianti Cafe, Julio's Barrio, Yiannis Taverna, Packrat Louie, and The Pourhouse Bier Bistro. Participants will choose one restaurant and sign up sheets will be available at the registration desk. Participants will be required to sign up by 2:00 pm Monday, May 27 to guarantee participation. Reservations will be limited to 10 per restaurant. Participants can choose to gather at the registration desk at 6:10 pm for a guided walk to the restaurant or make their own way to the restaurant for a 7:00 pm reservation. Participants will pay for their own meals which are subject to an automatic 18% gratuity.

De petits groupes se rendront dans divers restaurants du quartier Old Strathcona (2 à 2,8 km du campus Nord). Six restaurants ont été retenus : Murrieta's, Chianti Cafe, Julio's Barrio, Yiannis Taverna, Packrat Louie et le Pourhouse Bier Bistro. Des feuilles d'inscription seront disponibles au bureau d'inscription au congrès ; les participants devront y indiquer le restaurant de leur choix avant lundi 27 mai à 14h pour garantir leur participation. Les réservations seront limitées à 10 par restaurant. Les participants pourront se réunir au bureau des inscriptions à 18h10 pour une promenade guidée vers leur restaurant ou s'y retrouver à 19h. Les participants régleront leur propre repas, ainsi qu'un pourboire de 18 % automatique.

New Investigator Dinner • Souper des nouveaux chercheurs

Monday, May 27, 6:15 - 9:00 pm • Lundi 27 mai, 18h15 - 21h00 The Pourhouse Bier Bistro

The SSC New Investigators Committee organizes a causal get-together event for new researchers, as a part of the Old Strathcona Restaurant Night. If you are a new investigator, and want to meet other new investigators at the 41st Annual Meeting of the SSC, please indicate your preference when signing-up for the Old Strathcona Restaurant Night at the registration desk (choose The Pourhouse Bier Bistro, 10354 82 Avenue). Participants can choose to gather at the registration desk at 6:10 pm for a guided walk to the restaurant or make their own way to the restaurant for a 7:00 pm reservation. Participants will pay for their own meals which are subject to an automatic 18% gratuity.

Le comité des nouveaux chercheurs de la SSC organise une rencontre informelle pour les nouveaux chercheurs, dans le cadre de la soirée restaurant dans le Old Strathcona. Si vous êtes nouveau chercheur et souhaitez rencontrer d'autres nouveaux chercheurs pendant le 41^e congrès annuel de la SSC, veuillez indiquer votre préférence lors de l'inscription à la soirée restaurant au bureau des inscriptions (choisissez le Pourhouse Bier Bistro, 10354 82^e Avenue). Les participants pourront se réunir au bureau des inscriptions à 18h10 pour une promenade guidée vers leur restaurant ou s'y retrouver à 19h. Les participants régleront leur propre repas, ainsi qu'un pourboire de 18 % automatique.

Student Career Session • Séance carrières pour les étudiants

Tuesday, May 28, 12:00 - 1:30 pm • Mardi 28 mai, 12h00 - 13h30 ...Lister Centre, Aurora Room/Salle Aurora

The Biostatistics Section have organized a career session for all graduate students that focuses on skills needed to find, get and succeed in your first job. Following brief presentations from three senior biostatisticians employed in a variety of settings, faculty members will lead roundtable discussions for small groups of students on strategies to develop these skills while in graduate school. Participants were to have registered online and have received a ticket for the event with their registration package.

Le groupe de biostatistique a organisé une séance carrières pour les étudiants de cycle supérieur afin d'explorer les compétences requises pour décrocher et réussir à votre premier poste. Après de brèves présentations par trois biostatisticiens chevronnés qui travaillent dans divers environnements, des professeurs animeront des discussions en petits groupes d'étudiants sur la manière de développer ces compétences pendant ses études. Les participants devaient s'inscrire en ligne ; ils auront reçu un billet électronique pour cet événement.

NSERC Discovery Grant Application Workshop • Ateliers sur les demandes de subventions à la découverte du CRSNG

Tuesday, May 28, 5:00 - 6:00 pm • Mardi 28 mai, 17h00 - 18h00 Katz Group Centre, room/local 1-080

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.

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.

Banquet

Tuesday, May 28, 6:15 pm • Mardi 28 mai, 18h15 Fort Edmonton Park, Blatchford Hangar

The banquet will be held in the Blatchford Hangar at historic Fort Edmonton Park, with cocktails and a cash bar at 6:15 pm and dinner service beginning at 7:05 pm. All conference participants who have not declined the banquet will find a banquet ticket in their registration envelope. You are asked to place this on the table in front of you. If you have a ticket but do not wish to attend the banquet, please return your ticket to the registration desk by noon on Tuesday, May 28. Buses are arranged to transport participants from the University of Alberta (west side of Edmonton Clinic Health Academy) from 6:00 pm to 6:30 pm to Fort Edmonton (last bus at 6:30 pm) and return participants after the banquet (10:00 pm to 10:30 pm). Participants will be dropped off at the main gate and take a short walk to the Blatchford Hangar. We recommend that all participants use the bus service, and we ask that you show your banquet ticket prior to boarding the bus. Parking is also available at the general Fort Edmonton parking lot. If you have special needs or considerations that prevent use of the bus service or require delivery right to the Blatchford Hangar, please contact the Local Arrangements Chair (ssc2013@ssc.ca) for alternatives by May 15.

Le banquet se tiendra au hangar Blatchford dans le parc historique de Fort Edmonton, avec des cocktails et un bar payant à 18h15 et le souper à partir de 19h05. Les participants au congrès qui n'ont pas décliné l'offre trouveront un billet pour le banquet dans leur enveloppe d'inscription. Vous êtes prié de le placer sur la table devant vous. Si vous avez un billet et que vous ne souhaitez pas participer au banquet, veuillez le rendre au bureau des inscriptions avant mardi 28 mai à midi. Des bus sont organisés pour transporter les participants de l'Université de l'Alberta (entrée ouest de l'Edmonton Clinic Health Academy) entre 18h et 18h30 vers Fort Edmonton (dernier bus à 18h30), avec un retour après le banquet (de 22h à 22h30). Les participants seront déposés au portail principal et auront une courte marche à pied pour se rendre au hangar Blatchford. Nous recommandons à tous les participants de se prévaloir du transport et vous demandons de montrer votre billet lors de la montée dans le bus. Vous pouvez également vous garer au stationnement principal du parc Fort Edmonton. En cas de besoin spécial, si vous ne pouvez pas prendre le bus ou si vous devez être déposé au hangar Blatchford, veuillez contacter le président des arrangements locaux (ssc2013@ssc.ca) avant le 15 mai pour organiser une solution.

Statistics Coffee Time - A CIHR Funded Collaboration Event • Pause café Statistique - une collaboration financée par les IRSC

**Wednesday, May 29, 9:30 - 10:00 am • Mercredi 29 mai, 9h30 - 10h00 Katz Group Centre, 2nd floor
Atrium/Atrium du 2nd étage**

With funding received from a Canadian Institutes of Health Research Dissemination Events grant, a Statistics Coffee Time has been created. In an informal setting, participating health researchers and statistics and probability experts will socialize and discuss potential collaborations for research projects. Participants were to have registered online and have received an electronic ticket for the event.

Nous avons obtenu une bourse des Activités de dissémination des Instituts de recherche en santé du Canada pour créer une Pause café Statistique. Dans un cadre informel, chercheurs médicaux et experts en statistique et probabilité pourront se rencontrer et discuter de leurs projets de recherche et de possibilités de collaboration. Les participants devaient s'inscrire en ligne ; ils auront reçu un billet électronique pour cet événement.

Committees and Meetings • Comités et réunions

Saturday May 25

samedi 25 mai

18:00 - 22:30 Executive Committee 1 / Comité exécutif 1
Campus Tower Suite Hotel, Boardroom

Sunday May 26

dimanche 26 mai

9:30 - 12:00 Department Heads Meeting / Rencontre des directeurs de départements
ECHA 3-001

10:30 - 12:00 Publications Committee / Comité des publications
ECHA 1-131

12:00 - 16:00 Board of Directors 1 / Conseil d'administration 1
ECHA 4-036

Monday May 27

lundi 27 mai

12:00 - 13:30 Business and Industrial Statistics Section Executive Committee / Comité exécutif du groupe de statistique industrielle et de gestion
KATZ 2-153

12:00 - 13:30 Committee on SSC Bylaws / Comité de statuts de la SSC
ECHA 1-441

12:00 - 13:30 Biostatistics Section Executive Committee 1 / Comité exécutif de biostatistique industrielle et de gestion 1
KATZ 2-152

12:00 - 13:30 Probability Section Executive Committee / Comité exécutif du groupe de probabilité
KATZ 2-151

12:00 - 13:30 Statistical Education Committee / Comité d'éducation en statistique
KATZ 2-157

12:00 - 13:30 Statistical Education Section Executive Committee / Comité exécutif du groupe d'éducation en statistique
KATZ 2-157

12:00 - 13:30 Strategic Planning Committee / Comité de la planification stratégique
ECHA 1-447

12:00 - 13:30 Survey Methods Section Executive Committee / Comité exécutif du groupe de méthodes d'enquête
KATZ 2-154

12:00 - 13:30 Committee on Women in Statistics / Comité de promotion de la femme en statistique
ECHA 2-407

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
KATZ 2-155

17:00 - 18:00 SSC Annual General Meeting / Assemblée générale annuelle de la SSC
KATZ 1-080

Tuesday May 28

mardi 28 mai

12:00 - 13:30 CJS Editorial Board / Comité de rédaction de la RCS
ECHA 1-447

12:00 - 13:30 Bilingualism Committee / Comité du bilinguisme
ECHA 2-407

12:00 - 13:30	Census at School Committee / Comité de recensement à l'école <i>ECHA 1-131</i>
12:00 - 13:30	Office Committee / Comité du bureau <i>ECHA 1-152</i>
12:00 - 13:30	Public Relations Committee / Comité des relations publiques <i>ECHA 1-144</i>
12:00 - 13:30	Research Committee / Comité de recherche <i>ECHA 1-148</i>
12:00 - 13:30	Accreditation General Meeting / Assemblée générale des membres accrédités <i>ECHA 1-441</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>KATZ 2-155</i>
17:00 - 18:00	Business and Industrial Statistics Section General Meeting / Groupe de statistique industrielle et de gestion assemblée générale <i>Lister Centre, Evergreen room / salle Evergreen</i>
17:00 - 18:00	Student Research Presentation Award Committee / Comité du prix pour la présentation de recherche étudiante <i>ECHA 2-407</i>
17:00 - 18:00	Actuarial Science Section General Meeting / Groupe d'actuariat assemblée générale <i>ECHA L1-230</i>
17:00 - 18:00	Biostatistics Section General Meeting / Groupe de biostatistique assemblée générale <i>ECHA L1-140</i>
17:00 - 18:00	Probability Section General Meeting / Groupe de probabilité assemblée générale <i>ECHA 1-498</i>
17:00 - 18:00	Statistical Education Section General Meeting / Groupe d'éducation en statistique assemblée générale <i>ECHA L1-150</i>
17:00 - 18:00	Survey Methods Section General Meeting / Groupe de méthodes d'enquête assemblée générale <i>ECHA L1-220</i>

Wednesday May 29**mercredi 29 mai**

11:30 - 13:00	Accreditation Committee / Comité d'accréditation <i>ECHA 1-447</i>
11:30 - 13:00	Biostatistics Section Executive Committee 2 / Comité exécutif de biostatistique industrielle et de gestion 2 <i>Katz 2-156</i>
11:30 - 13:00	Program Committee / Comité du programme <i>ECHA 1-441</i>
11:30 - 13:00	Treasurers / Trésoriers <i>ECHA 2-407</i>
16:15 - 17:30	Board of Directors 2 / Conseil d'administration 2 <i>ECHA 4-036</i>
17:30 - 18:30	Executive Committee 2 / Comité exécutif 2 <i>ECHA 4-036</i>

Social and Information Events • Événements sociaux et informatifs

Sunday May 26

dimanche 26 mai

12:15	Afternoon Excursion - Fee Event / Excursion en après-midi - activité payante <i>Katz Group Centre Atrium</i>
16:00 - 18:00	Information on 2013 NSERC Discovery Grant Competition / Information sur le concours 2013 des subventions à la découverte du CRSNG <i>ECHA L1-140</i>
18:00 - 20:00	Welcome Reception / Réception de bienvenue <i>Centennial Centre for Interdisciplinary Science</i>
19:30 - 21:30	Dinner of the Caucus for Women in Statistics and the SSC Committee on Women in Statistics / Souper informel du Caucus for Women in Statistics et du comité de promotion de la femme en statistique de la SSC <i>Kyoto Japanese Cuisine (8709 109 Street / rue)</i>

Monday May 27

lundi 27 mai

9:45 - 10:30	Coffee Break / Pause-café <i>Katz Group Centre Atrium</i>
12:00 - 13:30	uWaterloo Math Alumni Lunch / Dîner des anciens étudiants en math de la uWaterloo <i>Hudsons Canadian Tap House (Campus)</i>
15:00 - 15:30	Coffee Break / Pause-café <i>Katz Group Centre Atrium</i>
18:00 - 20:00	Student BBQ - Free for graduate students / Barbecue des étudiants - Gratuit pour les étudiants de cycle supérieur <i>Room at the Top (RATT)</i>
18:15 - 21:00	New Investigator Dinner - Fee Event / Souper des nouveaux chercheurs - événement payant <i>Katz Group Centre Atrium</i>
18:15 - 21:00	Old Strathcona Restaurant Night - Fee Event / Soirée restaurant dans le quartier Old Strathcona - événement payant <i>Katz Group Centre Atrium</i>

Tuesday May 28

mardi 28 mai

9:45 - 10:30	Coffee Break / Pause-café <i>Katz Group Centre Atrium</i>
12:00 - 13:30	Student Career Session / Séance carrières pour les étudiants <i>Lister Centre, Aurora room / salle Aurora</i>
15:00 - 15:30	Coffee Break / Pause-café <i>Katz Group Centre Atrium</i>
17:00 - 18:00	NSERC Discovery Grant Application Workshop / Ateliers sur les demandes de subventions à la découverte du CRSNG <i>Katz Group Centre, room/local 1-080</i>
18:15 - 19:05	Banquet - Cocktails and Cash Bar / Banquet - Bar payant <i>Blatchford Hangar, Fort Edmonton Park / hangar Blatchford, parc Fort Edmonton</i>
19:05 - 22:00	Banquet - Dinner Service / Banquet - Service du repas <i>Blatchford Hangar, Fort Edmonton Park / hangar Blatchford, parc Fort Edmonton</i>

Wednesday May 29

mercredi 29 mai

9:30 - 10:00	Coffee Break / Pause-café <i>Katz Group Centre Atrium</i>
--------------	--

9:30 - 10:00	Statistics Coffee Time / Pause café Statistique <i>Katz Group Centre, second floor Atrium / Katz Group Centre, 2e étage Atrium</i>
14:30 - 15:00	Coffee Break / Pause-café <i>Katz Group Centre Atrium</i>

List of Sessions • Liste des sessions

- 1A-A1 Inaugural Session
Séance inaugurale
- 1A-A2 SSC Presidential Invited Address
Allocution de l'invitée du Président de la SSC
- 1B-V Statistics and the Environment
La statistique et l'environnement
- 1B-I1 Financial Mathematics
Mathématiques financières
- 1B-I2 New Developments in Statistical Neuroscience
Nouveaux développements en statistique de la neuroscience
- 1B-I3 Recent Development on Computer Experiments
Développements récents en expériences informatiques
- 1B-I4 Survey Sampling and Applied Mathematics
Échantillonnage et mathématiques appliquées
- 1B-C1 Bayesian Estimation and Inference
Estimation et inférence bayésiennes
- 1B-C2 Biostatistics - Theory and Methods I
Biostatistique - théorie et méthodes I
- 1B-C3 Clustering and Classification
Partitionnement et classification
- 1C-I1 Dependence Models with Insurance Applications
Modèles de dépendance avec applications en assurance
- 1C-I2 Fitting Complex Models Using Automatic Differentiation Model Builder (ADMB)
Ajustement de modèles complexes en utilisant ADMB
- 1C-I3 High-Dimensional Data Analysis for Brain Imaging
Analyse de données de grande dimension pour l'imagerie encéphalique
- 1C-I4 Identifiability: Pushing Data to the Limits
Identifiabilité : les données poussées à leurs limites
- 1C-I5 Moving Up in Academic Ranks: Challenges and Strategies for Statisticians Involved in Interdisciplinary Research
Gravir les échelons universitaires – défis et stratégies pour les statisticiens impliqués dans la recherche interdisciplinaire
- 1C-I6 Survival Analysis Methods and Applications
Méthodes et applications en analyse de survie
- 1C-C1 Biostatistics - Theory and Methods II
Biostatistique - théorie et méthodes II
- 1C-C2 Business and Industrial Statistics I
Statistique industrielle et de gestion I
- 1C-C3 Environmental Modeling: Applications and Case Studies
Modélisation en environnement : applications et études de cas
- 1C-C4 Probability I
Probabilité I
- 1C-C5 Survey: Theory, Methods and Applications
Échantillonnage : théorie, méthodes et applications
- 1C-C6 Variable Selection and Statistical Learning
Sélection de variables et apprentissage statistique
- 1C-D1 Case Study I: Exploring Spatial and Temporal Heterogeneity of Environmental Noise in Toronto
Étude de cas I : L'exploration de l'hétérogénéité spatiale et temporelle du bruit environnemental à Toronto

- 1C-D2 Case Study II: Study of the Impact of Human Disturbances on Habitat Selection and on Calf Survival for Forest-Dwelling Caribou
Étude de cas II : Étude de l'impact des perturbations causées par l'homme sur la sélection d'habitat et la survie des faons dans l'écosystème forestier du caribou des bois
- 1C-P Poster Session
Séance d'affichage
- 1D-I1 Advances in High-Dimensional Data Analysis
Avancées en matière d'analyse de données de grande dimension
- 1D-I2 Recent Developments in R packages
Développements récents dans les packages R
- 1D-I3 Some Thoughts on How to be More Successful
Quelques pistes pour connaître un plus grand succès
- 1D-I4 Statistics and Ecology
La statistique et l'écologie
- 1D-I5 Stochastic Processes Arising in Environmental Sciences
Processus stochastiques dans les sciences environnementales
- 1D-C1 Analysis of Survival and Clinical Trials Data
Analyse de données de survie et d'essais cliniques
- 1D-C2 Biostatistics - Applications and Case Studies
Biostatistique - applications et études de cas
- 1D-C3 Dependence I: Correlation and Copulas
Dépendance I : corrélation et copules
- 1D-C4 Estimation Methods
Méthodes d'estimation
- 1D-C5 Experimental and Sampling Design
Plan d'expérience et d'échantillonnage
- 2A-A Gold Medal Address
Allocution du récipiendaire de la médaille d'or
- 2B-V Methods to Correct Measures of Effect for Bias Due to Exposure Measurement Error
Méthodes pour corriger la mesure des effets en tenant compte du biais dû aux erreurs de mesure d'exposition
- 2B-I1 Effective Use of Laboratory Sessions in Undergraduate Statistics Teaching
Du bon usage des séances de laboratoire dans les cours de statistique de premier cycle
- 2B-I2 Functional Data Analysis
Analyse de données fonctionnelles
- 2B-I3 Machine Learning and Data Analytics
Apprentissage automatique et analytique de données
- 2B-I4 Stochastic Models in Insurance
Modèles stochastiques en assurance
- 2B-C1 Modelling the Spread of Infectious Diseases
Modélisation de la propagation des maladies infectieuses
- 2B-C2 Robust Estimation
Estimation robuste
- 2B-C3 Statistical Modeling
Modélisation statistique
- 2B-C4 Statistical Theory and Applications I
Statistique : théorie et applications I
- 2C-I1 An Introduction to Relative Survival Analysis
Introduction à l'analyse de survie relative
- 2C-I2 Design and Methodological Issues in Environmental Diseases
Problématiques liées à la méthodologie et aux plans d'expérience en matière de maladies liées à l'environnement
- 2C-I3 Environmental Sampling and Monitoring
Échantillonnage environnemental et surveillance

- 2C-I4 Skew Elliptical Distributions: The State of the Art
Les distributions asymétriques elliptiques : technologie de pointe
- 2C-I5 Statistical Learning
Apprentissage statistique
- 2C-I6 Statistics in Business and Finance
Statistique en gestion et finance
- 2C-C1 Probability II
Probabilité II
- 2C-C2 Statistical Theory and Applications II
Statistique : théorie et applications II
- 2D-A Isobel Loutit Invited Address
Allocution invitée Isobel Loutit
- 2D-I1 Environmental and Ecological Quality Assessment
Évaluation de qualité environnementale et écologique
- 2D-I2 Promoting Statistical Literacy - CensusAtSchool and Beyond
Promotion de la culture statistique - Recensement à l'école et au-delà
- 2D-I3 Recurrent Event Analysis
Analyse d'événements récurrents
- 2D-I4 Stochastic Models with Applications to Environmental Problems
Modèles stochastiques avec applications aux problèmes environnementaux
- 3A-A CRM-SSC Prize in Statistics Address
Allocution du récipiendaire du Prix CRM-SSC
- 3B-V Ranking and Selection in High-Dimensional Inference
Ordonnement et sélection dans l'inférence pour les données de grande dimension
- 3B-I1 Bootstrap and Change-Point Analysis with Applications
Bootstrap et analyse du point de rupture avec applications
- 3B-I2 Methods for Cross-Country and Cross-Cultural Surveys
Méthodes pour les enquêtes transnationales et transculturelles
- 3B-I3 Recommender Systems
Systèmes de recommandation
- 3B-C1 Dependence II
Dépendance II
- 3B-C2 Estimation and Hypothesis Testing
Estimation et tests d'hypothèses
- 3B-C3 Models Accounting for Measurement Error and Meta-Analysis
Modèles avec erreurs de mesure et méta-analyse
- 3C-A SSC Impact Award Address
Allocution du récipiendaire du prix pour impact de la SSC
- 3C-I1 Issues and Challenges Faced by New Researchers
Problèmes et défis auxquels font face les nouveaux chercheurs
- 3C-I2 Resources of Energy and Environmental Issues
Ressources en énergie et enjeux environnementaux
- 3C-I3 Statistical Methods for Modeling Biology and Ecology Processes
Méthodes statistiques pour la modélisation des processus en biologie et en écologie
- 3C-I4 Targeted Maximum Likelihood Estimation (TMLE) and Related Methods
Estimation ciblée par maximum de vraisemblance (TMLE) et méthodes s'y rattachant
- 3C-C1 Probability in Actuarial Science and Finance
Probabilité en actuariat et en finance
- 3C-C2 Variable Selection, Clustering and Classification
Sélection de variables, partitionnement et classification
- 3D-A1 CJS Award Address
Allocution du récipiendaire du Prix de la RCS
- 3D-A2 Pierre Robillard Award Address
Allocution du récipiendaire du Prix Pierre-Robillard

- 3D-C1 Analysis of Longitudinal Data
Analyse de données longitudinales
- 3D-C2 Business and Industrial Statistics II
Statistique industrielle et de gestion II
- 3D-C3 Competing Risks and Disease Progression
Risques concurrents et progression des maladies
- 3D-C4 Mixture Models
Modèles de mélange
- 3D-C5 Screening and Diagnostic Tests
Examens préventifs et tests de diagnostic
- 3D-C6 Statistics: Applications and Case Studies
Statistique : applications et études de cas

Program • Programme

Saturday May 25

samedi 25 mai

18:00 - 22:30

Campus Tower Suite Hotel, Boardroom

Executive Committee 1

Comité exécutif 1

Sunday May 26**dimanche 26 mai****9:00 - 16:00****ECHA L1-150****Workshop of the Biostats Section: Analysis of Life History Data with Multistate Models****Atelier du groupe de biostatistique : analyse de données de cycles biologiques au moyen de modèles multi-états**

Organizers/Responsables: Richard Cook and/et Jerry Lawless (University of Waterloo)

9:00 - 16:00 Richard Cook (University of Waterloo)

9:00 - 16:00 Jerry Lawless (University of Waterloo)

9:00 - 16:00**ECHA L1-230****Workshop of the Business and Industrial Statistics Section and WatRISQ: Mathematical Models for Counterparty Credit Risk****Atelier du groupe de statistique industrielle et de gestion et WatRISQ : modèles mathématiques pour le risque de crédit de contrepartie**

Organizer and Chair / Responsable et président: David Saunders (University of Waterloo)

9:00 - 16:00 David Saunders (University of Waterloo)

9:00 - 16:00**ECHA L1-220****Workshop of the Probability Section: Estimation and Modeling Problems in Financial Engineering****Atelier du groupe de probabilité : estimation et modélisation de problèmes en ingénierie financière**

Organizer and Chair / Responsable et président: Bruno Rémillard (HEC Montréal)

9:00 - 16:00 Bruno Rémillard (HEC Montréal)

9:30 - 12:00**ECHA 3-001****Department Heads Meeting****Rencontre des directeurs de départements****10:30 - 12:00****ECHA 1-131****Publications Committee****Comité des publications****12:00 - 16:00****ECHA 4-036****Board of Directors 1****Conseil d'administration 1****12:15****Katz Group Centre Atrium****Afternoon Excursion - Fee Event****Excursion en après-midi - activité payante****16:00 - 18:00****ECHA L1-140****Information on 2013 NSERC Discovery Grant Competition****Information sur le concours 2013 des subventions à la découverte du CRSNG**

18:00 - 20:00

Centennial Centre for Interdisciplinary Science

Welcome Reception

Réception de bienvenue

19:30 - 21:30

Kyoto Japanese Cuisine (8709 109 Street / rue)

Dinner of the Caucus for Women in Statistics and the SSC Committee on Women in Statistics

Souper informel du Caucus for Women in Statistics et du comité de promotion de la femme en statistique de la SSC

Monday May 27**lundi 27 mai****8:15 - 8:30 Session 1A-A1****CCIS 1-430****Inaugural Session
Séance inaugurale**

Special Session / Conférence spéciale

Chair/Présidente: Rhonda J. Rosychuk (University of Alberta)

Organizer/Responsable: Christian Léger (Université de Montréal)

8:15 - 8:20 Duane Szafron (Vice-Provost and Associate Vice-President (IT), University of Alberta / Vice-recteur et vice-président, l'Université de l'Alberta)

8:20 - 8:25 Christian Léger (SSC President / Président de la SSC)

8:25 - 8:30 Rhonda J. Rosychuk (SSC 2013 Local Arrangements Chair / Présidente du comité des arrangements locaux, SSC 2013)

8:30 - 9:45 Session 1A-A2



(abstract/résumé - page 66)

CCIS 1-430**SSC Presidential Invited Address****Allocution de l'invitée du Président de la SSC**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Christian Léger (Université de Montréal)

8:30 - 9:45 Sallie Keller (University of Waterloo)

Celebrating the Challenges of Our Profession / Célébrons les défis de notre profession  **9:45 - 10:30****Katz Group Centre Atrium****Coffee Break****Pause-café****10:30 - 12:00 Session 1B-V**



(abstract/résumé - page 67)

MedSci 231**Statistics and the Environment****La statistique et l'environnement**

Introductory Overview Lecture / Conférence de vue d'ensemble

Organizer and Chair / Responsable et présidente: Debbie J. Dupuis (HEC Montréal)

10:30 - 12:00 Christopher Wikle (University of Missouri)

Statistics and the Environment: Overview and Challenges / Statistique et environnement : vue d'ensemble et défis  **10:30 - 12:00 Session 1B-I1**

(abstracts/résumés - page 68)

ECHA L1-430**Financial Mathematics****Mathématiques financières**



Invited Paper Session / Présentations sur invitation





Organizer and Chair / Responsable et président: Rafal Kulik (University of Ottawa)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

10:30 - 11:00 Don McLeish (University of Waterloo)

Martingales, Likelihood, and Monte Carlo Methods for Continuous Time Models in Finance / Martingales, vraisemblance et méthodes de Monte Carlo pour les modèles en temps continu en finance  

- 11:00 - 11:30 Cody Hyndman & Polynice Oyono Ngou (Concordia University)
A Convolution Method for Numerical Solution of Backward Stochastic Differential Equations / Une méthode par convolution pour résoudre numériquement les équations différentielles stochastiques rétrogrades  
- 11:30 - 12:00 Tahir Choulli & Junfeng Ma (University of Alberta)
Stochastic Hellinger Distance: A Statistical Tool for a Financial Economics Problem / Distance stochastique de Hellinger : méthode statistique pour un problème en finance  

10:30 - 12:00 Session 1B-I2 (abstracts/résumés - page 70) **EdSouth 254**

New Developments in Statistical Neuroscience







Nouveaux développements en statistique de la neuroscience

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Shoja'eddin Chenouri (University of Waterloo)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 10:30 - 11:00 Sam Behseta (California State University)
Bayesian Modeling of Multiple Spike Trains / Modélisation bayésienne de suites de potentiels d'action multiples  
- 11:00 - 11:30 Xi (Rossi) Luo (Brown University), Steven Gee & Vikaas Sohal (University of California, San Francisco), Dylan Small (University of Pennsylvania)
A Point-Process Response Model for Optogenetic Experiments on Neural Circuits / Un modèle de réponse d'un processus ponctuel dans les expériences optogénétiques sur les circuits neuronaux  
- 11:30 - 12:00 Reza Ramezan, Shoja'eddin Chenouri & Paul Marriott (University of Waterloo)
The Skellam Process: A New Approach to Modelling Neural Spike Trains / Le processus de Skellam : une nouvelle approche pour la modélisation de suites de potentiels d'action neurale  







10:30 - 12:00 Session 1B-I3 (abstracts/résumés - page 72) **ECHA 1-498**

Recent Development on Computer Experiments

Développements récents en expériences informatiques

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Devon C. Lin (Queen's University)

- 10:30 - 11:00 Dave Woods (University of Southampton), Antony Overstall (University of St Andrews)
Bayesian Inference for Computationally Expensive Models with Application to the Estimation of Stem Cell Properties / Inférence bayésienne pour modèles coûteux en ressources et applications pour l'estimation des propriétés des cellules souches  
- 11:00 - 11:30 Jason Loeppky (University of British Columbia - Okanagan), Hao (Nelson) Chen & William Welch (University of British Columbia)
Bayesian Prediction and Inference in Analysis of Computer Experiments / Prévission et inférence bayésiennes en analyse d'expériences sur ordinateur  
- 11:30 - 12:00 Pritam Ranjan (Acadia University), Chunfang Devon Lin (Queen's University)
A New Multi-Stage Sequential Design Approach for Estimating Quantiles of Simulator Outputs / Approche conceptuelle séquentielle multi-étapes pour estimer les quantiles des résultats de simulateurs  

10:30 - 12:00 **Session 1B-I4** (abstracts/résumés - page 74) **ECHA L1-140**

Survey Sampling and Applied Mathematics

Échantillonnage et mathématiques appliquées







Invited Paper Session / Présentations sur invitation

Chair/Président: J.N.K. Rao (Carleton University)

Organizers/Responsables: Susana Rubin Bleuer (Statistics Canada) and/et Ioana Schiopu Kratina (University of Ottawa)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 10:30 - 11:00 Susana Rubin Bleuer (Statistics Canada), Pascal Rivière (Caisse Nationale d'Assurance Vieillesse)
Micro-strata / Micro-strates  
- 11:00 - 11:30 Lenka Mach & Jean-Marc Fillion (Statistics Canada), Ioana Schiopu-Kratina (University of Ottawa)
Sample Coordination Using Methods Based on Linear Programming / Coordination des échantillons à l'aide de méthodes basées sur la programmation linéaire  
- 11:30 - 12:00 Changbao Wu (University of Waterloo)
Bayesian Pseudo-EL Approach Using Survey Data: Overview and Application to Regression Analysis / Approche bayésienne de pseudo-vraisemblance empirique à l'aide de données d'enquête : aperçu et application à l'analyse de régression  









10:30 - 12:00 **Session 1B-C1** (abstracts/résumés - page 76) **EdSouth 228**

Bayesian Estimation and Inference

Estimation et inférence bayésiennes

Contributed Paper Session / Séance de communications libres

Chair/Président: Paul Gustafson (University of British Columbia)

- 10:30 - 10:45 Avik Halder & Glen Takahara (Queen's University)
Example of Bayesian Semi-Parametric Inference for the Frailty Model Using Lévy Process Prior / Un exemple d'inférence semi-paramétrique bayésienne du modèle de fragilité au moyen du processus de Lévy comme loi a priori  
- 10:45 - 11:00 Khader Khadraoui (Université Laval), Pierre Ribereau (Université Claude Bernard)
Bayesian Estimation with M-splines of the Spectral Measure of an Extreme-Value Distribution / Estimation bayésienne avec M-splines de la mesure spectrale d'une distribution de valeurs extrêmes  
- 11:00 - 11:15 Luai Al Labadi (UAE University), Mahmoud Zarepour (University of Ottawa)
A Bayesian Nonparametric Goodness of Fit Test for Right Censored Data Based on Approximate Samples from the Beta-Stacy Process / Test d'adéquation non-paramétrique bayésien pour des données censurées à droite sur des échantillons approximatifs provenant du processus beta-Stacy  
- 11:15 - 11:30 Saman Muthukumarana (University of Manitoba)
Bayesian Analysis and Model Assessment of Binary Item Response Data / Analyse bayésienne et évaluation de modèle de données de réponse d'item binaires  
- 11:30 - 11:45 Mohammed Obeidat & Juxin Liu (University of Saskatchewan)
Bayesian Estimation of Time Series of Counts / Estimation bayésienne de séries chronologiques avec données de dénombrement  
- 11:45 - 12:00 Farhad Shokoohi & Mahmoud Torabi (University of Manitoba)
Bayesian Small Area Estimation Using P-Spline and Time Series Models / Estimation bayésienne pour petits domaines au moyen de modèles de p-splines et de séries chronologiques  









10:30 - 12:00 Session 1B-C2 (abstracts/résumés - page 79) **EdSouth 206**
Biostatistics - Theory and Methods I
Biostatistique - théorie et méthodes I





Contributed Paper Session / Séance de communications libres
 Chair/Présidente: Jinko Graham (Simon Fraser University)

- 10:30 - 10:45 Abdelaziz Alsubie & Jianan Peng (Acadia University)
 Simultaneous Inference for Low Dose Risk Estimation with Quantal Data in Benchmark Analysis / L'inférence simultanée pour l'estimation du risque à faible dose avec des données quantales pour l'analyse d'un banc d'essais  
- 10:45 - 11:00 Caitlin Daly, Xiaofeng Liu, Angela Rutledge, Kristin Hauff, Vijay Grey & Jemila Hamid (McMaster University)
 A Novel Statistical Method for Transference of Paediatric Reference Intervals Using Hospital-Based Data / Une nouvelle méthode statistique pour le transfert d'intervalles de référence pédiatriques au moyen de données hospitalières  
- 11:00 - 11:15 Depeng Jiang & Robert Tate (University of Manitoba)
 Latent Class Joint Model for Longitudinal and Time-to-Event Data / Modèle conjoint de classes latentes pour des données longitudinales et de temps avant événement  
- 11:15 - 11:30 Kun Liang (University of Waterloo), Sunduz Keles (University of Wisconsin - Madison)
 Normalization of ChIP-seq Data with Control / Normalisation des données ChIP-Seq avec contrôle  
- 11:30 - 11:45 Pooneh Pordeli, Xuewen Lu & Murray Burke (University of Calgary), Peter X.-K. Song (University of Michigan)
 Polynomial Spline Estimation for Partially Linear Single-Index Transformation Models with Current Status Data / Estimation par splines polynomiales pour des modèles de transformation partiellement linéaires à indice simple avec données actualisées  
- 11:45 - 12:00 Yujie Zhong & Richard Cook (University of Waterloo)
 Analysis of Current Status Family Data Subject to Ascertainment Bias / Analyse de données familiales de statut présent sujettes à un biais de constatation  

10:30 - 12:00 Session 1B-C3 (abstracts/résumés - page 82) **EdSouth 176**
Clustering and Classification
Partitionnement et classification

Contributed Paper Session / Séance de communications libres
 Chair/Président: Hugh Chipman (Acadia University)

- 10:30 - 10:45 Reshu Agarwal & Pritam Ranjan (Acadia University)
 mBACT: A New Classifier for Satellite Images / Arbre de classification naïve bayésienne (mBACT) : un nouveau classificateur pour les images satellite  
- 10:45 - 11:00 Khatoon Alobaidan (Acadia University), Hugh Chipman (Acadia University)
 A Measure of Compactness for Hierarchical Clustering / Une mesure de compacité pour le regroupement hiérarchique  
- 11:00 - 11:15 Sanjeena Dang & Paul McNicholas (University of Guelph), Salvatore Ingrassia & Antonio Punzo (University of Catania)
 Cluster-Weighted Factor Analyzers / Analyseurs factoriels par pondération de grappes  
- 11:15 - 11:30 Thierry Chekouo Tekougang (UT MD Anderson Cancer Center), Alejandro Murua (Université de Montréal)
 Variable Selection With Overlapping Clustering / Sélection de variables avec groupements qui se chevauchent  

11:30 - 11:45	Cristina Tortora & Paul McNicholas (University of Guelph), Francesco Palumbo (Università di Napoli Federico II) Probabilistic and Distance Based Factor Clustering: A Review / Étude de regroupement de facteurs basé sur la distance et probabiliste : une revue	 
11:45 - 12:00	Irene Vrbik & Paul McNicholas (University of Guelph) Fractionally-Supervised Classification / Classification partiellement dirigée	 

12:00 - 13:30 **Hudsons Canadian Tap House (Campus)**
uWaterloo Math Alumni Lunch
Dîner des anciens étudiants en math de la uWaterloo

12:00 - 18:00 **KATZ 2-155**
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 **KATZ 2-152**
Biostatistics Section Executive Committee 1
Comité exécutif de biostatistique industrielle et de gestion 1

12:00 - 13:30 **KATZ 2-153**
Business and Industrial Statistics Section Executive Committee
Comité exécutif du groupe de statistique industrielle et de gestion

12:00 - 13:30 **ECHA 1-441**
Committee on SSC Bylaws
Comité de statuts de la SSC

12:00 - 13:30 **ECHA 2-407**
Committee on Women in Statistics
Comité de promotion de la femme en statistique

12:00 - 13:30 **KATZ 2-151**
Probability Section Executive Committee
Comité exécutif du groupe de probabilité

12:00 - 13:30 **KATZ 2-157**
Statistical Education Committee
Comité d'éducation en statistique

12:00 - 13:30 **KATZ 2-157**
Statistical Education Section Executive Committee
Comité exécutif du groupe d'éducation en statistique

12:00 - 13:30 **ECHA 1-447**
Strategic Planning Committee
Comité de la planification stratégique

12:00 - 13:30 **KATZ 2-154**
Survey Methods Section Executive Committee
Comité exécutif du groupe de méthodes d'enquête

13:30 - 15:00 **Session 1C-I1** (abstracts/résumés - page 85) **ECHA L1-430**
Dependence Models with Insurance Applications
Modèles de dépendance avec applications en assurance

Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et président: Gordon Willmot (University of Waterloo)
Session sponsored by the Actuarial Science Section
Séance parrainée par le Groupe d'actuariat

- 13:30 - 14:00 Jun Cai (University of Waterloo)
Dependence Notions Through Arrangement Increasing Joint Survival Functions with Applications to Optimal Allocation Problems / Notions de dépendance par le biais de fonctions de survie conjointes et applications aux problèmes d'imputation optimale E E
- 14:00 - 14:30 Christian Genest, Huijun Chen & David A. Stephens (McGill University)
Bayesian Hierarchical Modelling of Automobile Insurance Claim Reserves / Modélisation bayésienne hiérarchique des provisions pour sinistres en assurance automobile E E
- 14:30 - 15:00 Étienne Marceau & Hélène Cossette (Université Laval)
Recent Developments on Aggregation of Dependent Risks / Développements récents dans l'agrégation des risques dépendants E E
-

13:30 - 15:00 **Session 1C-I2** (abstracts/résumés - page 87) **EdSouth 206**
Fitting Complex Models Using Automatic Differentiation Model Builder (ADMB)
Ajustement de modèles complexes en utilisant ADMB

Invited Paper Session / Présentations sur invitation
Chair/Présidente: Joanna Mills Flemming (Dalhousie University)
Organizer/Responsable: Noel Cadigan (Memorial University)

- 13:30 - 14:00 Mollie Brooks, Thomas Bohrmann, Forrest Stevens, Andrew Hein, Mary Christman & Craig Osenberg (University of Florida), Benjamin Bolker (McMaster University)
Predicting Temporally Autocorrelated Pond Depths with Irregular Time Series Data / Prédire une profondeur de bassin avec autocorrélation temporelle à l'aide de données chronologiques irrégulières E E
- 14:00 - 14:30 Noel Cadigan (Memorial University)
Improved Estimation and Forecasts of Fish Stock Maturities Using Generalized Linear Mixed Effects Models / Modèles linéaires généralisés à effets mixtes : améliorer l'estimation et les prévisions de la maturité d'un stock de poissons E E
- 14:30 - 15:00 Anders Nielsen (Danish Technical University)
Introduction to AD Model Builder and State-Space Fish Stock Assessment Model / Introduction à AD Model Builder et modèles d'évaluation d'espaces d'états de stocks de poissons E E
-

13:30 - 15:00	Session 1C-I3	(abstracts/résumés - page 89)	EdSouth 176
High-Dimensional Data Analysis for Brain Imaging			
Analyse de données de grande dimension pour l'imagerie encéphalique			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Farouk Nathoo (University of Victoria)			
13:30 - 14:00	Mirza Faisal Beg (Simon Fraser University)	Computational Morphometric Biomarkers for Discrimination of Dementias / Biomarqueurs morphométriques informatiques pour la discrimination des démences	ⓔ ⓔ
14:00 - 14:30	Timothy D. Johnson (University of Michigan), Thomas E. Nichols & Tian Ge (University of Warwick)	A GPU Implementation of a Spatial GLMM: Assessing Spatially Varying Coefficients of Multiple Sclerosis Lesions / Implémentation GPU d'un MLGM spatial : évaluation de coefficients à variation spatiale de lésions de sclérose en plaque	ⓔ ⓔ
14:30 - 15:00	Mary Lesperance & Farouk Nathoo (University of Victoria), Andrew Lawson (Medical University of South Carolina), Charmaine Dean (University of Western Ontario)	Comparing Variational Bayes with Markov Chain Monte Carlo for Bayesian Computation in Neuroimaging / Comparaison du calcul bayésien variationnel et Monte Carlo par chaîne de Markov pour calcul bayésien en neuroimagerie	ⓔ ⓔ

13:30 - 15:00	Session 1C-I4	(abstracts/résumés - page 91)	ECHA 1-498
Identifiability: Pushing Data to the Limits			
Identifiabilité : les données poussées à leurs limites			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Luke Bornn (Harvard University)			
13:30 - 14:00	Paul Gustafson (University of British Columbia)	Bayesian Inference in Partially Identified Models: Exploring the Limits of Limited Data / Inférence bayésienne dans les modèles partiellement identifiés : aux limites des données limitées	ⓔ ⓔ
14:00 - 14:30	Wes Johnson (UC Irvine), Adam Branscum (Oregon State University), Tim Hanson (University of South Carolina)	Bayesian Semiparametric (Identifiable) Regression Modeling for ROC Analysis, Absent a Gold Standard / Modélisation de la régression semiparamétrique bayésienne (identifiable) pour l'analyse ROC en l'absence d'une norme de référence	ⓔ ⓔ
14:30 - 15:00	Stephen Walker (University of Kent)	Identifiability Issues with Mixtures of Normal Distributions and the Mixture of Dirichlet Process Model / Problèmes d'identification avec les mélanges de distributions normales et le modèle de mélange de processus de Dirichlet	ⓔ ⓔ

13:30 - 15:00	Session 1C-I5	(description - page 93)	EdSouth 254
Moving Up in Academic Ranks: Challenges and Strategies for Statisticians Involved in Interdisciplinary Research			
Gravir les échelons universitaires – défis et stratégies pour les statisticiens impliqués dans la recherche interdisciplinaire			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et présidente: Wendy Lou (University of Toronto)			
13:30 - 15:00	Joel A. Dubin (University of Waterloo)		
13:30 - 15:00	Terry Hyslop (Thomas Jefferson University & Kimmel Cancer Center)		
13:30 - 15:00	Karen A. Kopciuk (Alberta Health Services - Cancer Care & University of Calgary)		







- 13:30 - 15:00 A. John Petkau (University of British Columbia)
 13:30 - 15:00 Lehana Thabane (McMaster University)

13:30 - 15:00 Session 1C-I6 (abstracts/résumés - page 94) **ECHA L1-140**

Survival Analysis Methods and Applications
Méthodes et applications en analyse de survie

Invited Paper Session / Présentations sur invitation

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











- 13:30 - 14:00 Richard Cook (University of Waterloo)
 Assessment of Treatment Effects Following Intermediate Events in Cancer Trials / Évaluation des effets du traitement après événements intermédiaires dans les essais cliniques sur le cancer  
- 14:00 - 14:30 Yi Li (University of Michigan), Peirong Xu & Lixing Zhu (Hong Kong Baptist University)
 Ultrahigh Dimensional Time Course Feature Selection / Sélection de caractéristiques de période de dimension très élevée  
- 14:30 - 15:00 Ronghui (Lily) Xu & Gordon Honerkamp-Smith (University of California, San Diego)
 Explained Variation in Correlated Survival Data / Variation expliquée dans les données de survie corrélées  

13:30 - 14:45 Session 1C-C1 (abstracts/résumés - page 96) **EdSouth 265**

Biostatistics - Theory and Methods II
Biostatistique - théorie et méthodes II

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Mireille Schnitzer (Harvard School of Public Health)

- 13:30 - 13:45 Sayantee Jana & Narayanaswamy Balakrishnan (McMaster University), Dietrich von Rosen (Swedish University of Agricultural Sciences), Jemila Hamid (McMaster University)
 A Moderated Trace Test for the Growth Curve Model and Its Application for Gene-Filtering in Time-Course Microarray Data / Test de trace modéré pour le modèle de courbe de croissance et application au filtrage de gènes dans les données de micropuces de période  
- 13:45 - 14:00 Stefan Konigorski (University of Toronto), Yildiz E. Yilmaz & Shelley B. Bull (University of Toronto & Samuel Lunenfeld Research Institute of Mount Sinai Hospital)
 A Censored Regression Method to Adjust for Medication Effects in Genetic Analysis of a Quantitative Trait When the Treatment Received Depends on the Trait / Méthode de régression avec données censurées pour tenir compte des effets des médicaments dans l'analyse génétique d'un caractère quantitatif lorsque le traitement dépend du caractère  
- 14:00 - 14:15 Chel Hee Lee & Mikelis Bickis (University of Saskatchewan)
 Zero-Truncated Poisson Imprecise Probability Model for Quantifying Prior Ignorance / Modèle de probabilité imprécise de Poisson tronqué en zéro pour quantifier l'ignorance préalable  
- 14:15 - 14:30 Zihui (Amy) Liu, James Hanley, Nandini Dendukuri & Erin Strumpf (McGill University)
 Measuring the Yearly Mortality Reductions Due to a Cancer Screening Program: A Round-by-Round Approach / Mesure de la réduction de mortalité annuelle due à un programme de dépistage du cancer : approche par ronde  
- 14:30 - 14:45 Maryam Zolghadr (University of Gothenburg), Sergei Zuyev (Chalmers University of Technology)
 Optimal Design of Dilution Experiments under Volume Constraints / Plan optimal pour expériences de dilution sous contraintes de volume  
-

13:30 - 15:00 **Session 1C-C2** (abstracts/résumés - page 99) **EdSouth 221**

Business and Industrial Statistics I
Statistique industrielle et de gestion I









Contributed Paper Session / Séance de communications libres
 Chair/Président: Jason Loeppky (University of British Columbia)



- 13:30 - 13:45 Yumi Kondo & James V. Zidek (University of British Columbia)
 Bayesian Nonparametric Subset Selection Procedures with Weibull Components / Procé-
 dures de sélection de sous-ensembles non-paramétriques bayésiens avec les composantes
 de Weibull  
- 13:45 - 14:00 Mark Krzeminski & Xiaojian Xu (Brock University)
 Optimal Designs for Accelerated Life Tests with Multiple Stress Factors / Plans optimaux pour
 des tests de durée de vie accélérée avec plusieurs facteurs de stress  
- 14:00 - 14:15 Blake MacDonald, Pritam Ranjan & Hugh Chipman (Acadia University)
 GPfit: An R Package for Fitting Gaussian Process Models to Simulator Data / GPfit : un
 package R pour l'ajustement des modèles de processus gaussien pour données de simula-
 teur  
- 14:15 - 14:30 Nathaniel Payne & Tom Loughin (Simon Fraser University)
 Development of a Structural Wood Adhesive Testing Standard Using a Nested Generalized
 Linear Mixed Model / Développement d'une norme d'essai d'adhésif pour le bois de charpente
 à l'aide d'un modèle linéaire généralisé mixte emboîté  
- 14:30 - 14:45 Jimmy Poon & Yi Lu (Simon Fraser University)
 A Spatial Cross-Sectional Credibility Model with Dependence Among Risks / Modèle de cré-
 dibilité transversal spatial avec dépendance entre les risques  
- 14:45 - 15:00 Neil Spencer (Acadia University)
 Isomorphism Check for Two-Level Star-Based Factorial Designs with Randomization Restric-
 tions / Vérification de l'isomorphisme pour plans factoriels à deux niveaux et en étoile avec
 restrictions de randomisation  

13:30 - 14:45 **Session 1C-C3** (abstracts/résumés - page 102) **EdSouth 164**

Environmental Modeling: Applications and Case Studies
Modélisation en environnement : applications et études de cas

Contributed Paper Session / Séance de communications libres
 Chair/Président: Jean-François Angers (Université de Montréal)













- 13:30 - 13:45 Alisha Albert-Green (University of Western Ontario), Douglas Woolford (Wilfrid Laurier Univer-
 sity), W. John Braun (University of Western Ontario), David Martell (University of Toronto)
 Visualization Tools for Assessing the Markov Property: Sojourn Times in the Ontario Fire
 Weather Index / Outils de visualisation pour l'évaluation de la propriété markovienne : durées
 de séjour de l'indice forêt-météo en Ontario  
- 13:45 - 14:00 Wesley Burr (Queen's University), Hwashin H. Shin (Health Canada)
 Temperature and Mortality: a Canadian Perspective / Température et mortalité : point de vue
 canadien  
- 14:00 - 14:15 Laura Cowen (University of Victoria), Takis Besbeas & Byron Morgan (University of Kent), Carl
 Schwarz (Simon Fraser University)
 A Comparison of Abundance Estimates Obtained from Extended Batch Marking and Jolly-
 Seber Type Experiments / Comparaison d'estimés d'abondance obtenus par marquage de
 lots étendus et expériences de type Jolly-Seber  
- 14:15 - 14:30 Wenkai (Kevin) Ma, W. John Braun & Gregory Kopp (University of Western Ontario)
 A Simulation Model for Wind-Driven Fire-Spotting / Modèle de simulation pour la dissémina-
 tion du feu par le vent  

- 14:30 - 14:45 Shan Qiao (University of Saskatchewan)
Estimating the Effects of Air Pollutants on Recurrent Hospital Admission for Respiratory Diseases / Estimation des effets des polluants atmosphériques sur les admissions récurrentes à l'hôpital dues aux maladies respiratoires  

13:30 - 15:00 Session 1C-C4 (abstracts/résumés - page 104) **EdSouth 276**

**Probability I
Probabilité I**

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Gail Ivanoff (University of Ottawa)





- 13:30 - 13:45 Ibrahim Abdelrazeq, Gail Ivanoff & Rafal Kulik (University of Ottawa)
Inference for Discretely Observed Levy-Driven Ornstein-Uhlenbeck Processes / Inférence pour des processus d'Ornstein-Uhlenbeck observés discrètement et gouvernés par des processus de Lévy  
- 13:45 - 14:00 Farid Elktaibi & Gail Ivanoff (University of Ottawa)
Asymptotics for Linear Processes and Change-Point Detection / Comportement asymptotique pour des processus linéaires et la détection de points de rupture  
- 14:00 - 14:15 Craig Kainulainen (Lakehead University)
Average Distance Results in Geometric Figures and Compact Connected Spaces / Distances moyennes dans des figures géométriques et des espaces compacts connexes  
- 14:15 - 14:30 Deli Li (Lakehead University), Oleg Klesov (National Technical University of Ukraine Kiev Polytechnic Institute), George Stoica (University of New Brunswick)
On the Central Limit Theorem along Subsequences of Sums of I.I.D. Random Variables / À propos du théorème central limite pour des sous-suites de sommes de variables aléatoires i.i.d.  
- 14:30 - 14:45 Weiwei Liu, Reg Kulperger & Hao Yu (University of Western Ontario)
Asymptotic Theory for GARCH-in-Mean Processes / Théorie asymptotique pour des processus GARCH en moyenne  
- 14:45 - 15:00 Samira Sadeghi (University of Alberta)
On The Almost Sure Rates of Convergence of Linear Algorithms with Slowly Decreasing Gain / À propos des taux quasi-certains de convergence d'algorithmes linéaires avec gain en décroissance lente  







13:30 - 14:45 Session 1C-C5 (abstracts/résumés - page 107) **EdSouth 228**

Survey: Theory, Methods and Applications

Échantillonnage : théorie, méthodes et applications

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

- 13:30 - 13:45 Min Chen, Changbao Wu & Mary Thompson (University of Waterloo)
Pseudo Empirical Likelihood Method for Pretest-Posttest Studies Using Survey Data / Méthode de pseudo-vraisemblance empirique pour les études prétest/post-test avec données d'enquête  
- 13:45 - 14:00 Gauri Sankar Datta (University of Georgia, U.S. Bureau of the Census), William Bell & Jerzy Wieczorek (U.S. Bureau of the Census)
A Weighted Likelihood Approach to Model-Based Small Area Estimation with Unit-Level Data / Approche de vraisemblance pondérée pour l'estimation de modèles pour petits domaines avec données au niveau des unités  

- 14:00 - 14:15 Darren Gray (Statistics Canada)
A Multi-Constraint Sample Allocation for Detailed National Estimates Made Trickier by Provincial Buy-Ins / Une répartition d'échantillon sous contraintes multiples pour des estimations nationales détaillées complexifiée par l'ajout de demandes de partenaires provinciaux  
- 14:15 - 14:30 Pierre-Olivier Julien, Christian Gagné & Carlos A. Leon (Statistique Canada)
Negative Sample Coordination: An Overview of Some Methods / Coordination négative d'échantillons : un survol de quelques méthodes  
- 14:30 - 14:45 Christian Olivier Nambu & Mike Hidioglou (Statistique Canada)
Calibration Using Estimated Constraints / Calage utilisant des contraintes estimées  

13:30 - 15:00 Session 1C-C6 (abstracts/résumés - page 110)

EdSouth 177

Variable Selection and Statistical Learning

Sélection de variables et apprentissage statistique

Contributed Paper Session / Séance de communications libres

Chair/Président: John Koval (University of Western Ontario)

- 13:30 - 13:45 Nasrin Afzal & Xuewen Lu (University of Calgary)
Variable Selection in a Partially Linear Additive Hazard Model with the Elastic Net Penalty / Sélection de variables pour un modèle partiellement linéaire à risques additifs avec pénalité en filet élastique  
- 13:45 - 14:00 Brian Franczak, Ryan Browne & Paul McNicholas (University of Guelph)
Clustering and Classification of Asymmetric Data / Regroupement et classification de données asymétriques  
- 14:00 - 14:15 Longlong Huang & Xuewen Lu (University of Calgary), Karen A. Kopciuk (Alberta Health Services)
Jackknife Empirical Likelihood Based Variable Selection for Accelerated Failure Time Model / Sélection de variables basée sur la vraisemblance empirique jackknife pour un modèle de temps de défaillance accéléré  
- 14:15 - 14:30 Zi Zhen Liu, Reg Kulperger & Hao Yu (University of Western Ontario)
The Adaptive Positive LASSO Procedure for Modelling ARCH Processes / Modélisation des processus ARCH par la méthode du LASSO positif adaptatif  
- 14:30 - 14:45 Javed Hossain Tomal, William Welch & Ruben H. Zamar (University of British Columbia)
Ensembling Classification Models Based on Phalanxes of Variables with Applications in Drug Discovery / Apprentissage par ensemble pour des modèles de classification basés sur les phalanxes des variables avec application à la découverte de médicaments  
- 14:45 - 15:00 Yukun Zhang & Xuewen Lu (University of Calgary)
Coordinate Descent Methods for the Birnbaum-Saunders Regression Model with High-Dimensional Survival Data / Méthodes de descente des coordonnées pour le modèle de régression de Birnbaum-Saunders avec des données de survie de grande dimension  

13:30 - 15:30 Session 1C-D1

Second floor atrium, Katz Group Centre

Case Study I: Exploring Spatial and Temporal Heterogeneity of Environmental Noise in Toronto

Étude de cas I : L'exploration de l'hétérogénéité spatiale et temporelle du bruit environnemental à Toronto

Case Study / Étude de cas

Chair/Président: Lajmi Lakhal-Chaieb (Université Laval)

Organizer/Responsable: Lennon Li (Public Health Ontario & University of Toronto)

- 13:30 - 15:30 Shiva Ashta, Doug Cohen, Kelvin Hung & Jin Hyung Lee (University of Toronto)
- 13:30 - 15:30 Janie Coulombe, Kenza Sallier & Audrey-Anne Vallée (Université de Montréal)
- 13:30 - 15:30 Nicholas Fishbane, Seong-Hwan Jun, Yunlong Nie & Camila Casquilho Resende (University of British Columbia)

- 13:30 - 15:30 Charlotte Haley, Evi Lansimaki, Joshua Pohlkamp-Hartt, David Riegert & Aaron Springford (Queen's University)
- 13:30 - 15:30 Daming Lin, Shiyi Chen, Zihang Lu, Stefan Konigorski & Anastasia Teterina (University of Toronto)
- 13:30 - 15:30 Adrian Waddell & Qianyi Huang (University of Waterloo)

13:30 - 15:30 Session 1C-D2 Second floor atrium, Katz Group Centre

Case Study II: Study of the Impact of Human Disturbances on Habitat Selection and on Calf Survival for Forest-Dwelling Caribou

Étude de cas II : Étude de l'impact des perturbations causées par l'homme sur la sélection d'habitat et la survie des faons dans l'écosystème forestier du caribou des bois

Case Study / Étude de cas

Chair/Président: Lajmi Lakhali-Chaieb (Université Laval)

Organizer/Responsable: Thierry Duchesne (Université Laval)

- 13:30 - 15:30 Mohamad Elmasri, Marie-Pier Cote & Ievgenii Grebennikov (McGill University)
- 13:30 - 15:30 Ruth Joy & Zheng Sun (Simon Fraser University)
- 13:30 - 15:30 Anastasia Teterina, Priscila Pequeno & Zihang Lu (University of Toronto)

13:30 - 15:30 Session 1C-P (abstracts/résumés - page 113) Main floor atrium, Katz Group Centre

Poster Session

Séance d'affichage

Contributed Poster Session / Séance d'affichage libre

- 13:30 - 15:30 Taryn Athey, Paul McNicholas & Robert Hanner (University of Guelph)
Using Frequency Matrices for a Cross-Taxa Analysis of Errors in DNA Barcodes / Utilisation des matrices de fréquence pour une analyse des erreurs entre taxons dans le codage à barres moléculaire
- 13:30 - 15:30 Maggie Chu & Wenqing He (University of Western Ontario)
Cluster Analysis of Time-Course RNA-seq Data: Evaluation of Initialization / Analyse de regroupement pour la progression temporelle de données seq-ARN : évaluation et initialisation
- 13:30 - 15:30 Bolanle Dansu & Lisa M. Lix (University of Manitoba), Tolulope Sajobi (University of Calgary)
Statistical Inference for Discriminant Function Coefficients in Doubly Multivariate Data / Inférence statistique pour les coefficients de la fonction discriminante avec des données doublement multivariées
- 13:30 - 15:30 Spectrum Han (University of Western Ontario)
Models for Crown Fire Rate of Spread / Modèles pour la vitesse de propagation des feux de cimes
- 13:30 - 15:30 Zhiyong Jin (University of Manitoba)
A Fast Numerical Algorithm for Boundary Crossing Probabilities for Jump Diffusion Processes / Un algorithme numérique rapide pour les probabilités de traverser des frontières dans un processus de diffusion avec sauts
- 13:30 - 15:30 Yang Liu, Matías Salibián-Barrera, Ruben H. Zamar & James V. Zidek (University of British Columbia)
Lower Quantile Estimation via Subjectively Censored Weibull MLE with Bootstrap Threshold Selection / Estimation des quantiles inférieurs à l'aide d'un estimateur du maximum de vraisemblance pour le modèle Weibull censuré subjectivement dont la sélection des seuils est effectuée par le bootstrap
- 13:30 - 15:30 Nickchi Payman, Pezeshk Hamid & Shemehsavar Soudabeh (University of Tehran)
A Bidirectional Hidden Markov Model Together with External Hints for Gene Prediction / Un modèle de Markov caché bidirectionnel avec indices extérieurs pour la prévision de gènes

- 13:30 - 15:30 Adam Rahman (Western University), Duncan Murdoch (University of Western Ontario)
Sweavetest - Automatic Multiple Choice Test Analysis / Sweavetest - Analyse automatique des tests à choix multiples
- 13:30 - 15:30 Tianji Shi (University of British Columbia), Ciprian Pirvu (FPInnovations), William Welch & James V. Zidek (University of British Columbia)
Tension Proofloading on Fingerjointed Lumber: Benefits and Optimization / Essais de charge pour la tension sur les pièces de bois à entures multiples : bénéfiques et optimisation
- 13:30 - 15:30 Zheng Sun, Richard Lockhart & Michael Stephens (Simon Fraser University)
Testing Fit for Latent Variables / Tests d'ajustement pour des variables latentes
- 13:30 - 15:30 Yang Tang, Paul McNicholas & Ryan Browne (University of Guelph)
Common Factors For Clustering Data of Mixed Type / Facteurs communs pour le regroupement de données de type mixte
- 13:30 - 15:30 Jinkun Xiao & Ian McLeod (University of Western Ontario)
Improved Portmanteau Diagnostic Check for ARFIMA Time Series Models / Diagnostic de « Portmanteau » amélioré pour les modèles de séries chronologiques ARFIMA
- 13:30 - 15:30 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
- 13:30 - 15:30 Lin Xue & Depeng Jiang (University of Manitoba)
Longitudinal Analysis of Development and Well-Being of Canadian Children and Youth / Analyse longitudinale du développement et du bien-être des enfants et de la jeunesse canadienne

15:00 - 15:30
Coffee Break
Pause-café

Katz Group Centre Atrium

15:30 - 17:00 Session 1D-I1 (abstracts/résumés - page 119)

ECHA L1-140

Advances in High-Dimensional Data Analysis

Avancées en matière d'analyse de données de grande dimension

Invited Paper Session / Présentations sur invitation

Chair/Présidente: Xiaojian Xu (Brock University)

Organizer/Responsable: Sanjoy Sinha (Carleton University)

- 15:30 - 16:00 Christina Kendzioriski & Ning Leng (University of Wisconsin - Madison)
EBSeq: An Empirical Bayesian Approach for Inference in RNA-seq Experiments / EBSeq : une approche bayésienne empirique pour l'inférence dans les expériences de séquençage d'ARN  
- 16:00 - 16:30 Abdus Sattar (Case Western Reserve University), Shili Lin & Taslim Cenny (Ohio State University), Sheldon Bai & Sharoon Hanook (Case Western Reserve University)
Prediction of Genome-Wide Promoters and Enhancers Using a Flexible Semiparametric Mixed Effects Model / Préviation de promoteurs et amplificateurs génétiques via un modèle semi-paramétrique souple à effets mixtes  
- 16:30 - 17:00 Sanjoy Sinha (Carleton University), Abdus Sattar (Case Western Reserve University)
Robust Methods for Selecting Differential miRNAs in Breast Cancer / Méthodes robustes de sélection de micro-ARN différentiels dans le cancer du sein  







15:30 - 17:00 **Session 1D-I2** (abstracts/résumés - page 121) **EdSouth 176**

Recent Developments in R packages

Développements récents dans les packages R

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Duncan Murdoch (University of Western Ontario)

- 15:30 - 16:00 Peter Solymos, Khurram Nadeem & Subhash Lele (University of Alberta)
Complex Models in Ecology: Challenges and Solutions / Modèles complexes en écologie : difficultés et solutions  
- 16:00 - 16:30 Michael Friendly & Matthew Sigal (York University)
Recent Advances in Visualization for Multivariate Linear Models / Progrès récents en visualisation pour les modèles linéaires multivariés  
- 16:30 - 17:00 Paul Gilbert (retired)
2piQA: Crowd-Sourcing R Package Quality Assessment / 2piQA : Évaluation de la qualité du package R de « crowd-sourcing »  

15:30 - 17:00 **Session 1D-I3** (description - page 123) **EdSouth 254**

Some Thoughts on How to be More Successful

Quelques pistes pour connaître un plus grand succès

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Dena Schanzer (Public Health Agency of Canada)

Session sponsored by the Accreditation Committee

Séance parrainée par le Comité d'accréditation

- 15:30 - 17:00 Bob Starbuck







15:30 - 17:00 **Session 1D-I4** (abstracts/résumés - page 124) **ECHA 1-498**

Statistics and Ecology

La statistique et l'écologie

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Louis-Paul Rivest (Université Laval)

- 15:30 - 16:00 Joanna Mills Flemming (Dalhousie University), Eva Cantoni (University of Geneva), Alan Welsh (Australian National University)
Modelling and Prediction for Clustered Count Data with Excess Zeros / Modélisation et prévision pour les données de comptage regroupé avec zéros excessifs  
- 16:00 - 16:30 Subhash Lele & Khurram Nadeem (University of Alberta)
Population Viability Analysis in the Presence of Observation Error: Consequences and Statistical Inference / Analyse de viabilité de la population en présence d'erreur d'observation : conséquences et inférence statistique  
- 16:30 - 17:00 Thierry Duchesne, Louis-Paul Rivest & Daniel Fortin (Université Laval)
Statistical Analysis of Animal Movement / Analyse statistique du déplacement animal  
-

15:30 - 17:00 **Session 1D-I5** (abstracts/résumés - page 126) **ECHA L1-430**

Stochastic Processes Arising in Environmental Sciences







Processus stochastiques dans les sciences environnementales

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

- 15:30 - 16:00 Don Dawson (Carleton University)
Some Probabilistic Objects Motivated by Evolutionary Biology and Ecology / Quelques objets probabilistes motivés par la biologie évolutionnaire et l'écologie  
- 16:00 - 16:30 Gail Ivanoff (University of Ottawa)
Renewal Processes in Two Dimensions / Processus de renouvellement en deux dimensions  
- 16:30 - 17:00 Reg Kulperger & Irena Creed (University of Western Ontario)
Trend Analysis of Water Yields in Northern Catchments / Analyse de la tendance des apports d'eau dans les bassins versants nordiques  

15:30 - 17:00 **Session 1D-C1** (abstracts/résumés - page 128) **EdSouth 228**

Analysis of Survival and Clinical Trials Data

Analyse de données de survie et d'essais cliniques

Contributed Paper Session / Séance de communications libres

Chair/Président: Richard Cook (University of Waterloo)

- 15:30 - 15:45 Jeff Bakal, Yinngan Zheng, Cynthia M. Westerhout & Paul W. Armstrong (University of Alberta)
Refining the Analysis of Composite Endpoints in Clinical Trials / Raffinage de l'analyse des critères d'évaluation des essais cliniques  
- 15:45 - 16:00 Pierre-Jérôme Bergeron (University of Ottawa), Vahid Partovi Nia (École Polytechnique de Montréal), Jaime Younger (Institute for Clinical Evaluative Sciences), Ewa Makvandi (University of Ottawa)
Analyzing Length-Biased Survival Data Using the R Package lbiassurv / L'analyse de données censurées à droite avec biais de longueur en R avec le package lbiassurv  
- 16:00 - 16:15 Keyue Ding (Queen's University), Zhoushanyue He (Univeristy of Waterloo), Jianfeng Xu (Queen's University)
Discovery of Pre-Post-Treatment Predictive Biomarker / Découverte de bio-marqueurs de prédiction avant et après traitement  
- 16:15 - 16:30 Adel Elghafghuf & Henrik Stryhn (University of Prince Edward Island)
A Cross-Classification and Multiple Membership Model Applied to Time-to-Event Data / Un modèle de classification croisée et d'appartenance multiple appliqué à des données de temps avant événement  
- 16:30 - 16:45 Gabriel Escarela (Universidad Autónoma Metropolitana - Iztapalapa)
Regression Analysis of Competing Risks with Missing Covariate Data / Analyse de régression pour les risques concurrents avec des données manquantes pour covariables  
- 16:45 - 17:00 Michael McIsaac & Richard Cook (University of Waterloo), Melanie Poulin-Costello (Amgen Inc)
Misspecified Models in Analyses of Incomplete Clinical Trial Data / Modèles mal spécifiés dans l'analyse de données incomplètes provenant d'essais cliniques  
-













15:30 - 17:00 **Session 1D-C2** (abstracts/résumés - page 131) **EdSouth 265**

Biostatistics - Applications and Case Studies

Biostatistique - applications et études de cas

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Wendy Lou (University of Toronto)

- 15:30 - 15:45 Tanja Hoegg, Sylvia Esterby & Paramjit Gill (University of British Columbia - Okanagan), Cynthia Araujo & Rasika Rajapakshe (BC Cancer Agency)
Breast Cancer in British Columbia - Identification of High Risk Women Based on Breast Cancer Risk Modelling / Cancer du sein en Colombie-Britannique - identification des femmes à haut risque par modélisation du risque de cancer du sein  
- 15:45 - 16:00 Mohammad Ehsanul Karim, Paul Gustafson, A. John Petkau, Afsaneh Shirani, Yinshan Zhao, Elaine Kingwell & Helen Tremlett (University of British Columbia)
A Simulation Study of Methods Used to Reduce Variability in the Inverse Probability of Treatment Weights for Marginal Structural Cox Models / Étude par simulation des méthodes utilisées pour réduire la variabilité de la probabilité inverse des pondérations de traitement dans les modèles de Cox structurels marginaux  
- 16:00 - 16:15 Prosanta Mondal, Lilian Thorpe & Hyun Lim (University of Saskatchewan)
Predictors for Nursing Home Placement (NHP) and Mortality Among Intellectually Disabled Adults With and Without Down Syndrome (DS): Application of Competing Risks Models / Prédicteurs du placement en maison de soins (PMS) et de la mortalité chez les adultes présentant des déficiences intellectuelles avec et sans syndrome de Down (SD) : application de modèles de risques concurrents  
- 16:15 - 16:30 Yawen Xu & Sicheng Wu (York University), Zeny Feng & Xiaojian Yang (University of Guelph), Xiaogang Wang & Xin Gao (York University)
Multiple-Platform Data Integration Method with Application to Combined Analysis of Microarray and Proteomic Data / Méthode d'intégration de données de différentes plateformes et application à l'analyse combinée de données provenant de biopuces et protéomiques  
- 16:30 - 16:45 Xuechen Zhang (University of Alberta)
Power Comparison of Discrete Data Methods with Repeated Measurement Data in Small Sample / Comparaison de la puissance de méthodes basées sur des mesures répétées discrètes avec des échantillons de petite taille  
- 16:45 - 17:00 Nicholas Mitsakakis (Toronto Health Economics and Technology Assessment Collaborative), Karen Bremner (University Health Network, Toronto), Murray Krahn (Toronto Health Economics and Technology Assessment Collaborative)
Methodological Challenges in Mapping a Disease Specific Psychometric Instrument to a Disease Specific Utility Instrument / Défis méthodologiques de la cartographie d'un instrument psychométrique spécifique à une maladie en un instrument utilitaire spécifique à une maladie  



15:30 - 16:45 **Session 1D-C3** (abstracts/résumés - page 135) **EdSouth 221**









Dependence I: Correlation and Copulas

Dépendance I : corrélation et copules

Contributed Paper Session / Séance de communications libres

Chair/Président: Bruno Rémillard (HEC Montréal)

- 15:30 - 15:45 Ayi Ajavon & François Perron (Université de Montréal)
Copulas Extensions for Non Continuous Random Variables / Extensions continues pour les sous copules des variables aléatoires discrètes  

- 15:45 - 16:00 Niroshan Withanage & Alexander R. de Leon (University of Calgary)
Joint Modeling of Hierarchically Clustered Non-Gaussian Continuous Outcomes via the Gaussian Copula / Modélisation conjointe de résultats continus non gaussiens en grappes hiérisées au moyen d'une copule gaussienne  
- 16:00 - 16:15 Yanling Cai, James V. Zidek & William Welch (University of British Columbia)
Breaking the Same Board Twice: the Magic of Statistics! / Casser la même planche deux fois : la magie de la statistique !  
- 16:15 - 16:30 Zihang Lu (University of Toronto), Jenna Sykes & Melania Pintilie (Ontario Cancer Institute, Princess Margaret Hospital)
The Bias for Sub-Hazard Ratio when the Event Times are Correlated / Biais pour le rapport de sous-risques lorsque les temps d'événements sont corrélés  
- 16:30 - 16:45 Chafik Bouhaddioui (UAE University)
Statistical Inference in Infinite-Order Cointegrated Vector Autoregressive Processes under Uncorrelated but Dependent Errors / Estimation des paramètres d'un processus autorégressif cointégré sous des erreurs non-corrélées mais dépendantes  

15:30 - 16:45 Session 1D-C4 (abstracts/résumés - page 138)











EdSouth 177

Estimation Methods

Méthodes d'estimation

Contributed Paper Session / Séance de communications libres

Chair/Président: Jose Garrido (Concordia University)

- 15:30 - 15:45 Golshid Chatrchi (Carleton University)
Robust Estimation of Variance Components in Small Area Estimation / Estimation robuste des composantes de la variance dans le contexte de l'estimation sur petits domaines  
- 15:45 - 16:00 Christopher A. Gravel (Carleton University), Palash Ghosh & Anup Dewanji (Indian Statistical Institute), Patrick Farrell (Carleton University), Daniel Krewski (University of Ottawa)
Analysis of Electronic Health Records for Active Pharmacovigilance with Misclassification of the Study Variable / Analyse de dossiers électroniques en pharmacovigilance avec erreur de classification pour la variable à l'étude  
- 16:00 - 16:15 Caren Hasler & Yves Tillé (Université de Neuchâtel)
Balanced k -nearest Neighbor Imputation / Imputation équilibrée par les k plus proches voisins  
- 16:15 - 16:30 Feng He & Grace Yi (University of Waterloo)
Analysis of Progressive Multi-State Models with Misclassified States / Analyse de modèles multi-états progressifs avec états mal classés  
- 16:30 - 16:45 Huijing Wang & X. Joan Hu (Simon Fraser University), Mary McBride & Maria Lorenzi (British Columbia Cancer Agency)
Analysis of Counts with Two Latent Classes / Analyse de dénombrements avec deux classes latentes  

15:30 - 16:30 Session 1D-C5 (abstracts/résumés - page 141)









EdSouth 206

Experimental and Sampling Design

Plan d'expérience et d'échantillonnage

Contributed Paper Session / Séance de communications libres

Chair/Président: Derek Bingham (Simon Fraser University)

- 15:30 - 15:45 Yin Li (University of Alberta), Yuanyuan Liang (University of Texas at San Antonio), Jing Wang & Keumhee C. Carriere (University of Alberta)
Multiple-Objective Response-Adaptive Repeated Measurement Designs in Clinical Trials for Binary Responses / Plan d'expérience multi-critères s'adaptant aux données pour les études cliniques comportant des mesures binaires répétées  
- 15:45 - 16:00 Michael A Rotondi (York University), Allan Donner (University of Western Ontario)
A Confidence Interval Approach to Sample Size Estimation for Interobserver Agreement Studies with Multiple Raters and Outcomes / Approche par intervalle de confiance pour l'estimation de la taille d'échantillon dans des études sur la concordance inter-observateurs à évaluateurs et résultats multiples  
- 16:00 - 16:15 Chris Wild (University of Auckland), John M. Neuhaus (University of California, San Francisco), Alastair J. Scott (University of Auckland)
Likelihood-Based Analysis of Longitudinal Data from Outcome-Dependent Sampling Designs / Analyse basée sur la vraisemblance de données longitudinales provenant de plans d'échantillonnage dépendant des résultats  
- 16:15 - 16:30 Yanqing Yi (Memorial University of Newfoundland)
The Most Powerful Test and the Optimal Design of Response Adaptive Clinical Trials / Test le plus puissant et plan optimal pour essais cliniques adaptatifs à la réponse  

17:00 - 18:00

KATZ 1-080

SSC Annual General Meeting
Assemblée générale annuelle de la SSC

18:00 - 20:00

Room at the Top (RATT)

Student BBQ - Free for graduate students
Barbecue des étudiants - Gratuit pour les étudiants de cycle supérieur

18:15 - 21:00

Katz Group Centre Atrium

New Investigator Dinner - Fee Event
Souper des nouveaux chercheurs - événement payant

18:15 - 21:00

Katz Group Centre Atrium

Old Strathcona Restaurant Night - Fee Event
Soirée restaurant dans le quartier Old Strathcona - événement payant

Tuesday May 28**mardi 28 mai**

8:45 - 9:45

Session 2A-A

(abstract/résumé - page 143)



CCIS 1-430**Gold Medal Address****Allocution du récipiendaire de la médaille d'or**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Bovas Abraham (University of Waterloo)

8:45 - 9:45

Rob Tibshirani (Stanford University)

The Lasso: A Review and Some Recent Developments / Le Lasso : une revue et certains développements récents  

9:45 - 10:30

Katz Group Centre Atrium**Coffee Break****Pause-café**

10:30 - 12:00

Session 2B-V

(abstract/résumé - page 144)



MedSci 231**Methods to Correct Measures of Effect for Bias Due to Exposure Measurement Error****Méthodes pour corriger la mesure des effets en tenant compte du biais dû aux erreurs de mesure d'exposition**

Introductory Overview Lecture / Conférence de vue d'ensemble

Organizer and Chair / Responsable et présidente: Debbie J. Dupuis (HEC Montréal)

10:30 - 12:00

Donna Spiegelman (Harvard School of Public Health)

Methods to Correct Measures of Effect for Bias Due to Exposure Measurement Error / Méthodes pour corriger la mesure des effets en tenant compte du biais dû aux erreurs de mesure d'exposition  

10:30 - 12:00

Session 2B-I1

(abstracts/résumés - page 145)

ECHA 1-190**Effective Use of Laboratory Sessions in Undergraduate Statistics Teaching****Du bon usage des séances de laboratoire dans les cours de statistique de premier cycle**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Kevin E. Thorpe (University of Toronto)

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique



10:30 - 11:00

W. John Braun (University of Western Ontario)

Naive Analysis of Variance / Analyse usuelle de la variance  



11:00 - 11:30

Bruce Dunham & Gaitri Yapa (University of British Columbia)

Using Group-Based Lab Activities in Undergraduate Statistics Teaching / Utilisation des activités de laboratoire en groupe dans l'enseignement de la statistique au premier cycle  

11:30 - 12:00

Jim Stallard (University of Calgary)

The Use of Small-Group Activities in a Post-Calculus, Mathematical Statistics Course / Les activités en petits groupes dans le cadre d'un cours de statistique mathématique dont les étudiants ont déjà suivi un cours de calcul différentiel et intégral  

10:30 - 12:00 **Session 2B-I2** (abstracts/résumés - page 147) **ECHA 1-498**







Functional Data Analysis

Analyse de données fonctionnelles

Invited Paper Session / Présentations sur invitation

Chair/Président: Duncan Murdoch (University of Western Ontario)

Organizer/Responsable: Jiguo (Jack) Cao (University of Western Ontario)

- 10:30 - 11:00 David Campbell (Simon Fraser University), Frederic Ferraty & Philippe Vieu (Université Paul Sabatier)
 Inferring Nonlinear Differential Equation Models from Non-Parameteric Functional Estimators / Comment inférer des modèles d'équations différentielles non linéaires à partir d'estimateurs fonctionnels non paramétriques  
- 11:00 - 11:30 Giles Hooker (Cornell University)
 Selecting the Domain of Integration in a Functional Linear Model / Sélection du domaine d'intégration dans un modèle fonctionnel linéaire  
- 11:30 - 12:00 James O. Ramsay (McGill University)
 Representing and Quantifying Complexity in Functional Observations and Parameters / La représentation et quantification de complexité des observations et paramètres fonctionnels  







10:30 - 12:00 **Session 2B-I3** (abstracts/résumés - page 149) **EdSouth 254**

Machine Learning and Data Analytics

Apprentissage automatique et analytique de données

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Ali Ghodsi (University of Waterloo)

- 10:30 - 11:00 Abbas Khalili (McGill University)
 Sparse Mixture-of-Experts Models in High-Dimensional Classification Problems / Modèles limités de mélange d'experts pour les problèmes de classification de grande dimension  
- 11:00 - 11:30 Csaba Szepesvari, Arash Afkanpour, Andras Gyorgy & Michael Bowling (University of Alberta)
 A Randomized Mirror Descent Algorithm for Large Scale Multiple Kernel Learning / Algorithme de descente en miroir randomisé pour l'apprentissage de noyaux multiples de grande échelle  
- 11:30 - 12:00 Andras Gyorgy (University of Alberta)
 Universal Prediction in Changing Environments / Prévission universelle dans des environnements changeants  

10:30 - 12:00 **Session 2B-I4** (abstracts/résumés - page 151) **ECHA L1-430**

Stochastic Models in Insurance





Modèles stochastiques en assurance



Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Sheldon Lin (University of Toronto)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe d'actuariat

- 10:30 - 11:00 Xiaowen Zhou (Concordia University)
 Lévy Risk Processes / Processus de risque de Lévy  
- 11:00 - 11:30 Christiane Lemieux, David Landriault & Gordon Willmot (University of Waterloo)
 Adaptive Premium Policies in Risk Theory / Taux de prime adaptatif en théorie du risque  

- 11:30 - 12:00 Taehan Bae (University of Regina)
Pricing a Motor Insurance Loss Rate Swaption / Établir le prix du taux de perte d'une option d'échange d'une assurance automobile  









10:30 - 11:30 Session 2B-C1 (abstracts/résumés - page 153) **EdSouth 228**

Modelling the Spread of Infectious Diseases

Modélisation de la propagation des maladies infectieuses

Contributed Paper Session / Séance de communications libres

Chair/Président: Patrick Brown (University of Toronto)

- 10:30 - 10:45 Hsing-Ming Chang & Rhonda J. Rosychuk (University of Alberta)
Simulation Study for a Spatial Scan Test Based on Compound Poisson Model / Étude de simulation pour un test de balayage spatial au moyen d'un modèle de Poisson composé  
- 10:45 - 11:00 Rob Deardon, Rajat Malik & Pui Sze Kwong (University of Guelph)
Parameterizing Individual-Level Models of Infectious Disease Spread Using Sampling-Based Likelihood Approximations / Établissement des paramètres de modèles au niveau de l'individu pour la propagation de maladies infectieuses par des approximations de la vraisemblance basées sur l'échantillonnage  
- 11:00 - 11:15 Ashok Krishnamurthy (Mount Royal University), Loren Cobb & Jan Mandel (University of Colorado), Jonathan Beezley (CERFACS)
Tracking Infectious Disease Epidemics Using Bayesian Data Assimilation / Suivi d'épidémies de maladies infectieuses à l'aide de l'assimilation de données bayésiennes  
- 11:15 - 11:30 Mahmoud Torabi (University of Manitoba)
Spatio-Temporal Modeling for Disease Mapping / Modélisation spatio-temporelle pour la cartographie de maladies  









10:30 - 11:30 Session 2B-C2 (abstracts/résumés - page 155) **ECHA L1-220**

Robust Estimation

Estimation robuste

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Julie Zhou (University of Victoria)













- 10:30 - 10:45 Sepideh Farsinezhad, Russell Steele & Masoud Asgharian (McGill University)
Doubly Robust G-Estimation for Time Varying Outcome via the Kalman Filter / G-Estimation doublement robuste au moyen du filtre de Kalman pour des résultats variant dans le temps  
- 10:45 - 11:00 Gun Ho Jang & Lakshmi Muthuswamy (Ontario Institute for Cancer Research)
Copy Number Variation Detection with Estimated Parameters Using Robust Distribution / Détection de la variation du nombre de copies avec paramètres estimés d'une distribution robuste  
- 11:00 - 11:15 Alexandre Leblanc, Stephane Durocher & Matthew Skala (University of Manitoba)
The Projection Median: A Multivariate Robust Adaptive Weighted Average / La médiane projective : une moyenne multivariée pondérée adaptative robuste  
- 11:15 - 11:30 Olli Saarela, Léo Raymond-Belzile & David A. Stephens (McGill University)
Can There be Bayesian Double Robustness? / La double robustesse bayésienne existe-t-elle?  

10:30 - 12:00 Session 2B-C3 (abstracts/résumés - page 157) **EdSouth 206**

Statistical Modeling**Modélisation statistique**

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Laura Cowen (University of Victoria)











- 10:30 - 10:45 Rashid Ahmed & Steve Brown (University of Waterloo)
Methods for Missing Data in Discrete Hierarchical Models / Méthodes pour les données manquantes dans les modèles hiérarchiques discrets  
- 10:45 - 11:00 Eshetu G. Atenafu (University Health Network, University of Toronto), Wendy Lou (University of Toronto)
Multistate Markov Models for Joint Modeling / Modèles de Markov multi-état pour modélisation conjointe  
- 11:00 - 11:15 Narges Nazeri Rad & Jerry Lawless (University of Waterloo)
Multi-State Models of Intermittently Observed Processes with Internal Covariates / Modèles multi-états de processus observés de façon intermittente avec covariables internes  
- 11:15 - 11:30 Arindam RoyChoudhury (Columbia University)
A Composite Likelihood Method for Estimating Phylogenetic Tree from Dependent Loci / Méthode d'estimation d'arbres phylogénétiques par vraisemblance composite à partir de loci dépendants  
- 11:30 - 11:45 Tyler Williamson (Queen's University), Misha Eliasziw (Tufts University), Gordon Fick (University of Calgary)
Log-Binomial Models: Concavity and Estimation / Modèles log-binomiaux : concavité et estimation  
- 11:45 - 12:00 Ximing Xu & Joanna Mills Flemming (Dalhousie University), Eva Cantoni (University of Geneva), Chris Field (Dalhousie University)
State-Space Modelling of Fish Stock Maturities / Modélisation à espace d'états de la maturité des stocks de poissons  


10:30 - 12:00 Session 2B-C4 (abstracts/résumés - page 160) **EdSouth 276**

Statistical Theory and Applications I**Statistique : théorie et applications I**

Contributed Paper Session / Séance de communications libres

Chair/Président: Richard Lockhart (Simon Fraser University)

- 10:30 - 10:45 Martin Lysy (University of Waterloo), Aleksander Labuda (Asylum Research)
Stochastic Simulation of a Nanoscale Experiment / Simulation stochastique à l'échelle nanoscopique  
- 10:45 - 11:00 Maria Elena Rivera Mancia (McGill University)
Bayesian Methods in the Study of Extremes / Méthodes bayésiennes pour l'étude des extrêmes  
- 11:00 - 11:15 Jeffrey Picka (University of New Brunswick)
Doubt, Science, and Statistics / Doute, science et statistique  
- 11:15 - 11:30 Susan Zhe Sheng (University of Western Ontario)
Univariate and Bivariate Polynomially Adjusted Saddlepoint Density / Densité en point de selle univariée et bivariée ajustée par polynôme  
- 11:30 - 11:45 Nana Xing (University of Calgary)
Application of Maximum Lq-Likelihood Method / Application de la méthode du maximum de vraisemblance-Lq  

11:45 - 12:00 Alwell Oyet (Memorial University)
 Estimation of a Lag 2 Dynamic Model with Measurement Error in Covariates / Estimation
 d'un modèle dynamique à décalage de deux périodes avec erreurs de mesure dans les co-
 variables 

12:00 - 13:30 **Lister Centre, Aurora room / salle Aurora**
Student Career Session
Séance carrières pour les étudiants

12:00 - 13:30 **ECHA 1-441**
Accreditation General Meeting
Assemblée générale des membres accrédités

12:00 - 18:00 **KATZ 2-155**
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 **ECHA 2-407**
Bilingualism Committee
Comité du bilinguisme

12:00 - 13:30 **ECHA 1-447**
CJS Editorial Board
Comité de rédaction de la RCS

12:00 - 13:30 **ECHA 1-131**
Census at School Committee
Comité de recensement à l'école

12:00 - 13:30 **ECHA 1-152**
Office Committee
Comité du bureau



12:00 - 13:30 **ECHA 1-144**
Public Relations Committee
Comité des relations publiques

12:00 - 13:30 **ECHA 1-148**
Research Committee
Comité de recherche

13:30 - 15:00 Session 2C-I1 (abstract/résumé - page 163) **ECHA 1-190**

An Introduction to Relative Survival Analysis
Introduction à l'analyse de survie relative



Invited Paper Session / Présentations sur invitation
 Organizer and Chair / Responsable et président: Russell Steele (McGill University)
 Session sponsored by the Biostatistics Section
 Séance parrainée par le Groupe de biostatistique



13:30 - 15:00 Noori Akhtar-Danesh (McMaster University)
 An Introduction to Relative Survival Analysis / Une introduction à l'analyse de survie relative  



13:30 - 15:00 Session 2C-I2 (abstracts/résumés - page 164) **EdSouth 254**

Design and Methodological Issues in Environmental Diseases
Problématiques liées à la méthodologie et aux plans d'expérience en matière de maladies liées à l'environnement

Invited Paper Session / Présentations sur invitation
 Organizer and Chair / Responsable et président: Lehana Thabane (McMaster University)

13:30 - 14:00 Judy-Anne Chapman (Queen's University)
 Personal Environment: Design and Analysis of Breast Cancer Trials / Environnement personnel : conception et analyse d'essais cliniques sur le cancer du sein  



14:00 - 14:30 Afisi Ismaila (GlaxoSmithKline & McMaster University)
 Design and Analyses Issues in Asthma and COPD Clinical Trials / Problèmes de conception et d'analyse dans les essais cliniques sur l'asthme et la MPOC  



14:30 - 15:00 Tim Ramsay & Lorelle Weiss (University of Ottawa), Tye Arbuckle & Mandy Fisher (Health Canada)
 Temporal Variability in Urinary Concentrations of Triclosan among Pregnant Women in Canada / Variabilité temporelle de la concentration urinaire de triclosan chez les femmes enceintes au Canada  



13:30 - 15:00 Session 2C-I3 (abstracts/résumés - page 166) **ECHA L1-430**

Environmental Sampling and Monitoring
Échantillonnage environnemental et surveillance

Invited Paper Session / Présentations sur invitation
 Chair/Président: Martin Lysy (University of Waterloo)
 Organizer/Responsable: Yulia R. Gel (University of Waterloo)
 Session sponsored by the Survey Methods Section
 Séance parrainée par le Groupe de méthodes d'enquête

13:30 - 14:00 Sylvia Esterby (University of British Columbia - Okanagan)
 Issues in the Extraction of Information from Water Quality Monitoring Data / Problèmes dans l'extraction de l'information provenant de données sur la surveillance de la qualité de l'eau  







14:00 - 14:30 Brian Gray and Vyacheslav Lyubchich (US Geological Survey / University of Waterloo)
 Regression Estimation of Trends in Temperature when Time and Date of Sampling are Haphazard / Estimation par la régression des tendances de température lorsque l'heure et la date de l'échantillonnage sont fortuites  

- 14:30 - 15:00 Carolyn Huston (CSIRO), Steve Thompson (Simon Fraser University)
Where? When? How? Why? Choosing a Sampling Strategy to Support a Bayesian/Complex Model / Où ? Quand ? Comment ? Pourquoi ? Choisir une stratégie d'échantillonnage pour supporter un modèle bayésien/complexe  

13:30 - 15:00 Session 2C-I4 (abstracts/résumés - page 168) **ECHA 1-498**

Skew Elliptical Distributions: The State of the Art
Les distributions asymétriques elliptiques : technologie de pointe







Invited Paper Session / Présentations sur invitation
Chair/Président: Christian Genest (McGill University)
Organizer/Responsable: C. J. Adcock (Sheffield University)

- 13:30 - 14:00 C. J. Adcock (Sheffield University)
The Multivariate Extended Normal-Gamma Distribution / Distribution normale - gamma étendue multivariée  
- 14:00 - 14:30 Adelchi Azzalini (Università di Padova), Giuliana Regoli (Università di Perugia)
Perturbation of Symmetry in Non-Standard Settings / Perturbation de la symétrie dans les situations non standard  
- 14:30 - 15:00 Marc G. Genton (King Abdullah University of Science and Technology), Yanyuan Ma & Mijeong Kim (Texas A&M University)
Semiparametric Efficient and Robust Estimation of the Center of an Unknown Symmetric Population under Arbitrary Sample Selection Bias / Estimation efficace et robuste semiparamétrique du centre d'une population symétrique inconnue sous biais de sélection d'échantillon arbitraire  

13:30 - 15:00 Session 2C-I5 (abstracts/résumés - page 170) **EdSouth 276**

Statistical Learning
Apprentissage statistique







Invited Paper Session / Présentations sur invitation
Chair/Président: Rob Tibshirani (Stanford University)
Organizer/Responsable: Ryan Tibshirani (Carnegie Mellon University)

- 13:30 - 14:00 Jacob Bien (Cornell University), Jonathan Taylor & Rob Tibshirani (Stanford University)
A Lasso for Hierarchical Interactions / Un lasso pour les interactions hiérarchiques  
- 14:00 - 14:30 Julian Wolfson (University of Minnesota)
Risk Prediction from Electronic Health Record Data: A Naive Bayes Approach / Préviation des risques à partir de dossiers de santé électroniques : une approche bayésienne naïve  
- 14:30 - 15:00 Ryan Tibshirani (Carnegie Mellon University)
Fast Stagewise Algorithms for Approximate Regularization Paths / Algorithmes par étapes rapides pour chemins de régularisation approximatifs  

13:30 - 15:00 Session 2C-I6 (abstracts/résumés - page 172) **EdSouth 206**

Statistics in Business and Finance
Statistique en gestion et finance

Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et président: Mark Reesor (University of Western Ontario)
Session sponsored by the Business and Industrial Statistics Section
Séance parrainée par le Groupe de statistique industrielle et de gestion

- 13:30 - 14:00 Adam Metzler (Wilfrid Laurier University)
Modeling Stochastic, State-Dependent Correlations / Modélisation des corrélations stochastiques, dépendantes de l'état  
- 14:00 - 14:30 Jesse Coull (ATB Financial)
Credit Risk and Stress Testing / Risque de crédit et test de stress  
- 14:30 - 15:00 Yuri Lawryshyn (University of Toronto)
Applied Research Program with Financial Institutions / Programme de recherche appliquée dans les institutions financières  

13:30 - 15:00 **Session 2C-C1** (abstracts/résumés - page 174) **EdSouth 228**

Probability II
Probabilité II





Contributed Paper Session / Séance de communications libres
Chair/Président: Hao Yu (University of Western Ontario)









- 13:30 - 13:45 Hassan Omid Firouzi (Université de Montréal), Hirbod Assa (Concordia University), Manuel Morales (Université de Montréal)
Capital Allocation Problem for a New Ruin-Based Coherent Risk Measure / Problème d'allocation de capital pour une nouvelle mesure cohérente du risque basée sur la ruine  
- 13:45 - 14:00 Clarence D. Kalitsi, Jan Vrbik & Wai Kong Yuen (Brock University)
Approximate Sampling Distributions of the Parameter Estimators in the AR(1)-Model / Distributions échantillonales approximées pour les estimateurs des paramètres dans un modèle AR(1)  
- 14:00 - 14:15 Jingya Li (University of Western Ontario), Adam Metzler (Wilfrid Laurier University), Mark Reesor (University of Western Ontario)
Valuation of Contingent Capital Bonds in First-Passage Structural Models / Évaluation d'obligations de capital conditionnel dans des modèles structurels de premier passage  
- 14:15 - 14:30 Vahed Maroufy & Paul Marriott (University of Waterloo)
Computational Aspects of Inference in Local Mixture Models / Aspects computationnels de l'inférence pour des modèles de mélanges locaux  
- 14:30 - 14:45 Harsha Perera & Tim Swartz (Simon Fraser University), Paramjit Gill (University of British Columbia - Okanagan)
Declaration Guidelines in Test Cricket / Directives pour la déclaration au Test cricket  
- 14:45 - 15:00 Jose Garrido, Ionica Groparu & Xiaowen Zhou (Concordia University)
The Finite-Time Gerber-Shiu Function as a Risk Measure / La fonction de Gerber-Shiu comme mesure de risque  

13:30 - 15:00 **Session 2C-C2** (abstracts/résumés - page 177) **ECHA L1-220**

Statistical Theory and Applications II
Statistique : théorie et applications II

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Mary Lesperance (University of Victoria)

- 13:30 - 13:45 Oksana Chkrebtii (Simon Fraser University), Erin Cameron (University of Alberta), David Campbell (Simon Fraser University), Erin Bayne (University of Alberta)
Transdimensional ABC for Inferring Introduction and Spread Rates of Invasive Species / Calculs bayésiens approximatifs transdimensionnels pour l'inférence des taux d'introduction et de propagation d'espèces envahissantes  
- 13:45 - 14:00 Camila de Souza & Nancy Heckman (University of British Columbia)
Switching Nonparametric Regression Models / Modèles de régression non paramétriques à changements d'états  

- 14:00 - 14:15 Shirin Golchi & Richard Lockhart (Simon Fraser University)
A Decision Theoretic Approach for Hypothesis Testing in Particle Physics / Approche décisionnelle théorique pour les tests d'hypothèses en physique des particules  
- 14:15 - 14:30 Jonathan Lee, Reg Kulperger & Hao Yu (University of Western Ontario)
Parallelization of Single Chain MCMC with Applications in Spatial Statistics / Parallélisation de MCMC à chaîne unique pour des applications en statistiques spatiales  
- 14:30 - 14:45 Joseph Francois Tagne Tatsinkou, Pierre Duchesne & Pierre Lafaye de Micheaux (Université de Montréal)
Smooth Test of Goodness-of-Fit of Normality for the Errors of an ARMA Process with Unknown Mean / Test lisse d'ajustement pour les erreurs d'un modèle ARMA avec moyenne inconnue  
- 14:45 - 15:00 Kasra Yousefi & Tim Swartz (Simon Fraser University)
Advanced Putting Metrics in Golf / Mesures avancées pour coups roulés au golf  

15:00 - 15:30
Coffee Break
Pause-café

Katz Group Centre Atrium

15:30 - 17:00 Session 2D-A (abstract/résumé - page 180) **MedSci 231**

Isobel Loutit Invited Address



Allocution invitée Isobel Loutit

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Executive committee of BISS & Mu Zhu (University of Waterloo)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 15:30 - 17:00 Yehuda Koren (Google Research)
Challenges in Recommender Systems: Science, Art and Pragmatics / Défis des systèmes recommandeurs : science, art et pragmatique  

15:30 - 17:00 Session 2D-I1 (abstracts/résumés - page 181) **ECHA L1-430**

Environmental and Ecological Quality Assessment





Évaluation de qualité environnementale et écologique



Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Abdel H. El-Shaarawi (American University in Cairo and the National Water Research Institute)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 15:30 - 16:00 Chris Field (Dalhousie University), Alan Welsh (Australian National University)
Estimating Bird Abundance: Approaches and Issues / Estimer l'abondance des oiseaux : approches et enjeux  
- 16:00 - 16:30 Jonah Keim (Matrix Solutions), Subhash Lele (University of Alberta)
Statistical Analysis of Camera Trap Data to Study the Effect of Human Activity on Resource Use Patterns by Wildlife: Spatial Recurrent Event Process and Inference / Analyse statistique des données recueillies à l'aide de caméras-pièges pour étudier l'effet de l'activité humaine sur les modèles d'utilisation des ressources par la faune : processus d'événements récurrents spatiaux et inférence  

15:30 - 17:00 Steve Thompson (Simon Fraser University)
Dynamic Spatial and Network Sampling / Échantillonnage spatial dynamique et par réseau  

15:30 - 17:00 Session 2D-I2 (description - page 183) **EdSouth 254**

Promoting Statistical Literacy - CensusAtSchool and Beyond
Promotion de la culture statistique - Recensement à l'école et au-delà



Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et président: John Brewster (University of Manitoba)
Session sponsored by the Statistical Education Committee
Séance parrainée par le Comité d'éducation en statistique



15:30 - 17:00 Alison Gibbs (University of Toronto)
15:30 - 17:00 Angela McCanny (Algonquin College, formerly Statistics Canada)
15:30 - 17:00 Roxy Peck (California Polytechnic State University)
15:30 - 17:00 Chris Wild (University of Auckland)



15:30 - 17:00 Session 2D-I3 (abstracts/résumés - page 184) **ECHA 1-190**

Recurrent Event Analysis
Analyse d'événements récurrents

Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et président: Pierre-Jérôme Bergeron (University of Ottawa)
Session sponsored by the Biostatistics Section
Séance parrainée par le Groupe de biostatistique

15:30 - 16:00 Cecilia Cotton (University of Waterloo)
Causal Inference with Recurrent Events / Inférence causale des événements récurrents  



16:00 - 16:30 Jerry Lawless (University of Waterloo)
Deconstructing Event History Data / Déconstruction des données de l'historique des événements  



16:30 - 17:00 Russell Steele (McGill University), Pierre-Jérôme Bergeron (University of Ottawa), Ian Shrier (Jewish General Hospital)
Comparing Approaches for Causal Inference for Recurrent Injury Data / Comparaison des approches d'inférence causale sur des données de blessures récurrentes  


15:30 - 17:00 Session 2D-I4 (abstracts/résumés - page 186) **ECHA 1-498**

Stochastic Models with Applications to Environmental Problems
Modèles stochastiques avec applications aux problèmes environnementaux

Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et président: Hao Yu (University of Western Ontario)
Session sponsored by the Probability Section
Séance parrainée par le Groupe de probabilité

15:30 - 16:00 Priscilla Greenwood (University of British Columbia), Mark McDonnell (University of South Australia), Lawrence Ward (University of British Columbia)
Dynamics of Gamma Bursts in Local Field Potentials / Dynamiques des éclatements gamma dans les potentiels locaux de champs  

16:00 - 16:30 Yuliya Martsynyuk (University of Manitoba)
Consistent Estimators without Identifiability Conditions in Linear Functional and Structural Errors-in-Variables Models / Estimateurs cohérents sans conditions d'identifiabilité dans des modèles linéaires fonctionnels et structurels d'erreurs dans les variables  

16:30 - 17:00 Natalia Stepanova (Carleton University), Yuri Ingster (Yuri Ingster passed away in Aug. 2012. He was a Professor at St. Petersburg State Electrotechnical University.)
Adaptive Variable Selection in Nonparametric Sparse Regression / Sélection adaptative des variables d'une régression parcimonieuse non-paramétrique 

17:00 - 18:00 **Katz Group Centre, room 1-080**
NSERC Discovery Grant Application Workshop
Ateliers sur les demandes de subventions à la découverte du CRSNG

17:00 - 18:00 **ECHA L1-230**
Actuarial Science Section General Meeting
Groupe d'actuariat assemblée générale

17:00 - 18:00 **ECHA L1-140**
Biostatistics Section General Meeting
Groupe de biostatistique assemblée générale

17:00 - 18:00 **Lister Centre, Evergreen room**
Business and Industrial Statistics Section General Meeting
Groupe de statistique industrielle et de gestion assemblée générale

17:00 - 18:00 **ECHA 1-498**
Probability Section General Meeting
Groupe de probabilité assemblée générale

17:00 - 18:00 **ECHA L1-150**
Statistical Education Section General Meeting
Groupe d'éducation en statistique assemblée générale

17:00 - 18:00 **ECHA 2-407**
Student Research Presentation Award Committee
Comité du prix pour la présentation de recherche étudiante

17:00 - 18:00 **ECHA L1-220**
Survey Methods Section General Meeting
Groupe de méthodes d'enquête assemblée générale

18:15 - 19:05 **Blatchford Hangar, Fort Edmonton Park**
Banquet - Cocktails and Cash Bar
Banquet - Bar payant



19:05 - 22:00 **Blatchford Hangar, Fort Edmonton Park**
Banquet - Dinner Service
Banquet - Service du repas

Wednesday May 29**mercredi 29 mai**

8:30 - 9:30 **Session 3A-A** (abstract/résumé - page 188) **MedSci 231**

CRM-SSC Prize in Statistics Address
Allocution du récipiendaire du Prix CRM-SSC

Special Session / Conférence spéciale
 Organizer and Chair / Responsable et président: David A. Stephens (McGill University)

8:30 - 9:30 Derek Bingham (Simon Fraser University)
 Prediction and Extrapolation Using Calibrated Computer Models / Prédiction et extrapolation
 à l'aide de modèles informatiques calibrés  



9:30 - 10:00 **Katz Group Centre Atrium**
Coffee Break
Pause-café

9:30 - 10:00 **Katz Group Centre, second floor Atrium**
Statistics Coffee Time
Pause café Statistique

10:00 - 11:30 **Session 3B-V** (abstract/résumé - page 189) **MedSci 231**

Ranking and Selection in High-Dimensional Inference
Ordonnement et sélection dans l'inférence pour les données de grande dimension



Introductory Overview Lecture / Conférence de vue d'ensemble
 Organizer and Chair / Responsable et présidente: Debbie J. Dupuis (HEC Montréal)

10:00 - 11:30 Michael Newton & Nicholas Henderson (University of Wisconsin - Madison)
 Ranking and Selection in High-Dimensional Inference / Ordonnement et sélection dans
 l'inférence pour les données de grande dimension  



10:00 - 11:30 **Session 3B-I1** (abstracts/résumés - page 190) **ECHA 2-150**

Bootstrap and Change-Point Analysis with Applications
Bootstrap et analyse du point de rupture avec applications

Invited Paper Session / Présentations sur invitation
 Organizer and Chair / Responsable et président: Murray Burke (University of Calgary)
 Session sponsored by the Probability Section
 Séance parrainée par le Groupe de probabilité

10:00 - 10:30 Edit Gombay (University of Alberta), Abdulkadir Hussein (University of Windsor), Konstantinos
 Fokianos (University of Cyprus)
 Change Detection for Binary Time Series / Détection du changement dans les séries tempo-
 relles binaires  

10:30 - 11:00 Miklos Csorgo (Carleton University)
 Random Walking around the Bootstrap I / Marche aléatoire autour du bootstrap I  

11:00 - 11:30 Masoud Nasari (Carleton University)
 Random Walking around the Bootstrap II / Marche aléatoire autour du bootstrap II  

10:00 - 11:30 Session 3B-I2 (abstracts/résumés - page 192) **ECHA L1-430**







Methods for Cross-Country and Cross-Cultural Surveys
Méthodes pour les enquêtes transnationales et transculturelles

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Christian Boudreau (University of Waterloo)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 10:00 - 10:30 Jean Dumais (Statistics Canada), Marc Joncas & Sylvie LaRoche (Statistique Canada)
 Standardisation of Sampling Designs and Quality Assurance in International Comparative Studies / Standardisation des plans de sondage et assurance de la qualité dans les enquêtes comparatives  
- 10:30 - 11:00 Celia Huang, Dagmar Mariaca Hajducek & Mary Thompson (University of Waterloo)
 Cross-Country Comparisons in International Surveys: Examples Using Frequentist and Bayesian Approaches / Comparaisons entre les pays dans les enquêtes internationales : exemples employant les approches fréquentistes et bayésiennes  
- 11:00 - 11:30 Beth-Ellen Pennell (University of Michigan)
 Toward Best Practice in Cross-National Data Collection: Development of Guidelines / Vers des pratiques exemplaires dans la collecte de données transnationales : développement de lignes directrices  

10:00 - 11:30 Session 3B-I3 (abstracts/résumés - page 194) **ECHA 1-498**





Recommender Systems
Systèmes de recommandation

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Xiaotong Shen (University of Minnesota)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion





- 10:00 - 10:45 Mu Zhu (University of Waterloo), Peter Forbes (University of Oxford), Jennifer Nguyen (University College London)
 Content-Boosted Matrix Factorization Techniques for Recommender Systems / Techniques de factorisation de matrices à contenu amélioré dans les systèmes recommandeurs  
- 10:45 - 11:30 Yunzhang Zhu, Xiaotong Shen & Changqing Ye (University of Minnesota)
 Partial Latent Models and Sparse Over-Complete Factorizations with Missing Values / Modèles latents partiels et factorisations parcimonieuses sur-complètes avec données manquantes  




10:00 - 11:30 Session 3B-C1 (abstracts/résumés - page 196) **EdSouth 276**

Dependence II
Dépendance II

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Johanna Nešlehová (McGill University)

- 10:00 - 10:15 Hela Romdhani (McGill University), Louis-Paul Rivest & M'hamed Lajmi Chaieb-Lakhal (Université Laval)
 Kendall's tau for hierarchical data / Tau de Kendall pour des données hiérarchiques  
- 10:15 - 10:30 Liqun Diao, Richard Cook & Ker-Ai Lee (University of Waterloo)
 A New Copula Model for Marked Point Processes / Nouveau modèle de copule pour les processus ponctuels marqués  

- 10:30 - 10:45 Méлина Mailhot (Concordia University), Étienne Marceau (Université Laval), Mhamed Mesfioui (Université du Québec à Trois-Rivières)
Bivariate Value-at-Risk and Tail Value-at-Risk / VaR et TVaR bivariées  
- 10:45 - 11:00 Wayne Oldford & Derek Wang (University of Waterloo)
Visual Hypothesis Tests of Independence / Tests d'indépendance visuels  
- 11:00 - 11:15 Antonio Punzo (Università di Catania), Luca Bagnato (Università Cattolica del Sacro Cuore), Lucio De Capitani (Università di Milano-Bicocca)
Improving the Autodependogram Using the Kullback-Leibler Divergence / Amélioration de l'autodépendogramme à l'aide de la divergence de Kullback-Leibler  
- 11:15 - 11:30 Jules Joseph S. de Tibeiro (Université de Moncton), Bernard Colin (Université de Sherbrooke)
Optimal Quantization: An Information Theory Approach / Le codage optimal : un point de vue de la théorie de l'information  

10:00 - 11:30 Session 3B-C2 (abstracts/résumés - page 199) **EdSouth 228**

Estimation and Hypothesis Testing

Estimation et tests d'hypothèses

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Joan Hu (Simon Fraser University)

- 10:00 - 10:15 Mohamed Amezziane (DePaul University), Syed Ejaz Ahmed (Brock University)
The Limiting Distribution of a Class of Shrinkage Estimators and Related Inference / La distribution asymptotique des estimateurs obtenus par rétrécissement  
- 10:15 - 10:30 Song Cai & Jiahua Chen (University of British Columbia)
Dual Empirical Likelihood Ratio Test under Semi-Parametric Density Ratio Models for Multiple Samples / Test dual du rapport de vraisemblances empirique sous les modèles de rapport de densité semi-paramétriques pour échantillons multiples  
- 10:30 - 10:45 Shakhawat Hossain & Melody Ghahramani (University of Winnipeg)
Shrinkage Estimation in Linear Regression with AR(m)-GARCH(p,q) Errors / Estimation par rétrécissement dans des modèles de régression linéaire avec erreurs AR(m)-GARCH(p,q)  
- 10:45 - 11:00 Yingli Qin (University of Waterloo), Weiming Li (Beijing University of Posts and Telecommunications)
Hypothesis Testing for Large Dimensional Covariance Matrices / Tests d'hypothèses pour des matrices de covariance de grande taille  
- 11:00 - 11:15 Joel Sango & Pierre Duchesne (Université de Montréal)
Modelization of Multivariate Cointegrated Time Series / Modélisation des séries chronologiques multivariées et cointégrées  
- 11:15 - 11:30 Ekaterina Tuzov & Yuliya Martsynyuk (University of Manitoba)
Exploring Functional CLT Confidence Intervals for a Population Mean in the Domain of Attraction of the Normal Law / Exploration d'intervalles de confiance du TCL fonctionnel pour la moyenne d'une population dans le domaine d'attraction de la loi normale  









10:00 - 11:00 Session 3B-C3 (abstracts/résumés - page 202) **EdSouth 206**

Models Accounting for Measurement Error and Meta-Analysis

Modèles avec erreurs de mesure et méta-analyse

Contributed Paper Session / Séance de communications libres

Chair/Président: Duncan Murdoch (University of Western Ontario)

- 10:00 - 10:15 Mireya Diaz (Henry Ford Hospital)
Performance of Multivariate Logistic Regression for Meta-Analysis of Diagnostic Accuracy / Performance de la régression logistique multivariée pour la méta-analyse de la précision diagnostique  
- 10:15 - 10:30 Maria Esther Perez Trejo & Robert Platt (McGill University)
Statistical Modeling of Outcomes Subject to Measurement Error Induced by Double Clustering via a Bayesian Hierarchical Model / La modélisation statistique des réponses exposées à une erreur de mesure induite pour double congglomération avec un modèle hiérarchique bayésien  
- 10:30 - 10:45 Ying Yan & Grace Yi (University of Waterloo)
Corrected Profile Likelihood for Cox Model with Covariate Measurement Error / Profil de vraisemblance corrigé pour le modèle de Cox avec erreur de mesure des covariables  
- 10:45 - 11:00 Qiaohao Zhu (Alberta Health, Government of Alberta), Kimmie C. Chough (University of Alberta)
Detecting and Correcting Publication Bias in Meta-Analysis: A Maximum Likelihood Approach / Détection et correction du biais de publication dans les méta-analyses : une méthode basée sur le maximum de vraisemblance  

11:30 - 13:00

ECHA 1-447

Accreditation Committee
Comité d'accréditation

11:30 - 13:00

Katz 2-156

Biostatistics Section Executive Committee 2
Comité exécutif de biostatistique industrielle et de gestion 2

11:30 - 13:00

ECHA 1-441

Program Committee
Comité du programme

11:30 - 13:00

ECHA 2-407

Treasurers
Trésoriers

13:00 - 14:30

Session 3C-A

(abstracts/résumés - page 204)

MedSci 231**SSC Impact Award Address****Allocution du récipiendaire du prix pour impact de la SSC**

Special Session / Conférence spéciale



Organizer and Chair / Responsable et présidente: Georgia Roberts (Statistics Canada)

Session in honour of the 2012 SSC Impact Award winner, David Binder

Séance en l'honneur du récipiendaire du prix pour impact de la SSC 2012, David Binder



13:00 - 13:30



John Eltinge (Bureau of Labour Statistics)

Variance Function Models for Complex Sample Surveys / Modèles de fonction de la variance dans les enquêtes complexes  

13:30 - 14:00

J.N.K. Rao (Carleton University)

Weighted Estimating Equations Approach to Inference from Complex Survey Data: Impact of David Binder / Approche des équations estimantes pondérées à l'inférence de données d'enquêtes complexes : impact de David Binder  

14:00 - 14:30 Mary Thompson (University of Waterloo)
Measuring Causal Effects With Surveys / Mesure des effets causaux dans les enquêtes  

13:00 - 14:30 Session 3C-I1 (description - page 206) **EdSouth 254**

Issues and Challenges Faced by New Researchers

Problèmes et défis auxquels font face les nouveaux chercheurs

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Cecilia Cotton (University of Waterloo)

Session sponsored by the Committee on New Investigators

Séance parrainée par le Comité des nouveaux chercheurs

13:00 - 14:30 Shoja'eddin Chenouri (University of Waterloo)

13:00 - 14:30 Katherine Davies (University of Manitoba)

13:00 - 14:30 Richard Lockhart (Simon Fraser University)

13:00 - 14:30 Session 3C-I2 (abstracts/résumés - page 207) **ECHA 1-498**

Resources of Energy and Environmental Issues



Ressources en énergie et enjeux environnementaux



Invited Paper Session / Présentations sur invitation



Organizer and Chair / Responsable et président: Pritam Ranjan (Acadia University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

13:00 - 13:30 Derek Bingham & Shirin Golchi (Simon Fraser University), Chipman Hugh (Acadia University)
Monotone Computer Model Emulation / Émulation d'un modèle informatique monotone  

13:30 - 14:00 Vera Bulaevskaya, Sonia Wharton, Matthew Simpson, Don Lucas & Wayne Miller (Lawrence Livermore National Laboratory)
Statistical Methods for Wind Power Forecasting / Méthodes statistiques pour prédire l'énergie éolienne  

14:00 - 14:30 Richard Karsten (Acadia University), Tracey Kutney (Natural Resources Canada), Brain Polagye (University of Washington)
Quantifying the Uncertainty in Predicting Tidal Power Resources / Quantifier l'incertitude dans la prévision des ressources d'énergie marémotrice  

13:00 - 14:30 Session 3C-I3 (abstracts/résumés - page 209) **ECHA 2-150**

Statistical Methods for Modeling Biology and Ecology Processes

Méthodes statistiques pour la modélisation des processus en biologie et en écologie



Invited Paper Session / Présentations sur invitation



Chair/Présidente: Llwellyn Armstrong (Ducks Unlimited Canada)



Organizer/Responsable: Ying Zhang (Acadia University)

Session sponsored by the Committee on Women in Statistics

Séance parrainée par le Comité de promotion de la femme en statistique

13:00 - 13:30 Lihui Liu & Subhash Lele (University of Alberta)
Statistical Analysis of Serial Dilution Assays Using Estimating Functions and Data Cloning / Analyse statistique d'essais de dilution en série par fonctions d'estimation et clonage de données  

13:30 - 14:00 Ying Zhang, Jing Ding & Paul Cabilio (Acadia University)
Modelling Age-Specific Harvest Data / Modélisation des données sur les prises selon l'âge  

14:00 - 14:30 Liwen Zou (Memorial University), Eric Stone & Jeffrey Thorne (North Carolina State University)
Incorporating Human Disease Data into Models of Molecular Evolution / Inclusion de données
sur la maladie humaine dans les modèles d'évolution moléculaire  

13:00 - 14:30 Session 3C-I4 (abstracts/résumés - page 211) **ECHA L1-430**

Targeted Maximum Likelihood Estimation (TMLE) and Related Methods
Estimation ciblée par maximum de vraisemblance (TMLE) et méthodes s'y rattachant



Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Mireille Schnitzer (McGill University)



Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique



13:00 - 13:30 Susan Gruber (Harvard School of Public Health), Mark van der Laan (University of California, Berkeley)

Triple-Robust Targeted Minimum Loss-Based Estimation / Estimation triple robuste ciblée de la perte minimale  

13:30 - 14:00 Ori Stitelman, Claudia Perlich & Brian Dalessandro (Media6Degrees), Foster Provost (NYU - Stern School Of Business)

The Challenges of Estimating the Causal Effect of Online Display Advertising / Les défis de l'estimation de l'effet causal dans l'affichage publicitaire en ligne  

14:00 - 14:30 Mark van der Laan (University of California, Berkeley)

Targeted Minimum Loss-Based Estimation / Estimation ciblée de la perte minimale  

13:00 - 14:00 Session 3C-C1 (abstracts/résumés - page 213) **EdSouth 228**



Probability in Actuarial Science and Finance



Probabilité en actuariat et en finance



Contributed Paper Session / Séance de communications libres

Chair/Président: Don McLeish (University of Waterloo)

13:00 - 13:15 Alexander Melnikov & Hao Li (University of Alberta)
Orthogonal Polynomials and their Applications in Financial and Actuarial Modeling / Polynômes orthogonaux et leurs applications en modélisation financière et actuarielle  

13:15 - 13:30 Shanoja Naik & Bovas Abraham (University of Waterloo)
The Fractional Diffusion Equation and a New Distribution with Heavy Tails / Équation de diffusion fractionnelle et nouvelle distribution à ailes lourdes  

13:30 - 13:45 Ruodu Wang (University of Waterloo)
Completely Mixable Distributions / Distributions complètement mélangeables  

13:45 - 14:00 Taehan Bae (University of Regina), Ian Iscoe (IBM), Changki Kim (Korea University)
Valuing Retail Credit Tranches under Conditional Dependence / Valorisation de tranches de crédit de détail sous l'hypothèse de dépendance conditionnelle  



13:00 - 14:30 Session 3C-C2 (abstracts/résumés - page 215) **EdSouth 206**

Variable Selection, Clustering and Classification

Sélection de variables, partitionnement et classification

Contributed Paper Session / Séance de communications libres

Chair/Président: Jean-François Plante (HEC Montréal)

13:00 - 13:15 Folly Adjogou & Alejandro Murua (Université de Montréal)
Functional Model-Based Clustering for Longitudinal Data / Analyse de données longitudinales selon une approche fonctionnelle de la classification non supervisée  

- 13:15 - 13:30 Anne-Sophie Charest (Université Laval), Russell Steele (McGill University)
Imputation for Nonresponse via Clustering of High-Dimensional Categorical Variables / Imputation pour la non-réponse via un classement basé sur plusieurs variables catégoriques  
- 13:30 - 13:45 Kevin Keen (University of British Columbia), William Petrlich (University of Northern British Columbia), Stephan van Eeden & James Dunne (University of British Columbia)
Bumping Bagged Classification Trees and Random Forests / Bumping d'arbres de décision et de forêts aléatoires agrégés par bootstrap  
- 13:45 - 14:00 Maciak Matus & Ivan Mizera (University of Alberta)
Change in Behaviour with Lasso Methods / Changement de comportement avec les méthodes Lasso  
- 14:00 - 14:15 Paula Murray, Richard Browne & Paul McNicholas (University of Guelph)
Model-Based Clustering Using Mixtures of Skew-t Factor Analyzers / Modèle de regroupement en classes au moyen de mélanges d'analyseurs de facteurs t asymétriques  
- 14:15 - 14:30 Monica H.T. Wong & Paul McNicholas (University of Guelph)
Parsimonious Biclustering of Gene Expression Data with Applications in Nutrigenomics / Classification double parcimonieuse de données d'expression génique avec applications en nutriginomique  

14:30 - 15:00**Katz Group Centre Atrium****Coffee Break****Pause-café****15:00 - 16:00****Session 3D-A1**

(abstract/résumé - page 218)

EdSouth 254**CJS Award Address****Allocution du récipiendaire du Prix de la RCS**



Special Session / Conférence spéciale

Chair/Président: David A. Stephens (McGill University)

Organizer/Responsable: Denis Larocque (HEC Montréal)

15:00 - 16:00

Jing Qin (National Institute of Allergy and Infectious Diseases), Guan Zhong (Indiana University South Bend), Biao Zhang (University of Toledo)

Information Borrowing Methods for Covariate-Adjusted ROC Curve / Méthodes d'extraction d'information pour courbes ROC ajustées à l'effet de covariables  **15:00 - 16:00****Session 3D-A2**

(abstract/résumé - page 219)



MedSci 231**Pierre Robillard Award Address****Allocution du récipiendaire du Prix Pierre-Robillard**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et présidente: Johanna Nešlehová (McGill University)

15:00 - 16:00

Luke Bornn (Harvard University)









Nonstationary Modeling Through Dimension Expansion / Modélisation non stationnaire par l'agrandissement de la dimension  **15:00 - 16:00****Session 3D-C1**

(abstracts/résumés - page 220)

EdSouth 276**Analysis of Longitudinal Data****Analyse de données longitudinales**

Contributed Paper Session / Séance de communications libres









Chair/Président: Joel Dubin (University of Waterloo)

- 15:00 - 15:15 Cynthia Struthers & Don McLeish (University of Waterloo)
A Multivariate Diffusion Model for AIDS Data / Modèle de diffusion multivarié pour des données sur le SIDA  
- 15:15 - 15:30 Mireille Schnitzer, Judith J. Lok & Ronald J. Bosch (Harvard School of Public Health)
Determining the Predictors for Negative HIV Outcomes under a Suppressive ART Regime / Détermination des prédicteurs de résultats négatifs du VIH sous un régime de traitement antirétroviral suppressif  
- 15:30 - 15:45 Eleanor M Pullenayegum (McMaster University)
Analysis of Longitudinal Data Subject to Irregular Follow-Up: Diagnosing and Correcting for Small-Sample Bias in Inverse-Intensity Weighted Estimating Equations / Analyse de données longitudinales sujettes à un suivi irrégulier : diagnostic et correction de biais dans des équations d'estimation pondérées par l'intensité inverse pour de petits échantillons  
- 15:45 - 16:00 Jemila Hamid & WeiLiang Huang (McMaster University)
Residuals in the Growth Curve Model and their Application in the Analysis of Longitudinal Data / Les résidus dans le modèle de courbe de croissance et application de ceux-ci dans l'analyse de données longitudinales  

15:00 - 16:00 **Session 3D-C2** (abstracts/résumés - page 222) **EdSouth 221**

Business and Industrial Statistics II
Statistique industrielle et de gestion II



Contributed Paper Session / Séance de communications libres
Chair/Président: Fernando Camacho (Damos)







- 15:00 - 15:15 Amber Corkum, Ying Zhang & Paul Cabilio (Acadia University)
Confidence Interval Construction for Short Time Series with Applications in Administrative Health Databases / Construction de l'intervalle de confiance pour les séries chronologiques courtes et applications dans des bases de données administratives dans le domaine de la santé  
- 15:15 - 15:30 Veeresh Gadag & Asokan Mulayath Variyath (Memorial University), Uditha Balasooriya (Nanyang Technical University)
On Tests for Multiple Contaminants in Shocks Model and Labelled Slippage Models / Tests pour des modèles de chocs et de glissement marqué avec plusieurs contaminants  
- 15:30 - 15:45 Tom Loughin (Simon Fraser University Surrey), Andrew Henrey (Simon Fraser University)
Joint Model Selection for Location and Dispersion Effects in Un-Replicated Factorial Experiments / Sélection de modèle conjoint pour les effets de localisation et de dispersion dans des expériences factorielles non répétées  
- 15:45 - 16:00 Xiang Yan & Mei Ling Huang (Brock University)
On Confidence Intervals of High Quantiles / Sur les intervalles de confiance pour quantiles supérieures  

15:00 - 16:00 **Session 3D-C3** (abstracts/résumés - page 224) **EdSouth 228**

Competing Risks and Disease Progression
Risques concurrents et progression des maladies

Contributed Paper Session / Séance de communications libres
Chair/Président: Thierry Duchesne (Université Laval)

- 15:00 - 15:15 Yun-Hee Choi (Western University)
On Combining Population-Based and Clinic-Based Time-to-Event Family Data in Disease Risk Estimation / Sur la combinaison de données de temps avant événement familiales pour les populations cliniques et temporelles dans l'estimation du risque de maladies  

- 15:15 - 15:30 Utkarsh Dang & Paul McNicholas (University of Guelph)
Cluster-Weighted Modeling for Time-to-Event Data / La modélisation par pondération de grappes pour des données de temps avant événement  
- 15:30 - 15:45 Yan Yan Wu & Laurent Briollais (Samuel Lunenfeld Research Institute, Mount Sinai Hospital), Augustine Wong (York University)
An Accurate Method for Testing Genetic Heterogeneity in Families from Studies of Rare Genetic Variants / Une méthode précise pour tester l'hétérogénéité génétique dans des familles à partir d'études de variantes génétiques rares  
- 15:45 - 16:00 Ying Wu & Richard Cook (University of Waterloo)
Selection of Genetic Risk Factors for Times of Disease Progression Subject to Interval-censoring / Sélection des facteurs de risque génétique pour les durées de progression de maladie censurées par intervalles  









15:00 - 16:00 **Session 3D-C4** (abstracts/résumés - page 227) **EdSouth 213**

Mixture Models

Modèles de mélange

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Anne-Sophie Charest (Université Laval)

- 15:00 - 15:15 Katherine Morris & Paul McNicholas (University of Guelph)
Non-Gaussian Clustering and Dimension Reduction / Regroupement non gaussien et réduction de dimension  
- 15:15 - 15:30 Jeffrey L. Andrews, Paul McNicholas & Ryan Browne (University of Guelph)
Robust Latent Variable Mixture Models for Clustering Data of Mixed Type / Modèles de mélanges robustes à variables latentes pour des données en grappes de type mixte  
- 15:30 - 15:45 Mohammad Jafari Jozani & Armin Hatefi (University of Manitoba), Omer Ozturk (Ohio State University)
Mixture Model Analysis of Partially Rank Ordered Set Samples: Estimating the Age-Groups of Fish from Length-Frequency Data / Analyse de modèles de mélanges pour des échantillons d'ensembles partiellement ordonnés : estimation des groupes d'âge de poissons à partir de la fréquence des longueurs  
- 15:45 - 16:00 Tasnima Abedin & Jingjing Wu (University of Calgary)
A Study on Two-Sample Mixture Model / Étude sur un modèle de mélange à deux échantillons  





15:00 - 16:00 **Session 3D-C5** (abstracts/résumés - page 229) **EdSouth 206**





Screening and Diagnostic Tests

Examens préventifs et tests de diagnostique

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Cecilia Cotton (University of Waterloo)

- 15:00 - 15:15 Amay Cheam & Paul McNicholas (University of Guelph)
Modelling Receiver Operating Characteristics (ROC) Curve Using Gaussian Mixture / Modélisation de la courbe des caractéristiques receveur-opérateur (ROC) avec le modèle de mélange gaussien  
- 15:15 - 15:30 James Hanley & Zhihui (Amy) Liu (McGill University)
Statistical Animations in Understanding Mortality Reductions Due to Cancer Screening / Animations statistiques dans la compréhension des réductions de mortalité grâce au dépistage du cancer  

- 15:30 - 15:45 Khokan C. Sikdar, Yutaka Yasui & Marcy Winget (University of Alberta)
Diagnostic Intervals of Colorectal Cancer in Alberta: A Population-Based Study / Intervalles diagnostiques pour le cancer colorectal en Alberta : une étude basée sur la population  
- 15:45 - 16:00 Wanhua Su (Grant MacEwan University), Mu Zhu (University of Waterloo), Yan Yuan (University of Alberta)
Average Precision Analysis for Screening and Diagnostic Tests / Analyse de précision moyenne pour les tests de dépistage et de diagnostic  








15:00 - 16:00 Session 3D-C6 (abstracts/résumés - page 231) **EdSouth 265**

Statistics: Applications and Case Studies

Statistique : applications et études de cas

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Ying Zhang (Acadia University)

- 15:00 - 15:15 David Ross Brillinger (University of California Berkeley)
Aligning Some Nicholson Sheep-Blowfly Data Sets with System Input / Aligner quelques jeux de données de Nicholson sur la mouche à viande aux intrants du système  
- 15:15 - 15:30 Theodoro Koulis & Saman Muthukumarana (University of Manitoba)
A Bayesian Stochastic Model for Batting Performance Evaluation in One-Day Cricket / Un modèle stochastique bayésien pour l'évaluation de la performance au bâton dans le cricket d'une journée  
- 15:30 - 15:45 Tulsı Paudel (University of Guelph), Dawn M. VanLeeuwen & Betsy Bender Stringam (New Mexico State University)
Iterative Model Building Using Regression Diagnostics and Mean Shift Parameters in Logistic Regression for the Analysis of a Large Hotel Rating Data Set / Construction d'un modèle itératif par des diagnostics de régression et de paramètres de décalage moyen dans la régression logistique pour l'analyse d'un grand ensemble de données de cotation d'hôtels  
- 15:45 - 16:00 Mark Reesor & Almas Naseem (University of Western Ontario)
Risk Analysis of the Smith Manoeuvre for Re-advanceable Canadian Mortgages / Analyse des risques de la manœuvre de Smith pour les prêts sur la valeur nette des maisons canadiennes  

16:15 - 17:30 **ECHA 4-036**

Board of Directors 2

Conseil d'administration 2

17:30 - 18:30 **ECHA 4-036**

Executive Committee 2

Comité exécutif 2

Abstracts • Résumés

1A-A2: SSC Presidential Invited Address
Allocution de l'invitée du Président de la SSC

Organizer and Chair / Responsable et président: Christian Léger (Université de Montréal)

Room/Salle: CCIS 1-430

Abstract/Résumé

SALLIE KELLER (University of Waterloo)

[Monday May 27 / lundi 27 mai, 8:30]

Celebrating the Challenges of Our Profession

Célébrons les défis de notre profession

Do we deserve an International Year of Statistics? Yes! It has been recognized by leaders in science, engineering, business, government, the social sciences and policy that the data complexity – the big data revolution - of today's global realities needs our leadership. Are we ready to step outside the comfort of our offices and engage? As a field that has spent more than a century unraveling the mystery of uncertainty and applying this to evidence based decisions, we need to be prepared to embrace this revolution and allow it to shape our future.

Méritons-nous une année mondiale de la statistique ? Oui ! Des dirigeants en science, génie, administration, gouvernement, sciences sociales et politique ont reconnu que la complexité des données - la révolution des grandes données - des réalités globales d'aujourd'hui a besoin de notre direction. Sommes-nous prêts à quitter le confort de nos bureaux et à nous engager ? En tant que domaine ayant passé plus d'un siècle à révéler le mystère de l'incertitude et à l'appliquer aux décisions basées sur l'évidence, nous devons nous préparer à adopter cette révolution et lui permettre de façonner notre avenir.

1B-V: Statistics and the Environment La statistique et l'environnement

Organizer and Chair / Responsable et présidente: Debbie J. Dupuis (HEC Montréal)

Room/Salle: MedSci 231

Abstract/Résumé

CHRISTOPHER WIKLE (University of Missouri)

[Monday May 27 / lundi 27 mai, 10:30]

Statistics and the Environment: Overview and Challenges

Statistique et environnement : vue d'ensemble et défis

The earth's environment is undergoing unprecedented change due to the influences of climate, land use, pressures from growing population, etc. Our understanding of the underlying processes, and their interactions, is somewhat limited. Yet, we have increasingly large amounts of environmental information from high-resolution deterministic model simulations, remotely sensed observations (e.g., satellites), and increasingly sophisticated (e.g., wireless) monitoring networks. Currently, environmental statistics is concerned with understanding variation and uncertainty in high dimensional, multivariate, nonlinear systems in the presence of such data and their uncertainties. This talk presents an overview of current and future challenges and model-based solutions in environmental statistics.

L'environnement de la terre subit un changement sans précédent à cause de l'influence du climat, de l'utilisation du terrain, des pressions d'une population grandissante, etc. Notre compréhension des processus sous-jacents et de leurs interactions est quelque peu limitée. Pourtant, nous avons une quantité de plus en plus grande de renseignements environnementaux des simulations de modèles déterministes à haute résolution, d'observations de télédétection (p. ex. satellites) et de réseaux de surveillance de plus en plus sophistiqués (p. ex. sans fil). Les statistiques environnementales s'intéressent actuellement à comprendre la variation et l'incertitude de systèmes en plusieurs dimensions, multivariés et non-linéaires en présence de telles données et de leur incertitude. Cet exposé présente une vue d'ensemble des défis actuels et futurs et des solutions basées sur un modèle en statistique environnementale.

1B-I1: Financial Mathematics Mathématiques financières

Organizer and Chair / Responsable et président: Rafal Kulik (University of Ottawa)

Session sponsored by the Probability Section
Séance parrainée par le Groupe de probabilité

Room/Salle: ECHA L1-430

Abstracts/Résumés

DON MCLEISH (University of Waterloo)

[Monday May 27 / lundi 27 mai, 10:30]

Martingales, Likelihood, and Monte Carlo Methods for Continuous Time Models in Finance

Martingales, vraisemblance et méthodes de Monte Carlo pour les modèles en temps continu en finance

We will discuss modelling continuous time processes using products and finite approximations to product integrals. These define a Radon-Nikodym derivative with respect to a base measure or process, chosen for its ease of simulation. Since Radon-Nikodym derivatives such as that of Girsanov are necessarily martingales, a related question is when such products form martingales. The goal is a general computational, likelihood-friendly framework for the construction and simulation of models for continuous-time processes, and parameter estimation. Examples of stochastic volatility models and jump diffusions are given. Parts of this talk are based on joint work with Carole Bernard and Zhenyu Cui.

Nous discuterons la modélisation des processus en temps continu en utilisant des produits et des approximations finies d'intégrales de produits. Ceux-ci définissent une dérivée de Radon-Nikodym par rapport à une mesure de base ou un processus, choisi pour l'aisance de simulations. Puisque les dérivées de Radon-Nikodym telles que celles de Girsanov sont nécessairement des martingales, une question apparentée se pose à savoir quand de tels produits forment des martingales. Le but est un cadre de calcul général permettant la vraisemblance pour construire et simuler des modèles pour des processus en temps continu et l'estimation des paramètres. Nous donnons des exemples de modèles de volatilité stochastique et de distribution des chocs. Des parties de cet exposé proviennent de travaux réalisés conjointement avec Carole Bernard et Zhenyu Cui.

CODY HYNDMAN & POLYNICE OYONO NGOU (Concordia University)

[Monday May 27 / lundi 27 mai, 11:00]

A Convolution Method for Numerical Solution of Backward Stochastic Differential Equations

Une méthode par convolution pour résoudre numériquement les équations différentielles stochastiques rétrogrades

We propose a new method for the numerical solution of backward stochastic differential equations (BSDEs) which finds its roots in Fourier analysis. The method consists of an Euler time discretization of the BSDE with certain conditional expectations expressed in terms of Fourier transforms and computed using the fast Fourier transform (FFT). The problem of error control is addressed, we consider the extension of the method to reflected BSDEs, and some numerical examples are consid-

Nous proposons une nouvelle méthode pour résoudre numériquement les équations différentielles stochastiques rétrogrades qui trouvent leurs racines dans l'analyse de Fourier. La méthode consiste en une discrétisation temporelle d'Euler des équations différentielles stochastiques rétrogrades sous certaines espérances conditionnelles exprimées en terme de transformées de Fourier et calculées en utilisant la transformée de Fourier rapide. Nous traitons le problème du contrôle de l'erreur, nous considérons le prolongement de la méthode aux équations différentielles sto-

1B-I1: Financial Mathematics Mathématiques financières

ered from finance demonstrating the performance of the method.

chastiques rétrogrades reflétées et quelques exemples numériques en finance pour démontrer la performance de la méthode.

TAHIR CHOULLI & JUNFENG MA (University of Alberta)

[Monday May 27 / lundi 27 mai, 11:30]

Stochastic Hellinger Distance: A Statistical Tool for a Financial Economics Problem

Distance stochastique de Hellinger: méthode statistique pour un problème en finance

Consider an agent possessing non-tradable asset and investing in stock market. Her goal is to find optimal portfolio and optimal time to liquidate all her assets. This problem can be solved if we understand how optimal portfolio depends on horizon. Our approach relies on extending Hellinger distance to positive martingales. This extension allows us identifying agents/investors whose optimal portfolio is horizon-unbiased. Our results contributed independently with other researchers to the birth of Forward Utilities. An advanced extension of Hellinger distance leads to Minimal Hellinger Deflator which is related to the Non-arbitrage theory.

Notre considérons un investisseur possédant un bien et investissant dans le marché financier. Son but est de déterminer le portefeuille optimal et le temps optimal pour liquider tous ses actifs (bien inclus). Nous procédons par étendre la distance de Hellinger, et nous aboutissons ainsi à la notion d'utilités forward que d'autres chercheurs ont réussi à définir indépendamment. Notre analyse décrit explicitement la classe d'utilités forward et leurs portefeuilles optimaux. Une extension approfondie de la distance de Hellinger nous permet de définir le déflateur minimal de Hellinger. Cette notion de déflateur est intimement liée à la théorie d'arbitrage.

1B-I2: New Developments in Statistical Neuroscience Nouveaux développements en statistique de la neuroscience

Organizer and Chair / Responsable et président: Shoja'eddin Chenouri (University of Waterloo)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: EdSouth 254

Abstracts/Résumés

SAM BEHSETA (California State University)

[Monday May 27 / lundi 27 mai, 10:30]

Bayesian Modeling of Multiple Spike Trains

Modélisation bayésienne de suites de potentiels d'action multiples

Neurophysiological studies often involve in-vivo examination of the spiking activity of neurons following various external stimuli. These investigations typically attempt to identify neurons that have condition-related differences among firing-rate functions; population-level differences across conditions; and neurons that exhibit interesting patterns of temporal difference across conditions. We discuss a series of nonparametric Bayesian modeling strategies to address the inferential problem of comparing multiple neuronal spiking patterns, including simultaneously recorded neurons. We showcase our proposed Bayesian paradigm in conjunction with data recorded from the primary motor cortex, prefrontal cortex and supplementary eye field of brain.

Les études neurologiques nécessitent souvent l'étude in-vivo de l'activité de pointe des neurones en réaction à différents stimuli externes. Ces études tentent habituellement d'identifier les neurones dont les différences entre les fonctions de taux d'allumage sont reliées à la condition; les différences entre les conditions au sein de la population et les neurones qui affichent des modèles intéressants de différence temporelle entre les conditions. Nous traitons une série de stratégies de modélisation bayésiennes non-paramétriques pour aborder le problème inférentiel de comparaison de plusieurs modèles de suites de potentiels d'action neurale, incluant les neurones enregistrés simultanément. Nous présentons le modèle bayésien que nous proposons en l'appliquant à des données du cortex moteur primaire, du cortex préfrontal et de parties supplémentaires du domaine oculaire du cerveau.

XI (ROSSI) LUO (Brown University), **STEVEN GEE & VIKAAS SOHAL** (University of California, San Francisco), **DYLAN SMALL** (University of Pennsylvania)

[Monday May 27 / lundi 27 mai, 11:00]

A Point-Process Response Model for Optogenetic Experiments on Neural Circuits

Un modèle de réponse d'un processus ponctuel dans les expériences optogénétiques sur les circuits neuronaux

Optogenetics is a new tool to stimulate genetically targeted neuronal circuits using light flashes. The data from such experiments present a statistical challenge of modeling a high frequency point process (neuronal spikes) while the input is another high frequency point process (light flashes). We will describe a new probabilistic approach to model the relationships between two point processes us-

L'optogénétique est un nouvel outil utilisé pour stimuler génétiquement des circuits neuronaux ciblés à l'aide de flashes lumineux. Les données de telles expériences présentent un défi statistique de modélisation d'un processus ponctuel à haute fréquence (pointes neuronales) alors que l'intrant est un autre processus ponctuel à haute fréquence (flashes lumineux). Nous décrivons une nouvelle approche probabiliste pour modéliser les relations entre les deux pro-

1B-I2: New Developments in Statistical Neuroscience Nouveaux développements en statistique de la neuroscience

ing additive point-process response functions. A real dataset and simulations demonstrate the advantages of our model. Comparing the model parameters enables neuroscientists to further study how neural circuits are altered under various disease conditions and/or experimental manipulations.

cessus ponctuels à l'aide de fonctions additives de réponse aux processus ponctuels. Un ensemble de données réelles et des simulations démontrent les avantages de notre modèle. La comparaison des paramètres du modèle permet aux neuroscientifiques d'étudier davantage la manière dont les circuits neuronaux sont modifiés sous plusieurs conditions de maladie et/ou des manipulations expérimentales.

REZA RAMEZAN, SHOJA'EDDIN CHENOURI & PAUL MARRIOTT (University of Waterloo)

[Monday May 27 / lundi 27 mai, 11:30]

The Skellam Process: A New Approach to Modelling Neural Spike Trains

Le processus de Skellam : une nouvelle approche pour la modélisation de suites de potentiels d'action neurale

A Spike train is a sequence of neural action potentials generated by a neuron. How the brain decodes the information content of neural spike trains is still an important open question. Using a Poisson process is a popular approach in modelling such data, however, its basic assumptions are usually violated due to biological phenomena such as bursting or refractory period. We have developed the Skellam Process with Resetting (SPR) as an alternative model, which is biologically justifiable, and has flexible multivariate generalizations for multiple-neuron modelling. An efficient estimation algorithm is developed for parameter estimation.

Une suite de potentiels d'action est une séquence de potentiel d'actions neurales généré par un neurone. La manière dont le cerveau décode le contenu de l'information de suites de potentiels d'action neurale reste toujours une question en suspens. L'utilisation d'un processus de Poisson se veut une approche populaire de modélisation de telles données. Cependant, ses hypothèses de base sont habituellement violées à cause des phénomènes biologiques tels qu'un éclatement ou une période réfractaire. Nous avons développé le processus de Skellam avec réinitialisation comme modèle alternatif qui est justifiable de manière biologique et qui permet des généralisations multivariées flexibles de la modélisation de neurones multiples. Un algorithme efficace d'estimation est développé pour l'estimation du paramètre.

1B-I3: Recent Development on Computer Experiments Développements récents en expériences informatiques

Organizer and Chair / Responsable et président: Devon C. Lin (Queen's University)

Room/Salle: ECHA 1-498

Abstracts/Résumés

DAVE WOODS (University of Southampton), **ANTONY OVERSTALL** (University of St Andrews)

[Monday May 27 / lundi 27 mai, 10:30]

Bayesian Inference for Computationally Expensive Models with Application to the Estimation of Stem Cell Properties
Inférence bayésienne pour modèles coûteux en ressources et applications pour l'estimation des propriétés des cellules souches

Bayesian inference for statistical models that depend on the evaluation of a computationally expensive simulator is often complicated by the limited number of likelihood evaluations that are feasible. We present a novel inference strategy which exploits a Gaussian process approximation of the simulator and allows computationally efficient MCMC inference. Our methodology is less reliant on the determination of tuning parameters than previous approaches, and allows model diagnostics without requiring additional simulator evaluations. We demonstrate the methods through application to a simulator that describes the properties of human embryonic stem cells using data from optical trapping experiments.

Dans les modèles statistiques qui dépendent de l'évaluation d'un simulateur coûteux en ressources, l'inférence bayésienne est souvent compliquée par le nombre limité d'évaluations de la vraisemblance qui sont réalisables. Nous présentons une stratégie d'inférence inédite qui exploite une approximation du simulateur par processus gaussien et permet une inférence MCMC efficace en matière de calcul informatique. Notre méthodologie dépend moins de la détermination des paramètres de réglage que les approches précédentes et permet une évaluation du modèle sans recourir à d'autres évaluations par le biais de simulateur. Nous démontrons la pertinence de ces méthodes par une application sur un simulateur qui décrit les propriétés de cellules souches embryonnaires humaines à l'aide de données tirées d'expériences de piège optique.

JASON LOEPPKY (University of British Columbia - Okanagan), **HAO (NELSON) CHEN & WILLIAM WELCH** (University of British Columbia)

[Monday May 27 / lundi 27 mai, 11:00]

Bayesian Prediction and Inference in Analysis of Computer Experiments
Prévision et inférence bayésiennes en analyse d'expériences sur ordinateur

Gaussian Processes (GP) are commonly used in the analysis of data from a computer experiment. Ideally, the analysis will provide accurate predictions with correct coverage probabilities. A Bayesian method can, in principle, capture all sources of uncertainty and hence give valid inference. Several implementations are available in the literature, differing in choice of priors, etc. We compare them in terms of accuracy and coverage probability for real computer codes and via simulation. Based on experiments varying the features seen

Les processus gaussiens (PG) sont couramment utilisés pour analyser les données provenant d'expériences sur ordinateur. Idéalement, ces analyses fournissent des prévisions précises et des probabilités de couverture correctes. Une méthode bayésienne peut, en principe, tenir compte de toutes les sources d'incertitude et donc donner une inférence valide. Il en existe plusieurs applications dans la littérature, avec divers a priori, etc. Nous en comparons la précision et la probabilité de couverture pour des codes informatiques réels et par le biais de simulations. Avec des expériences qui font varier les caractéristiques des applica-

1B-I3: Recent Development on Computer Experiments Développements récents en expériences informatiques

in the various implementations, we identify the important factors for a robust Bayesian method.

tions existantes, nous identifions les facteurs les plus importants d'une méthode bayésienne robuste.

PRITAM RANJAN (Acadia University), **CHUNFANG DEVON LIN** (Queen's University)

[Monday May 27 / lundi 27 mai, 11:30]

A New Multi-Stage Sequential Design Approach for Estimating Quantiles of Simulator Outputs

Approche conceptuelle séquentielle multi-étapes pour estimer les quantiles des résultats de simulateurs

Computer simulators are often used to model extremely expensive or unobservable complex physical systems. The estimation of the 100p-th percentile of the simulator responses is an important yet challenging problem. A naive approach is to use a sequential design scheme, where the follow-up points are chosen by optimizing the d-dimensional expected improvement (EI) criterion for contour estimation. We propose a new three-stage procedure for estimating a pre-specified percentile. The new approach will require a 1-dimensional EI-optimization instead of the d-dimensional EI-optimization in the naïve approach. Several test examples are used to compare the performance of the two approaches.

Les simulateurs informatiques sont couramment utilisés pour modéliser des systèmes physiques complexes très chers ou non observables. L'estimation du 100p-ème percentile des réponses du simulateur constitue un problème important et complexe. Une approche naïve serait d'employer un plan séquentiel, où les points de suivi sont choisis en optimisant le critère d'amélioration escomptée (AE) à d dimensions pour l'estimation du contour. Nous proposons une procédure en trois étapes permettant d'estimer un percentile prédéterminé. Cette nouvelle approche nécessite une optimisation AE à une dimension, plutôt qu'à d dimensions. Nous donnons plusieurs exemples de tests pour comparer la performance des deux approches.

1B-14: Survey Sampling and Applied Mathematics Échantillonnage et mathématiques appliquées

Chair/Président: J.N.K. Rao (Carleton University)

Organizers/Responsables: Susana Rubin Bleuer (Statistics Canada) and/et Ioana Schiopu
Kratina (University of Ottawa)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: ECHA L1-140

Abstracts/Résumés

SUSANA RUBIN BLEUER (Statistics Canada), **PASCAL RIVIÈRE** (Caisse Nationale d'Assurance Vieillesse)

[Monday May 27 / lundi 27 mai, 10:30]

Micro-strata
Micro-strates

We present a mathematical algorithm for the optimization of sample overlap. The topic is of great concern to statistical agencies. Indeed, negative sample coordination (minimization of sample overlap) for example, is the statistical solution to alleviate response burden. And response burden can lead to non-response, response error, etc. Controlling sample overlap is also fundamental for updating panels in surveys and reducing costs of data collection. We concentrate on the micro-strata methodology, which accounts for the real burden experienced by respondents. We describe the algorithm, show why permutations need to be performed within the micro-strata, and discuss its advantages and disadvantages.

Nous présentons un algorithme mathématique d'optimisation du chevauchement d'échantillons. Ce sujet révèle un grand intérêt pour les agences statistiques. En fait, la coordination négative d'échantillons (minimisation du chevauchement de l'échantillon), par exemple, est une solution statistique pour atténuer le fardeau de réponse. Et le fardeau de réponse peut entraîner la non-réponse, l'erreur de réponse, etc. Contrôler le chevauchement des échantillons est également essentiel pour mettre à jour les panels des enquêtes et pour réduire les coûts de collecte des données. Nous nous concentrons sur la méthodologie des micro-strates qui représente le vrai fardeau des répondants. Nous décrivons l'algorithme, montrons pourquoi les permutations doivent être réalisées dans les micro-strates et discutons ses avantages et inconvénients.

LENKA MACH & JEAN-MARC FILLION (Statistics Canada), **IOANA SCHIOPU-KRATINA** (University of Ottawa)

[Monday May 27 / lundi 27 mai, 11:00]

Sample Coordination Using Methods Based on Linear Programming

Coordination des échantillons à l'aide de méthodes basées sur la programmation linéaire

Sample coordination problems can be formulated mathematically as optimization problems and solved using linear programming (LP) techniques. We formulate an important sample coordination problem, solve it using LP methods as well as using classical statistical methods for minimizing variances, and obtain a solution that has optimal con-

Les problèmes de coordination des échantillons peuvent être formulés mathématiquement comme des problèmes d'optimisation et résolus à l'aide de techniques de programmation linéaire (PL). Nous formulons un problème de coordination d'échantillons, le résolvons à l'aide de méthodes PL et de méthodes statistiques classiques pour minimiser les variances et obtenons une solution qui possède des proprié-

1B-14: Survey Sampling and Applied Mathematics Échantillonnage et mathématiques appliquées

ditional properties. To support the mathematical methods, we construct a discrete probability space within which other sample coordination problems can be solved. We illustrate the method on examples and compare our solution with solutions obtained by non-LP methods.

tés conditionnelles optimales. Pour supporter les méthodes mathématiques, nous construisons un espace de probabilité discret à l'intérieur duquel d'autres problèmes de coordination des échantillons peuvent être résolus. Nous illustrons la méthode à l'aide d'exemples et comparons notre solution à celles obtenues avec des méthodes non-PL.

CHANGBAO WU (University of Waterloo)

[Monday May 27 / lundi 27 mai, 11:30]

Bayesian Pseudo-EL Approach Using Survey Data: Overview and Application to Regression Analysis

Approche bayésienne de pseudo-vraisemblance empirique à l'aide de données d'enquête: aperçu et application à l'analyse de régression

This talk provides an overview of Bayesian pseudo empirical likelihood method for complex surveys. We first review the two different approaches described in Rao and Wu (2010), one based on the profile empirical likelihood and the other treating the nonparametric probability masses as generic parameters. We then show that the latter approach can be used for analysis of survey data under much broader settings. We demonstrate that the proposed Bayesian method can be used for regression analysis with survey data involving complex sampling designs. This is joint work with Jon Rao of Carleton University.

Cet exposé présente un aperçu de la méthode bayésienne de pseudo-vraisemblance empirique pour les enquêtes complexes. Nous examinons d'abord les deux approches décrites par Rao et Wu (2010) : l'une est basée sur le profil de la vraisemblance empirique et l'autre traite les masses de probabilités non paramétriques comme des paramètres génériques. Nous montrons ensuite que cette deuxième approche peut être utilisée pour l'analyse de données d'enquête sous des conditions beaucoup plus larges. Nous démontrons que la méthode bayésienne proposée peut être utilisée pour l'analyse de régression de données d'enquête comportant des plans d'échantillonnage complexes. Il s'agit de travaux effectués conjointement avec Jon Rao de l'Université Carleton.

1B-C1: Bayesian Estimation and Inference

Estimation et inférence bayésiennes

Chair/Président: Paul Gustafson (University of British Columbia)

Room/Salle: EdSouth 228

Abstracts/Résumés

AVIK HALDER & GLEN TAKAHARA (Queen's University)

[Monday May 27 / lundi 27 mai, 10:30]

Example of Bayesian Semi-Parametric Inference for the Frailty Model Using Lévy Process Prior

Un exemple d'inférence semi-paramétrique bayésienne du modèle de fragilité au moyen du processus de Lévy comme loi a priori

Lee et al. (2003) first proposed the use of a Lévy process prior for Bayesian inference of the cumulative hazard function in a proportional hazard model and we extend this model by incorporating a Gamma frailty component. The Lévy process prior generalizes that used by Sinha (1993), who does consider a frailty model. We characterize the joint posterior and use blocked Gibbs sampling for sampling from the joint posterior. We compared a special case of our result with Aslanidou et al. (1998)'s frailty model analysis. Our plan is to adapt Al Labadi (2012)'s tools for our model and use it for data analysis.

Lee et coll. (2003) ont d'abord proposé d'utiliser le processus de Lévy comme loi a priori pour déterminer l'inférence bayésienne de la fonction de danger cumulative dans un modèle de danger proportionnel. Nous avons étendu ce modèle en incorporant une composante de fragilité gamma. La loi a priori de Lévy généralise la loi utilisée par Sinha (1993) qui considère le modèle de fragilité. Nous caractérisons la loi a posteriori conjointe et utilisons l'échantillonnage de Gibbs par blocs pour faire un échantillonnage à partir de cette loi. Nous comparons un cas particulier de nos résultats avec l'analyse du modèle de fragilité d'Aslanidou et coll. (1998). Notre plan consiste à adapter les outils d'Al Labadi (2012) à notre modèle et à l'utiliser pour analyser des données.

KHADER KHADRAOUI (Université Laval), **PIERRE RIBEREAU** (Université Claude Bernard)

[Monday May 27 / lundi 27 mai, 10:45]

Bayesian Estimation with M-splines of the Spectral Measure of an Extreme-Value Distribution

Estimation bayésienne avec M-splines de la mesure spectrale d'une distribution de valeurs extrêmes

We consider a Bayesian estimation with M-splines of the spectral measure of an extreme-value distribution. The estimation of the spectral measure is proposed thanks to a Bayesian nonparametric estimator that is guaranteed to satisfy the moment constraint. We give the asymptotic behavior of the posterior distribution by studying the rate of convergence of the posterior measure in the M-spline case. The problem of routine calculation of posterior distributions for both coefficients and knots of M-spline is addressed using the RJMCMC simulation technique. A simulation study shows that the Bayesian estimator resulted from the M-spline prior setting provides significant improvement over other procedures.

Nous considérons une estimation bayésienne avec des M-splines de la mesure spectrale d'une distribution de valeurs extrêmes. L'estimation de la mesure spectrale est proposée grâce à un estimateur non-paramétrique bayésien qui est garanti pour satisfaire la contrainte du moment. Nous attribuons le comportement asymptotique de la distribution postérieure en étudiant le taux de convergence de la mesure a posteriori dans le cas de la M-spline. Nous contournons le problème de calcul de routine de distributions a posteriori pour les coefficients et les nœuds de la M-spline à l'aide d'une technique de simulation basée sur la méthode Monte Carlo par chaîne de Markov à sauts réversible. Une étude de simulation montre que l'estimateur bayésien provenant de la M-spline a priori offre des améliorations importantes comparativement aux autres procédures.

1B-C1: Bayesian Estimation and Inference Estimation et inférence bayésiennes

LUAI AL LABADI (UAE University), **MAHMOUD ZAREPOUR** (University of Ottawa)

[Monday May 27 / lundi 27 mai, 11:00]

A Bayesian Nonparametric Goodness of Fit Test for Right Censored Data Based on Approximate Samples from the Beta-Stacy Process

Test d'adéquation non-paramétrique bayésien pour des données censurées à droite sur des échantillons approximatifs provenant du processus beta-Stacy

The beta-Stacy process, a generalization of the Dirichlet process, is a fundamental tool in studying Bayesian non-parametric statistics. In this paper, we derive a simple, yet efficient way to simulate the beta-Stacy process. We compare the efficiency of the new approximation to several other well-known approximations, and we demonstrate a significant improvement. Using the Kolmogorov distance and samples from the beta-Stacy process, a Bayesian nonparametric goodness of fit test is proposed. The proposed test is very general in the sense that it can be applied to censored and non-censored observations. A few illustrative examples have been included.

Le processus beta-Stacy, une généralisation du processus de Dirichlet, est un outil fondamental dans l'étude des statistiques non-paramétriques bayésiennes. Dans cet article, nous obtenons une façon simple et désormais efficace, de simuler le processus beta-Stacy. Nous comparons l'efficacité de la nouvelle approximation à plusieurs autres approximations bien connues, et nous démontrons une amélioration considérable. Nous proposons un test d'adéquation non-paramétrique bayésien au moyen de la distance de Kolmogorov et d'échantillons provenant du processus beta-Stacy. Ce test est très général dans le sens qu'il peut aussi être appliqué à des observations non censurées. Des exemples illustrent cette présentation.

SAMAN MUTHUKUMARANA (University of Manitoba)

[Monday May 27 / lundi 27 mai, 11:15]

Bayesian Analysis and Model Assessment of Binary Item Response Data

Analyse bayésienne et évaluation de modèle de données de réponse d'item binaires

In this talk, I will introduce three binary item response models used in the literature within the Bayesian framework. These models allow us to capture the heterogeneity of data in different capacities with different features. For this reason, model assessment methods need to be developed for better inferential purposes. I will introduce an assessment criteria based on prior predictive simulations and illustrate the approach using actual scores from an introduction to statistics class at the University of Manitoba.

Dans cet article, nous présentons trois modèles de réponse d'item binaires utilisés dans la littérature dans le cadre bayésien. Ces modèles permettent de capturer l'hétérogénéité des données de différentes capacités avec diverses caractéristiques. C'est pour cette raison que les méthodes d'évaluation de modèle nécessitent d'être développées pour améliorer l'inférence. Je vais présenter les critères d'évaluation basés sur des simulations prédictives a priori et illustrer mon approche avec des résultats actuels provenant d'une présentation d'un cours de statistique à l'Université du Manitoba.

MOHAMMED OBEIDAT & JUXIN LIU (University of Saskatchewan)

[Monday May 27 / lundi 27 mai, 11:30]

Bayesian Estimation of Time Series of Counts

Estimation bayésienne de séries chronologiques avec données de dénombrement

We apply a Bayesian approach to analyzing time series of count data. The time dependence is mod-

Nous appliquons une méthode bayésienne pour l'analyse de séries chronologiques avec données de dénombrement.

1B-C1: Bayesian Estimation and Inference Estimation et inférence bayésiennes

elled by assuming a latent auto-regressive AR(p) Gaussian process for the residuals in the Poisson regression model. Different from the regular Gibbs sampling idea (i.e. update one variable at a time), we propose to sample the latent process as a block by using Sequential Monte Carlo method. By simulation studies, we find that SMC outperforms the regular Gibbs sampler in terms of smaller mean squared errors for the autoregressive parameters; meanwhile the mean square errors for other parameters are very competitive.

La dépendance temporelle est modélisée sous l'hypothèse d'un processus gaussien autorégressif AR(p) latent pour les résidus dans le modèle de régression de Poisson. Contrairement à la méthode régulière d'échantillonnage de Gibbs (qui met à jour une variable à la fois), nous proposons d'échantillonner le processus latent dans son ensemble à l'aide de la méthode de Monte Carlo séquentielle. Nous constatons, par des études de simulations, que cette méthode surpasse l'échantillon régulier de Gibbs avec des erreurs quadratiques moyennes plus petites pour les paramètres autorégressifs et très compétitives pour les autres paramètres.

FARHAD SHOKOOHI & MAHMOUD TORABI (University of Manitoba)

[Monday May 27 / lundi 27 mai, 11:45]

Bayesian Small Area Estimation Using P-Spline and Time Series Models

Estimation bayésienne pour petits domaines au moyen de modèles de p-splines et de séries chronologiques

Bayesian methods have been extensively used in small area estimation (SAE). Nonparametric linear model incorporating random effects and sampling errors was previously proposed in SAE using cross-sectional data. There are, however, many situations that we have time-related counts or proportions in SAE; for example monthly data on the number of incidences in small areas. We incorporate time-series models to the cross-sectional data using spline models. Then, we unify nonparametric analysis of both discrete and continuous data. The performance of the proposed approach is evaluated through simulation studies and by an application. Keywords: Bayesian computation; Hierarchical model; P-spline; Random effects; Time-series.

Les méthodes bayésiennes ont été largement utilisées dans l'estimation pour petits domaines (EPD). Un modèle linéaire non paramétrique intégrant des effets aléatoires et les erreurs d'échantillonnage a déjà été proposé pour l'EPD utilisant des données transversales. Cependant, plusieurs situations présentent des proportions ou des décomptes liés au temps dans les EPD, comme les données mensuelles sur l'incidence dans les petits domaines. Nous intégrons des modèles de séries chronologiques aux données transversales à l'aide de modèles de splines. Nous unifions ensuite l'analyse non paramétrique des données discrètes et continues. La performance de la méthode proposée est évaluée à l'aide d'études de simulation et d'une application. Mots-clés : Calcul bayésien ; modèle hiérarchique ; p-splines ; effets aléatoires ; séries chronologiques.

1B-C2: Biostatistics - Theory and Methods I Biostatistique - théorie et méthodes I

Chair/Présidente: Jinko Graham (Simon Fraser University)

Room/Salle: EdSouth 206

Abstracts/Résumés

ABDELAZIZ ALSUBIE & JIANAN PENG (Acadia University)

[Monday May 27 / lundi 27 mai, 10:30]

Simultaneous Inference for Low Dose Risk Estimation with Quantal Data in Benchmark Analysis

L'inférence simultanée pour l'estimation du risque à faible dose avec des données quantiques pour l'analyse d'un banc d'essais

Risk assessment studies where 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. Al-Saidy et al. (Biometrics 59:1056-1062, 2003) developed the hyperbolic band for low dose inference with quantal data. However, the shape of the constant width band may be more desirable than that of the hyperbolic band in some applications. In this talk, we will show that a constant width band should do better than the hyperbolic band with quantal data.

Les études d'évaluation du risque dans lesquelles les données sont utilisées pour établir des niveaux sécuritaires à faible dose d'un agent toxique constituent un bon défi étant donné que l'information des études est limitée à des niveaux de doses élevées d'agent. Al-Saidy et coll. (Biometrics 59 :1056-1062, 2003) ont développé une bande hyperbolique pour l'inférence à faible dose avec des données quantiques. Cependant, une forme de bande à largeur constante semble plus appropriée que la bande hyperbolique dans certaines applications. Dans cet article, nous montrerons qu'une bande à largeur constante devrait être plus efficace qu'une bande hyperbolique avec des données quantiques.

CAITLIN DALY, XIAOFENG LIU, ANGELA RUTLEDGE, KRISTIN HAUFF, VIJAY GREY & JEMILA HAMID (McMaster University)

[Monday May 27 / lundi 27 mai, 10:45]

A Novel Statistical Method for Transference of Paediatric Reference Intervals Using Hospital-Based Data

Une nouvelle méthode statistique pour le transfert d'intervalles de référence pédiatriques au moyen de données hospitalières

Clinical interpretation of laboratory tests requires comparison of the test results with published reference intervals (RIs). Establishing or verifying paediatric RIs for use in laboratories is difficult due to the challenge of collecting healthy samples. We propose a novel statistical method that allows the use of hospital-based data to determine the transferability of established RIs. Traditional transference may be conducted after accounting for the "healthy" proportion of hospital data using the Hoffman approach or Gaussian mixture distributions. Extensive simulations across various sce-

L'interprétation d'essais clinique en laboratoire nécessite une comparaison des résultats d'essai avec les intervalles de référence publiés (RI). L'établissement ou la vérification d'intervalles de référence publiés pour les laboratoires est complexe en raison de la difficulté à collecter des échantillons sains. Nous proposons une nouvelle méthode statistique qui permet d'utiliser des données d'hôpitaux pour déterminer la transférabilité des intervalles de référence établis. Le transfert traditionnel peut être effectué après avoir pris en compte la proportion « saine » des données hospitalières avec l'approche de Hoffman ou des mélanges de distributions gaussiennes. Des simulations étendues à di-

1B-C2: Biostatistics - Theory and Methods I Biostatistique - théorie et méthodes I

narios illustrate comparable results between traditional transference using data from healthy populations and the proposed method with mixture distributions.

vers scénarios illustrent les résultats comparables entre le transfert traditionnel avec des populations saines de données et la méthode proposée, ainsi que des distributions de mélange.

DEPENG JIANG & ROBERT TATE (University of Manitoba)

[Monday May 27 / lundi 27 mai, 11:00]

Latent Class Joint Model for Longitudinal and Time-to-Event Data

Modèle conjoint de classes latentes pour des données longitudinales et de temps avant événement

This study proposes a general latent variable approach for jointly analyzing longitudinal and time-to-event data. First, unobserved heterogeneity among subjects in developmental trajectories (e.g., mental health) can be captured using multiple latent classes, where each class was allowed to have its own survival function. Second, the general latent variable modeling framework makes it possible to place the survival analysis in a larger conceptual model in order to study the relationship of survival to other developmental trajectories. Data from the Manitoba Follow-up Study concerns about the linkages of trajectories of mental health and survival functions serve as an illustration.

Cette étude propose une approche générale à variable latente pour l'analyse conjointe de données longitudinales et de temps avant événement. D'abord, l'hétérogénéité non observée parmi les sujets dans les trajectoires développementales (comme la santé mentale) peut être saisie par des classes latentes multiples, où chaque classe possède sa propre fonction de survie. Ensuite, le cadre général de la modélisation à variable latente permet de placer l'analyse de survie dans un modèle conceptuel plus large afin d'étudier la relation entre la survie et les autres trajectoires développementales. Une illustration est présentée à l'aide des données de la Manitoba Follow-up Study qui se penche sur les liens entre les trajectoires de santé mentale et les fonctions de survie.

KUN LIANG (University of Waterloo), **SUNDUZ KELES** (University of Wisconsin - Madison)

[Monday May 27 / lundi 27 mai, 11:15]

Normalization of ChIP-seq Data with Control

Normalisation des données ChIP-Seq avec contrôle

ChIP-seq has become an important tool for identifying genome-wide protein-DNA interactions, including transcription factor binding and histone modifications. Proper normalization between the ChIP and control samples is an essential aspect of the statistical analysis of ChIP-seq data. We propose a novel method for estimating the normalization factor between the ChIP and the control samples and evaluate its statistical properties against existing methods. Simulation results show that our method is the most powerful one among methods that control FDR at nominal levels.

Les données ChIP-seq sont devenues un outil important d'identification des interactions de la protéine/ADN dans le génome, y compris du facteur de transcription liant et des modifications de l'histone. Une normalisation adéquate entre les échantillons ChIP et les échantillons témoins est un aspect essentiel de l'analyse statistique des données ChIP-seq. Nous proposons une nouvelle méthode pour estimer le facteur de normalisation entre les échantillons ChIP et les échantillons témoins et évaluons ses caractéristiques statistiques par rapport aux méthodes existantes. Les résultats de simulation montrent que notre méthode est plus efficace que celles qui contrôlent le taux de fausses découvertes à des niveaux nominaux.

1B-C2: Biostatistics - Theory and Methods I Biostatistique - théorie et méthodes I

POONEH PORDELI, XUEWEN LU & MURRAY BURKE (University of Calgary), **PETER X.-K. SONG** (University of Michigan)

[Monday May 27 / lundi 27 mai, 11:30]

Polynomial Spline Estimation for Partially Linear Single-Index Transformation Models with Current Status Data
Estimation par splines polynomiales pour des modèles de transformation partiellement linéaires à indice simple avec données actualisées

This paper considers polynomial spline estimation for partially linear single-index transformation models with current status data. We use polynomial splines to estimate both the transformation function with monotonicity constraint; and the link function of the single-index component with no such constraint. We implement the proposed estimation through a backfitting algorithm on generalized linear models. We conduct simulation studies to examine the finite sample performance of the proposed estimation method, and present an analysis of a real data set for illustration.

Dans cet article, il est question d'estimation par splines polynomiales pour des modèles de transformation partiellement linéaires à indice simple avec des données actualisées. Nous utilisons des splines polynomiales pour estimer à la fois la fonction de transformation sous contrainte de monotonie et la fonction de lien de la composante à indice simple sans cette contrainte. Nous mettons en oeuvre l'estimation proposée par le biais d'un algorithme de backfitting sur des modèles linéaires généralisés. Nous effectuons des études de simulation pour examiner la performance de la méthode proposée sur des échantillons finis, et nous l'illustrons par l'analyse d'un jeu de données réelles.

YUJIE ZHONG & RICHARD COOK (University of Waterloo)

[Monday May 27 / lundi 27 mai, 11:45]

Analysis of Current Status Family Data Subject to Ascertainment Bias
Analyse de données familiales de statut présent sujettes à un biais de constatation

Families are often sampled for inclusion in a study because a particular individual, called a proband, is a member of a disease registry. Statistical analysis of the family data must address the so-called ascertainment bias arising from this selection criterion. We describe and critically evaluate several approaches which address this issue in different ways. We focus on the setting in which only current status data are available for family members and consider parametric inference regarding of the marginal distributions and the association structure. Empirical properties of estimators are compared through simulation and their relative efficiency is examined.

Les familles sont souvent échantillonnées pour être incluses dans une étude parce qu'un des membres, appelé proposant, est inscrit sur le registre d'une maladie. L'analyse statistique des données familiales doit tenir compte du biais de constatation lié à ce critère de sélection. Nous effectuons une description et une évaluation critique de plusieurs méthodes qui traitent ce problème de différentes façons. Nous nous concentrons sur le fait que seules des données familiales de statut présent sont disponibles pour les membres des familles et nous examinons l'inférence paramétrique pour les distributions marginales et la structure d'association. Nous comparons les propriétés empiriques des estimateurs par voie de simulation et examinons leur efficacité relative.

1B-C3: Clustering and Classification Partitionnement et classification

Chair/Président: Hugh Chipman (Acadia University)

Room/Salle: EdSouth 176

Abstracts/Résumés

RESHU AGARWAL & PRITAM RANJAN (Acadia University)

[Monday May 27 / lundi 27 mai, 10:30]

mBACT: A New Classifier for Satellite Images

Arbre de classification naïve bayésienne (mBACT) : un nouveau classificateur pour les images satellite

In this paper, we propose a new method for classifying satellite images. The proposed classifier is a generalization of the Bayesian Additive Classification Tree (BACT) method originally developed in credit risk modeling for binary classification only. The new classifier, referred to as mBACT (multi-class BACT), outperforms the state-of-the-art classifiers in remote sensing. We use LANDSAT 5 TM data with all spectral bands, excluding the thermal channel, covering Annapolis Valley, Nova Scotia, Canada for performance comparison of the proposed methodology with the competing popular approaches.

Dans cet article, nous proposons une nouvelle méthode de classification d'images satellite. Le classificateur proposé est une généralisation de la méthode de classification additive bayésienne initialement développée dans la modélisation des risques de crédit pour la classification binaire seulement. Ce nouveau classificateur, appelé arbre de classification naïve bayésienne (à plusieurs classes) a surclassé les classificateurs les plus modernes en télédétection. Nous utilisons des données de satellite LANDSAT 5 TM avec toutes les bandes spectrales, à l'exception d'un canal thermique, couvrant la vallée d'Annapolis en Nouvelle-Écosse (Canada) pour comparer l'efficacité de la méthodologie proposée aux approches populaires concurrentes.

KHATOON ALOBAIDAN & HUGH CHIPMAN (Acadia University)

[Monday May 27 / lundi 27 mai, 10:45]

A Measure of Compactness for Hierarchical Clustering

Une mesure de compacité pour le regroupement hiérarchique

In bottom-up hierarchical clustering, a useful concept is a "mutual cluster": a group of points closer to each other than to any other points outside the group. Previous work has defined a mutual cluster via its diameter (largest within-group distance) and distances to points outside the group. We generalize the definition, requiring that a multiple of the diameter of the mutual cluster be less than the distance to the nearest outside point. Modifications of the mutual cluster algorithm enable identification of mutual clusters, which can annotate dendrograms and assist in their interpretation. We use real datasets to illustrate the method.

Dans la classification hiérarchique ascendante, un concept utile est une « classification mutuelle » : un groupe de points plus près les uns des autres que tous les autres points en dehors du groupe. Les travaux précédents ont défini une classification mutuelle par son diamètre (la plus grande distance dans un groupe) et par les distances jusqu'aux points en dehors du groupe. Nous généralisons cette définition, ce qui nécessite qu'un multiple du diamètre de la classification mutuelle soit plus petit que la distance du point externe le plus près. Des modifications de l'algorithme de classification mutuelle permettent d'identifier des classifications mutuelles qui peuvent illustrer des dendrogrammes et aider à les interpréter. Des jeux de données réels illustrent cette méthode.

1B-C3: Clustering and Classification Partitionnement et classification

SANJEENA DANG & PAUL MCNICHOLAS (University of Guelph), **SALVATORE INGRASSIA & ANTONIO PUNZO** (University of Catania)

[Monday May 27 / lundi 27 mai, 11:00]

Cluster-Weighted Factor Analyzers

Analyseurs factoriels par pondération de grappes

Cluster-weighted modelling is a flexible statistical framework for modelling local relationships in heterogeneous populations on the basis of weighted combinations of local models. We extend cluster weighted models to include an underlying latent factor structure of the independent variable, resulting in a family of parsimonious cluster-weighted factor analyzers. This provides a model with the flexibility of clustering of high-dimensional data. An expectation-maximization algorithm and the Bayesian information criterion are used for parameter estimation and model selection, respectively. Our approach is illustrated on real and simulated data.

La modélisation par pondération de grappes est un cadre statistique flexible pour la modélisation de relations locales dans des populations hétérogènes sur la base de combinaisons de modèles locaux. Nous généralisons les modèles par pondération de grappes pour inclure une structure de facteurs latents sous-jacents de variables indépendantes créant une famille d'analyseurs de facteurs de pondération de grappes. Cela fournit un modèle avec la flexibilité de la pondération des données de haute dimension. Un algorithme d'espérance-maximisation et le critère d'information bayésien sont utilisés respectivement pour l'estimation des paramètres et la sélection de modèles. Nous illustrons notre approche par des données réelles et simulées.

THIERRY CHEKOUGO TEKOUANG (UT MD Anderson Cancer Center), **ALEJANDRO MURUA** (Université de Montréal)

[Monday May 27 / lundi 27 mai, 11:15]

Variable Selection With Overlapping Clustering

Sélection de variables avec groupements qui se chevauchent

In high dimensional data, the number of covariates is considerably larger than the sample size. To analyze these data, we propose a new method which searches simultaneously for the cluster structure of the observations and the informative variables. Our model for clustering and variable selection is inspired by the plaid model of Lazzeroni and Owen (2002) for bi-clustering. The estimation is done via the Monte Carlo EM algorithm. Applications to real data and comparisons with other methods show the advantages of our method in terms of variable selection and clustering.

Pour les données de plusieurs dimensions, le nombre de covariables est beaucoup plus grand que la taille de l'échantillon. Pour analyser ces données, nous proposons une nouvelle méthode qui cherche simultanément la structure de la grappe des observations et des variables instructives. Notre modèle de regroupement et de sélection de variables est inspiré par le modèle de plaid de biclustering de Lazzeroni et Owen (2002). L'estimation est effectuée avec l'algorithme EM Monte Carlo. Les applications à des données réelles et des comparaisons avec les autres méthodes montrent les avantages de notre méthode en termes de sélection de variables et de regroupement.

CRISTINA TORTORA & PAUL MCNICHOLAS (University of Guelph), **FRANCESCO PALUMBO** (Università di Napoli Federico II)

[Monday May 27 / lundi 27 mai, 11:30]

Probabilistic and Distance Based Factor Clustering: A Review

Étude de regroupement de facteurs basé sur la distance et probabiliste : une revue

1B-C3: Clustering and Classification Partitionnement et classification

Non-hierarchical cluster analysis aims at identifying the optimal k group partition in a multivariate dataset. Distance-based and model-based methods offer good performance according to the underlying data structure. However, their performance can deteriorate as the number of variables increases. Factor clustering refers to methods that integrate the variable extraction and clustering into one iterative algorithm, which optimizes a unique criterion to find the groups in a reduced subspace. This talk presents a generalized formulation of some methods that belong to distance-based and model-based families; it compares their overall performance using simulated and real data.

Une analyse de groupement non-hiérarchique vise à identifier la partition des groupes k optimaux dans des jeux de données multivariées. Les méthodes basées sur une distance et sur un modèle sont efficaces en fonction de la structure des données. Cependant, leur efficacité peut se détériorer lorsque le nombre de variables augmente. Le groupement de facteurs renvoie aux méthodes qui intègrent l'extraction et le groupement de variables en un algorithme itératif qui optimise un critère unique pour trouver les groupes dans un sous-espace réduit. Cet article présente une formulation de certaines méthodes qui appartiennent à des familles basées sur une distance et un modèle. Il compare l'efficacité générale de ces méthodes avec des données simulées et réelles.

IRENE VRBIK & PAUL MCNICHOLAS (University of Guelph)

[Monday May 27 / lundi 27 mai, 11:45]

Fractionally-Supervised Classification
Classification partiellement dirigée

The aim of classification is to discover inherent groups within the data for which a subset of observations have known group labels. We propose a fractionally-supervised approach wherein the influences from the unlabelled observations are allowed to range from 0 to 1, inclusively. Our adaptive classification approach uses a weighted likelihood to combine the information from the unlabelled and labeled populations at varying degree so that unlabelled data can provide full influence parameter estimation (semi-supervised classification), no influence on parameter estimation (discriminant analysis), or somewhere in between.

La classification vise à découvrir les groupes inhérents dans des données pour lesquelles un sous-ensemble d'observations a des étiquettes de groupe connues. Nous proposons une approche partiellement dirigée dans laquelle l'influence des observations sans étiquette est permise entre 0 et 1 inclusivement. Pour notre approche de classification adaptative, nous utilisons une vraisemblance pondérée qui permet de combiner l'information provenant des populations avec et sans étiquettes à différents degrés. L'objectif est que les données sans étiquette puissent fournir une estimation de paramètres d'influence (classification semi-dirigée), aucune influence sur l'estimation des paramètres (analyse discriminante), ou quelque chose entre ces deux estimations.

1C-I1: Dependence Models with Insurance Applications Modèles de dépendance avec applications en assurance

Organizer and Chair / Responsable et président: Gordon Willmot (University of Waterloo)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe d'actuariat

Room/Salle: ECHA L1-430

Abstracts/Résumés

JUN CAI (University of Waterloo)

[Monday May 27 / lundi 27 mai, 13:30]

Dependence Notions Through Arrangement Increasing Joint Survival Functions with Applications to Optimal Allocation Problems

Notions de dépendance par le biais de fonctions de survie conjointes et applications aux problèmes d'imputation optimale

Dependence plays an important role in optimal allocation problems for insurance and finance. In this talk, we propose new dependence notions to model dependent risks. These notions, defined through arrangement increasing joint survival functions, include many common dependence structures. We develop properties of these dependence structures and explore their applications to optimal allocation problems in insurance and finance. This talk is based on joint works with Wei Wei.

La dépendance joue un rôle important dans les problèmes d'imputation optimale en assurance et finance. Nous proposons de nouvelles notions de dépendance pour modéliser les risques dépendants. Ces notions, définies par des fonctions de survie, incluent de nombreuses structures de dépendance communes. Nous développons les propriétés de ces structures de dépendance et explorons leurs applications aux problèmes d'imputation optimale en assurance et finance. Cette présentation est fondée sur des travaux en collaboration avec Wei Wei.

CHRISTIAN GENEST, HUIJUN CHEN & DAVID A. STEPHENS (McGill University)

[Monday May 27 / lundi 27 mai, 14:00]

Bayesian Hierarchical Modelling of Automobile Insurance Claim Reserves

Modélisation bayésienne hiérarchique des provisions pour sinistres en assurance automobile

Insurance companies typically base their reserve allocation on qualitative exposure estimates provided by adjusters. A statistical solution to this problem will be proposed and illustrated with data from a Canadian insurer. A multinomial logit model is first used to predict the frequency of different claim types. The claim type severity is then modelled using predictors. The dependence between claim types is accounted for by a copula. MCMC sampling is used to estimate the parameters. The model is updated through the claim life cycle. The properties of the predictive loss distribution are investigated through simulations.

Les assureurs s'appuient généralement sur l'avis qualitatif d'experts en sinistres pour déterminer leurs provisions. Une solution statistique à ce problème sera proposée et illustrée avec des données d'un assureur canadien. Un modèle logit multinomial permet d'abord de prédire la fréquence de divers types de réclamations. Leur gravité est ensuite modélisée à l'aide de prédicteurs. Une copule rend compte de la dépendance entre les types de réclamations. Les paramètres sont estimés au moyen d'un algorithme MCMC. Le modèle est actualisé au fil du traitement des réclamations. Les propriétés de la loi prédictive des pertes sont étudiées par simulation.

1C-I1: Dependence Models with Insurance Applications Modèles de dépendance avec applications en assurance

ÉTIENNE MARCEAU & HÉLÈNE COSSETTE (Université Laval)

[Monday May 27 / lundi 27 mai, 14:30]

Recent Developments on Aggregation of Dependent Risks

Développements récents dans l'agrégation des risques dépendants

In the context of actuarial science and quantitative risk management, we investigate risk aggregation problems for a portfolio of dependent risks whose bivariate distribution is defined with a copula and with continuous marginals. For specific copulas and for specific marginals, we obtain explicit expressions for the cumulative distribution function of the sum of two dependent risks. For other copulas and marginals, we propose numerical methods to compute the cumulative distribution function of the sum. The findings are illustrated with numerical examples.

Dans le contexte de la science actuarielle et de la gestion quantitative des risques, nous étudions les problèmes d'agrégation pour un portefeuille de risques dépendants dont la distribution bivariée est définie avec une copule et des marginales continues. Pour des copules spécifiques et pour des marginales spécifiques, on obtient des expressions explicites pour la fonction de répartition de la somme de deux risques dépendants. Pour les autres copules et d'autres marginales, nous proposons des méthodes numériques pour calculer la fonction de répartition de cette somme. Les résultats sont illustrés par des exemples numériques.

1C-I2: Fitting Complex Models Using Automatic Differentiation Model Builder (ADMB) Ajustement de modèles complexes en utilisant ADMB

Chair/Présidente: Joanna Mills Flemming (Dalhousie University)

Organizer/Responsable: Noel Cadigan (Memorial University)

Room/Salle: EdSouth 206

Abstracts/Résumés

MOLLIE BROOKS, THOMAS BOHRMANN, FORREST STEVENS, ANDREW HEIN, MARY CHRISTMAN & CRAIG OSENBERG (University of Florida), **BENJAMIN BOLKER** (McMaster University)

[Monday May 27 / lundi 27 mai, 13:30]

Predicting Temporally Autocorrelated Pond Depths with Irregular Time Series Data

Prédire une profondeur de bassin avec autocorrélation temporelle à l'aide de données chronologiques irrégulières

Long-term environmental monitoring projects can produce irregular time series data due to changes in sampling regime or priorities. Another complication is that predictors may be monitored on different time scales than responses. Both of these problems typically require processing data before standard time series models are fit. We present these problems in an irregular twenty-year data set of ephemeral pond depths and associated covariates. We present an AR(1) model fit using MLE in ADMB, and tailored to predict daily pond-depths.

Les projets à long terme de suivis environnementaux peuvent produire des données chronologiques irrégulières lorsque les régimes ou priorités d'échantillonnage évoluent. De plus, les indicateurs sont parfois suivis selon d'autres échelles de temps que les réponses. En raison de ces deux problèmes, il faut généralement traiter les données avant de pouvoir ajuster des modèles de séries chronologiques standards. Nous explorons ces problèmes sur un ensemble de données représentant des profondeurs de bassin et des covariables associées, collectées de façon irrégulière sur une période de vingt ans. Nous présentons un ajustement de modèle AR(1) par estimation du maximum de vraisemblance (EMV) dans ADMB, qui est adapté pour prédire les profondeurs de bassin quotidiennes.

NOEL CADIGAN (Memorial University)

[Monday May 27 / lundi 27 mai, 14:00]

Improved Estimation and Forecasts of Fish Stock Maturities Using Generalized Linear Mixed Effects Models

Modèles linéaires généralisés à effets mixtes : améliorer l'estimation et les prévisions de la maturité d'un stock de poissons

Long time-series of Binomial-type data for age-specific numbers examined and proportion mature have been developed for some fish stocks. This information is required to estimate spawning stock biomass (SSB) which is an indicator of the stocks productivity potential. A common model used to estimate the proportions is Binomial logistic regression which is often applied separately for each cohort in a time-series. However, there can be large retrospective variation with this approach. I show how to use generalized linear mixed models with

De longues séries chronologiques de données binomiales (nombres de poissons examinés par âge, proportion de poissons adultes) ont été créées pour certains stocks de poissons. Ces informations sont nécessaires pour estimer la biomasse du stock reproducteur, un bon indicateur du potentiel de productivité du stock. Pour estimer cette proportion, on applique souvent la régression logistique binomiale à chaque cohorte d'une série chronologique. Toutefois, cette approche conduit souvent à une grande variation rétrospective. Je montre comment utiliser des modèles linéaires généralisés à effets mixtes avec des effets aléatoires autocor-

1C-I2: Fitting Complex Models Using Automatic Differentiation Model Builder (ADMB) Ajustement de modèles complexes en utilisant ADMB

autocorrelated random effects and marginal maximum likelihood via ADMB to improve estimates and forecasts.

rélés et le maximum de vraisemblance marginale via ADMB pour améliorer les estimations et les prévisions.

ANDERS NIELSEN (Danish Technical University)

[Monday May 27 / lundi 27 mai, 14:30]

Introduction to AD Model Builder and State–Space Fish Stock Assessment Model

Introduction à AD Model Builder et modèles d'évaluation d'espaces d'états de stocks de poissons

AD Model Builder is a flexible and efficient tool for developing and optimizing non-linear statistical models. AD Model Builder consists of a library for Automatic Differentiation (AD), a powerful optimizer, helpful container classes, tools for MCMC and profile likelihood computations, and finally Laplace approximation and importance sampling for dealing with random effects models. Single-species and age-structured fish stock assessments still remains the main tool for managing fish stocks. A simple state-space assessment model is presented as alternative to deterministic procedures and the full parametric statistical catch at age models.

AD Model Builder est un outil souple et efficace pour développer et optimiser des modèles statistiques non linéaires. AD Model Builder consiste en une bibliothèque de différentiation automatique (AD), un optimiseur puissant, des classes de conteneurs utiles, des outils pour les calculs MCMC et de vraisemblance de profil, ainsi que l'approximation de Laplace et l'échantillonnage par importance pour traiter les modèles à effets aléatoires. Les évaluations structurées selon l'espèce et selon l'âge sont le principal outil de gestion des stocks de poissons. Nous présentons un modèle d'évaluation d'espaces d'états simple comme alternative aux procédures déterministes et aux modèles statistiques paramétriques de prise selon l'âge.

1C-I3: High-Dimensional Data Analysis for Brain Imaging

Analyse de données de grande dimension pour l'imagerie encéphalique

Organizer and Chair / Responsable et président: Farouk Nathoo (University of Victoria)

Room/Salle: EdSouth 176

Abstracts/Résumés

MIRZA FAISAL BEG (Simon Fraser University)

[Monday May 27 / lundi 27 mai, 13:30]

Computational Morphometric Biomarkers for Discrimination of Dementias

Biomarqueurs morphométriques informatiques pour la discrimination des démences

The computerized identification of neuroanatomical changes from MR images that occur due to disease and distinguishing those from those that occur due to normal variation in the population is a difficult problem that is the goal of Computational Anatomy (CA). One of the key challenges is that making detailed measurements describing the brain structure leads invariably to very high dimensional data. I will discuss few tools from the Computational Anatomy toolbox and show a few results of applying these tools towards the problem of discrimination between structural changes in the brain in dementias such as Alzheimer's and Frontotemporal dementias.

Identifier les modifications neuroanatomiques dues à la maladie sur les images par RM et distinguer entre ces modifications et la variation normale de la population, voilà un problème difficile que l'« anatomie computationnelle » se propose de traiter. L'un des principaux défis réside dans les données de très grande dimension produites dans le cadre de mesures de la structure du cerveau. Je présente quelques outils offerts par l'« anatomie computationnelle » et je montre comment ces outils s'appliquent au problème de distinguer entre les modifications structurelles du cerveau dans le cas de démences comme la maladie d'Alzheimer et la démence fronto-temporale.

TIMOTHY D. JOHNSON (University of Michigan), **THOMAS E. NICHOLS** & **TIAN GE** (University of Warwick)

[Monday May 27 / lundi 27 mai, 14:00]

A GPU Implementation of a Spatial GLMM: Assessing Spatially Varying Coefficients of Multiple Sclerosis Lesions

Implémentation GPU d'un MLGM spatial : évaluation de coefficients à variation spatiale de lésions de sclérose en plaque

In this talk I present a parallel programming approach for the estimation of spatially varying coefficients in a spatial GLMM analyzing Multiple Sclerosis lesion images. The model is a spatial probit model of binary image data with subject specific covariates. The spatial coefficients for these covariates are spatially dependent and are a priori modeled using a multivariate conditional autoregressive model. The large size of these images and coefficient maps, along with spatial dependence between voxels, makes this problem extremely computationally intense. Code parallelization and GPU implementation results in an executable that is 50 times faster than a serial implementation.

Je présente une approche par programmation parallèle de l'estimation de coefficients à variation spatiale dans un MLGM spatial pour analyser des images de lésions de sclérose en plaque. Le modèle est un modèle probit spatial de données d'images binaires, avec des covariables spécifiques au sujet. Les coefficients spatiaux de ces covariables sont dépendants de l'espace et modélisés a priori à l'aide d'un modèle autorégressif conditionnel multivarié. La grande taille de ces images et de ces cartes de coefficients, ainsi que la dépendance spatiale entre voxels, exige une intensité de calcul particulièrement forte. Grâce à la parallélisation des codes et l'implémentation GPU, l'exécution est 50 fois plus rapide qu'une implémentation en série.

1C-I3: High-Dimensional Data Analysis for Brain Imaging Analyse de données de grande dimension pour l'imagerie encéphalique

MARY LESPERANCE & FAROUK NATHOO (University of Victoria), **ANDREW LAWSON** (Medical University of South Carolina), **CHARMAINE DEAN** (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 14:30]

Comparing Variational Bayes with Markov Chain Monte Carlo for Bayesian Computation in Neuroimaging

Comparaison du calcul bayésien variationnel et Monte Carlo par chaîne de Markov pour calcul bayésien en neuroimagerie

We consider Bayesian computation within the context of brain imaging studies, focussing on the neuroelectromagnetic inverse problem in electroencephalography. The relationship between observed scalp-level data measured from electrodes and unobserved neural activity within the brain is modelled by an underdetermined dynamic linear model. We discuss variational approximations for fitting hierarchical models where parameters represent the unknown neural sources of interest. While variational methods have been widely adopted in neuroimaging, they have received little attention in the statistical literature where Markov chain Monte Carlo (MCMC) is often used. We compare the approximations to MCMC using both synthetic and real electroencephalography data.

Nous étudions le calcul bayésien dans le contexte des études d'imagerie cérébrale, en mettant l'accent sur le problème inverse neuroélectromagnétique en électroencéphalographie. Nous modélisons la relation entre les données mesurées par électrode au niveau du scalp et l'activité neuronale non observée dans le cerveau par le biais d'un modèle linéaire dynamique sous-déterminé. Nous discutons d'approximations variationnelles pour l'ajustement de modèles hiérarchiques où les paramètres représentent des sources d'intérêt neuronales inconnues. Les méthodes variationnelles sont aujourd'hui largement adoptées en neuroimagerie, mais elles ne sont que peu étudiées dans la littérature, contrairement à la méthode de Monte Carlo par chaîne de Markov (MCMC). Nous comparons ces approximations de type MCMC à l'aide de données EEG à la fois synthétiques et réelles.

1C-14: Identifiability: Pushing Data to the Limits
Identifiabilité : les données poussées à leurs limites

Organizer and Chair / Responsable et président: Luke Bornn (Harvard University)

Room/Salle: ECHA 1-498

Abstracts/Résumés

PAUL GUSTAFSON (University of British Columbia)

[Monday May 27 / lundi 27 mai, 13:30]

Bayesian Inference in Partially Identified Models: Exploring the Limits of Limited Data

Inférence bayésienne dans les modèles partiellement identifiés : aux limites des données limitées

Roughly put, partial identification is the grey zone between learning everything as the data accumulate, and learning nothing. In this talk, I review how partial identification can arise in practical problems, and how the performance of Bayesian inference can be characterized in such settings. I also discuss challenges arising as models become more complex, in terms of both characterizing partial identification and computing posterior quantities.

On entend par identification partielle cette zone grise entre apprendre de tout à rien lorsque les données s'accumulent. Dans ma présentation, j'explore les problèmes pratiques dans lesquels peut survenir un cas d'identification partielle et comment caractériser le rendement de l'inférence bayésienne dans de telles situations. Je discute aussi des défis qui se présentent, au fur et à mesure que les modèles deviennent plus complexes, pour caractériser l'identification partielle et calculer les quantités postérieures.

WES JOHNSON (UC Irvine), **ADAM BRANSCUM** (Oregon State University), **TIM HANSON** (University of South Carolina)

[Monday May 27 / lundi 27 mai, 14:00]

Bayesian Semiparametric (Identifiable) Regression Modeling for ROC Analysis, Absent a Gold Standard

Modélisation de la régression semiparamétrique bayésienne (identifiable) pour l'analyse ROC en l'absence d'une norme de référence

A semiparametric regression model is developed to evaluate a continuous biomarker, absent perfect diagnosis. To compensate we incorporate covariate information by modeling the probability of disease as a function of covariates; and we model biomarker outcomes to depend on covariates. Researchers can thus quantify the impact of covariates on the accuracy of a test; for instance, it may be easier to diagnose disease in older individuals. Flexible error distributions are used. The resulting regression model is identifiable under mild conditions. We obtain inferences about covariate-specific test accuracy and illustrate the model using data on an age-adjusted biomarker for lung cancer.

Nous développons un modèle de régression semiparamétrique pour évaluer un biomarqueur continu en l'absence d'un diagnostic parfait. Pour compenser, nous incorporons des informations sur les covariables en modélisant la probabilité de la maladie comme fonction des covariables; puis nous modélisons les résultats des biomarqueurs pour dépendre des covariables. Les chercheurs peuvent ainsi quantifier l'impact des covariables sur la précision d'un test; par exemple, il peut être plus facile de diagnostiquer la maladie chez les individus plus âgés. Nous utilisons des distributions d'erreurs flexibles. Ce modèle de régression est identifiable sous des hypothèses faibles. Nous obtenons des inférences concernant la précision du test spécifique à la covariable et illustrons le modèle avec des données sur un biomarqueur du cancer du poumon ajusté selon l'âge.

1C-14: Identifiability: Pushing Data to the Limits Identifiabilité : les données poussées à leurs limites

STEPHEN WALKER (University of Kent)

[Monday May 27 / lundi 27 mai, 14:30]

Identifiability Issues with Mixtures of Normal Distributions and the Mixture of Dirichlet Process Model

Problèmes d'identification avec les mélanges de distributions normales et le modèle de mélange de processus de Dirichlet

The mixture of Dirichlet process model, commonly used in Bayesian nonparametric modeling, is known to suffer from identifiability issues regarding the parameters, i.e. the weights and locations. This talk will describe an alternative procedure for updating the parameters given the data, based on sound Bayesian principles, which attempts to provide some interpretation for the parameters.

Le modèle de mélange de processus de Dirichlet, communément utilisé en modélisation non paramétrique bayésienne, souffre de problèmes d'identification concernant les paramètres, à savoir les poids et les positions. Cette présentation décrira une procédure alternative pour la mise à jour des paramètres sur la base des données, fondée sur de bons principes bayésiens, qui vise à offrir une interprétation des paramètres.

1C-15: Moving Up in Academic Ranks: Challenges and Strategies for Statisticians Involved in Interdisciplinary Research
Gravir les échelons universitaires – défis et stratégies pour les statisticiens impliqués dans la recherche interdisciplinaire

Organizer and Chair / Responsable et présidente: Wendy Lou (University of Toronto)

Room/Salle: EdSouth 254

Description

Statistical researchers in interdisciplinary environments face special challenges at performance evaluation, promotion, and tenure application. The extra effort needed to highlight creative/original statistical contributions to interdisciplinary research can be discouraging. Other issues include: (i) Reduced productivity while training oneself/students in new disciplines; (ii) Lack of appreciation for non-statistical journals or non-traditional research output; (iii) Conflicting expectations under joint appointments; (iv) Higher service demands including grant reviews and consulting. The panel will discuss strategies for showcasing successfully one's interdisciplinary research contributions. The panelists will share their perspectives, including several administrative viewpoints, gathered from various research institutions in Canada and the US.

Les statisticiens oeuvrant dans des milieux interdisciplinaires font face à des défis particuliers à l'égard de l'évaluation du rendement, des promotions et de la demande de permanence. L'effort supplémentaire nécessaire pour mettre en valeur les contributions statistiques créatives et inédites à la recherche interdisciplinaire peut être décourageant. D'autres questions importantes comprennent (i) une productivité réduite découlant de sa propre formation ou de celle d'étudiants dans de nouvelles disciplines, (ii) le manque de reconnaissance des journaux non statistiques ou des résultats de recherche non traditionnels, (iii) des attentes inconciliables dans le cadre de nominations conjointes et (iv) des demandes plus nombreuses du Service, y compris de la consultation et des évaluations de demandes de subventions. Le panel discutera de stratégies pour réussir à mettre en valeur ses contributions en recherche interdisciplinaire. Les panélistes exposeront leur vision, y compris plusieurs points de vue administratifs recueillis de divers établissements de recherche au Canada et aux États-Unis.

1C-I6: Survival Analysis Methods and Applications Méthodes et applications en analyse de survie

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

Room/Salle: ECHA L1-140

Abstracts/Résumés

RICHARD COOK (University of Waterloo)

[Monday May 27 / lundi 27 mai, 13:30]

Assessment of Treatment Effects Following Intermediate Events in Cancer Trials

Évaluation des effets du traitement après événements intermédiaires dans les essais cliniques sur le cancer

Cancer trials are routinely designed to assess the effect of treatment on progression-free survival times, but real interest lies in identifying treatments that prolong survival. This talk will present challenges arising in the study of treatment effects on survival following progression as well as overall survival, in trials designed based on a progression-free survival endpoint. Issues of causal inference will be considered including methods based on propensity score analysis and marginal structural models. These approaches will be reviewed and application to a cancer trial will be used for illustration.

Les essais cliniques sur le cancer sont généralement conçus pour évaluer l'effet d'un traitement sur les temps de survie sans progression, mais il est surtout intéressant d'identifier les traitements qui prolongent la survie. Nous explorons les défis soulevés par l'étude des effets du traitement sur la survie après progression et la survie globale, dans les essais conçus sur la base d'un paramètre de survie sans progression. Nous étudions les problèmes d'inférence causale, les méthodes fondées sur l'analyse des coefficients de propension et les modèles structurels marginaux. Nous passons en revue ces approches et les illustrons par une application à un essai clinique sur le cancer.

YI LI (University of Michigan), **PEIRONG XU & LIXING ZHU** (Hong Kong Baptist University)

[Monday May 27 / lundi 27 mai, 14:00]

Ultrahigh Dimensional Time Course Feature Selection

Sélection de caractéristiques de période de dimension très élevée

Statistical challenges arise from modern biomedical studies that produce time course genomic data with ultrahigh dimensions. We propose a novel GEE-based screening procedure that only pertains to the specifications of the first two marginal moments and a working correlation structure. The newly proposed procedure merely involves making a single evaluation of GEE function and thus effectively reduces the dimensionality of covariates. The new method is robust with respect to the misspecification of correlation structure and enjoys theoretical readiness, which is further verified via intensive Monte Carlo simulations.

Les études biomédicales modernes qui produisent sur une période des données génomiques de dimension très élevée présentent plusieurs défis statistiques. Nous proposons une nouvelle procédure de dépistage fondée sur les équations d'estimation généralisées (EEG) qui ne porte que sur les spécifications des deux premiers moments marginaux et une structure de corrélation provisoire. Cette procédure ne requiert qu'une seule évaluation de la fonction EEG, réduisant ainsi la dimensionnalité des covariables. La nouvelle méthode est robuste aux erreurs de spécification dans la structure de corrélation et jouit d'une bonne capacité théorique, ce que nous vérifions par le biais de multiples simulations Monte Carlo.

1C-I6: Survival Analysis Methods and Applications Méthodes et applications en analyse de survie

RONGHUI (LILY) XU & GORDON HONERKAMP-SMITH (University of California, San Diego)

[Monday May 27 / lundi 27 mai, 14:30]

Explained Variation in Correlated Survival Data

Variation expliquée dans les données de survie corrélées

We consider explained variation in correlated survival data, specifically under the proportional hazards mixed-effects modeling (PHMM) of such data. The concept of explained variation can be formulated in different ways, and we discuss the formulations that are commonly encountered in the literature. Explained variation is particularly relevant under the mixed-effects modeling, as variance components are inherently defined, and their contributions to the total variation can be easily understood. We study the proposed measures both in theory and through simulation, and application examples will be provided.

Nous étudions la variation expliquée dans les données de survie corrélées, notamment dans les modèles de risques proportionnels à effets mixtes de telles données. Le concept de la variation expliquée peut se formuler de diverses manières ; nous discutons des formulations courantes dans la littérature. La variation expliquée est particulièrement pertinente dans les modèles à effets mixtes, car les composantes de la variance sont définies de manière intrinsèque et leur contribution à la variation totale est facile à comprendre. Nous étudions les mesures proposées en théorie et par le biais de simulation et en donnons des exemples d'applications.

1C-C1: Biostatistics - Theory and Methods II Biostatistique - théorie et méthodes II

Chair/Présidente: Mireille Schnitzer (Harvard School of Public Health)

Room/Salle: EdSouth 265

Abstracts/Résumés

SAYANTEE JANA & NARAYANASWAMY BALAKRISHNAN (McMaster University), **DIETRICH VON ROSEN** (Swedish University of Agricultural Sciences), **JEMILA HAMID** (McMaster University)

[Monday May 27 / lundi 27 mai, 13:30]

A Moderated Trace Test for the Growth Curve Model and Its Application for Gene-Filtering in Time-Course Microarray Data

Test de trace modéré pour le modèle de courbe de croissance et application au filtrage de gènes dans les données de micropuces de période

Traditional tests for analyzing longitudinal data are not applicable in high-dimensional setup (when $n < p$). We propose a moderated test using the Growth Curve Model, which is useful for analyzing high-dimensional longitudinal data. Extensive simulations demonstrated that our moderated test is unbiased, and is monotone with respect to sample size and parameter value, regardless of the covariance. We illustrated the application of our method for gene-filtering in time-course micro-array data using lung cancer data, where measurements were taken from normal and chemically treated human lung tissues. We were able to identify 1053 (4.7%) significant ones out of 22,227 genes.

Les tests traditionnels permettant l'analyse de données longitudinales ne sont pas applicables dans les situations de dimensions élevées (où $n < p$). Nous proposons un test modéré à l'aide du modèle de courbe de croissance, qui est utile pour analyser les données longitudinales de dimensions élevées. Des simulations exhaustives ont prouvé que notre test modéré est non biaisé et monotone par rapport à la taille de l'échantillon et à la valeur des paramètres, quelle que soit la covariance. Nous avons illustré l'application de notre méthode de filtrage des gènes aux données de micropuces de période par le biais des données sur le cancer du poumon, avec des mesures prises sur des tissus pulmonaires humains normaux et ayant subi un traitement chimique. Nous avons pu identifier 1 053 (4,7 %) gènes significatifs sur 22 227.

STEFAN KONIGORSKI (University of Toronto), **YILDIZ E. YILMAZ & SHELLEY B. BULL** (University of Toronto & Samuel Lunenfeld Research Institute of Mount Sinai Hospital)

[Monday May 27 / lundi 27 mai, 13:45]

A Censored Regression Method to Adjust for Medication Effects in Genetic Analysis of a Quantitative Trait When the Treatment Received Depends on the Trait

Méthode de régression avec données censurées pour tenir compte des effets des médicaments dans l'analyse génétique d'un caractère quantitatif lorsque le traitement dépend du caractère

In some observational studies, receiving a treatment or not is response-dependent. For example, the use of antihypertensive medication to lower blood pressure (BP) depends on the individual's pre-treatment BP measure. In genetic association analysis of BP, adjusting BP for the effect of medication is crucial when the objective is to iden-

Dans certaines études d'observation, le fait de recevoir ou non un traitement dépend de la réponse. Par exemple, l'utilisation de médicaments antihypertenseurs pour abaisser la tension dépend de la mesure de tension de l'individu avant le traitement. Dans une analyse d'association génétique de la tension, il est essentiel d'ajuster la tension en fonction de l'effet des médicaments si l'objectif est d'iden-

1C-C1: Biostatistics - Theory and Methods II

Biostatistique - théorie et méthodes II

tify genes associated with high or low BP. A naive analysis based on modeling BP with the treatment as a covariate leads to biased estimates of genetic effects. We propose a new method based on censored regression, assuming that a treated individual's true underlying BP is higher than the observed.

tifier les gènes associés à une tension élevée ou réduite. Une analyse naïve, fondée sur une modélisation de la tension avec le traitement comme simple covariable, induirait un biais dans l'estimation des effets génétiques. Nous proposons une nouvelle méthode fondée sur la régression avec données censurées, en prenant pour hypothèse que la tension réelle sous-jacente d'un individu traité est supérieure à celle observée.

CHEL HEE LEE & MIKELIS BICKIS (University of Saskatchewan)

[Monday May 27 / lundi 27 mai, 14:00]

Zero-Truncated Poisson Imprecise Probability Model for Quantifying Prior Ignorance

Modèle de probabilité imprécise de Poisson tronqué en zéro pour quantifier l'ignorance préalable

Prevalence estimates based on the intensity of usage of medical services are biased because zero counts are not observed. Such zero counts can be accommodated using a zero-truncated Poisson model. We quantify our epistemic ignorance by applying the imprecise probability inferential paradigm of Walley to this model. Normal and log-gamma priors are studied by constructing a three parameter exponential family of distributions which includes both priors. Finally, we present a visualization of updating as a translation on the hyperparameter space of this family of posteriors and graphically demonstrate the reduction of imprecision.

Les estimations de la prévalence fondées sur l'intensité de l'utilisation des services médicaux sont biaisées parce que les zéros ne sont pas observés. Ces zéros peuvent être inclus à l'aide d'un modèle de Poisson tronqué en zéro. Nous quantifions notre ignorance épistémique en appliquant le paradigme inférentiel de probabilité imprécise de Walley à ce modèle. Nous étudions les a priori normaux et log-normaux en construisant une famille de distributions exponentielles à trois paramètres qui inclut les deux a priori. Enfin, nous présentons une visualisation de la mise à jour comme une translation sur l'espace des hyperparamètres de cette famille des a posteriori et démontrons graphiquement la réduction de l'imprécision.

ZHIHUI (AMY) LIU, JAMES HANLEY, NANDINI DENDUKURI & ERIN STRUMPF (McGill University)

[Monday May 27 / lundi 27 mai, 14:15]

Measuring the Yearly Mortality Reductions Due to a Cancer Screening Program: A Round-by-Round Approach

Mesure de la réduction de mortalité annuelle due à un programme de dépistage du cancer : approche par ronde

Summary statistics from cancer screening trials are of limited use for projecting the mortality reduction pattern expected from a sustained screening program. Earlier, to project the timing, duration, and magnitude of the reductions under any specified regimen, we had reformulated a 3-parameter model, characterizing the impact of each round of screening, that can be fitted to year-by-year mortality data. Now, we extend it to (i) accommodate trials with >1 screening arm, (ii) combine information across trials with different regimens, (iii) account for age-specific regimens. We quantify how precisely the sustained asymptote in the mortality

Les statistiques sommaires des essais de dépistage du cancer sont d'une utilité limitée pour projeter l'effet de réduction de la mortalité escompté d'un programme de dépistage continu. Pour projeter le moment, la durée et l'amplitude des réductions résultant d'un régime spécifique, nous avons formulé un modèle à trois paramètres, caractérisant l'impact de chaque ronde de dépistage, qui pouvait être adapté aux données de mortalité par année. Nous étendons maintenant ce modèle pour (i) tenir compte des essais avec branche >1 de dépistage, (ii) combiner des informations provenant d'essais ayant des régimes différents, (iii) tenir compte des régimes spécifiques à l'âge. Nous quantifions la précision de mesure de l'asymptote soutenue dans la courbe de réduc-

1C-C1: Biostatistics - Theory and Methods II

Biostatistique - théorie et méthodes II

reduction curve can be measured.

tion de la mortalité.

MARYAM ZOLGHADR (University of Gothenburg), **SERGEI ZUYEV** (Chalmers University of Technology)

[Monday May 27 / lundi 27 mai, 14:30]

Optimal Design of Dilution Experiments under Volume Constraints

Plan optimal pour expériences de dilution sous contraintes de volume

We develop methods to construct a one-stage design of dilution experiment under the total available volume constraint typical for bio-medical applications. We consider different optimality criteria based on the Fisher information in both non-Bayesian and Bayesian settings. It turns out that the optimal design is one atomic, meaning that all the dilutions should be of the same size. Our proposed approach to solve such optimization problems is a variational analysis of functionals of a measure. The advantage of the measure optimization approach is that additional requirements like a total cost of experiment can be easily incorporated into the goal function.

Nous développons des méthodes pour concevoir un plan d'expérience de dilution en une étape sous la contrainte de volume maximal disponible pour les applications biomédicales. Nous examinons divers critères d'optimalité basés sur l'information de Fisher dans les contextes bayésien et non bayésien. Il s'avère que le plan optimal est atomique, ce qui signifie que toutes les dilutions devraient être de la même taille. L'approche proposée pour résoudre ce type de problème d'optimisation consiste en une analyse variationnelle des fonctionnelles d'une mesure. L'avantage de cette approche est que les exigences supplémentaires, comme le coût total de l'expérience, peuvent facilement être intégrées dans la fonction cible.

1C-C2: Business and Industrial Statistics I Statistique industrielle et de gestion I

Chair/Président: Jason Loeppky (University of British Columbia)

Room/Salle: EdSouth 221

Abstracts/Résumés

YUMI KONDO & JAMES V. ZIDEK (University of British Columbia)

[Monday May 27 / lundi 27 mai, 13:30]

Bayesian Nonparametric Subset Selection Procedures with Weibull Components

Procédures de sélection de sous-ensembles non-paramétriques bayésiens avec les composantes de Weibull

Our paper presents two Bayesian solutions to the problem of selecting a subset of populations of lumber breaking strengths from a specified finite set of populations so as to contain the one with the smallest (or equivalently from a mathematical perspective) largest α^{th} quantile. Based on independent random samples from each, estimates of these quantiles are obtained and used to specify conservative design values for engineering applications. In conformance with the ASTM, Inc standards requirement that requires a nonparametric approach in this application, semiparametric and nonparametric alternatives are described and empirically compared. Conclusions are given.

Notre article présente deux solutions bayésiennes au problème de sélection de sous-ensembles de populations de résistances à la rupture du bois provenant d'un ensemble fini de populations pour contenir celui avec le plus petit (ou de la même façon avec une perspective mathématique) ou le plus grand quantile. À partir d'échantillons aléatoires provenant de chaque ensemble, on obtient les estimations de ces quantiles qui permettent de spécifier les valeurs de calcul conservatrices à des fins d'ingénierie. Conformément aux normes de l'American Society for Testing and Materials Inc (ASTM) qui exige une approche non-paramétrique dans cette application, on décrira et comparera de façon empirique des alternatives semi-paramétriques et non-paramétriques. Nous présenterons nos conclusions.

MARK KRZEMINSKI & XIAOJIAN XU (Brock University)

[Monday May 27 / lundi 27 mai, 13:45]

Optimal Designs for Accelerated Life Tests with Multiple Stress Factors

Plans optimaux pour des tests de durée de vie accélérée avec plusieurs facteurs de stress

We present optimal designs for time-censored two-factor ALT to estimate percentiles of product life at usage conditions. We assume a Weibull life distribution, log-linear life-stress relationships with non-constant shape parameter for both stress factors. We consider both cases when the parameters in the relationship between the shape parameter and stress-factors are prespecified; and when these parameters must be estimated. The primary optimality criterion is minimization of asymptotic variance of MLE of life percentile at usage stress. We also discuss secondary criteria: D-optimality and A-optimality. Designs are illustrated with examples and a comparison study presented.

Nous présentons des plans optimaux de tests de durée de vie accélérée à deux facteurs censurés dans le temps permettant d'estimer les centiles de la durée de vie d'un produit utilisé. Nous supposons une distribution de vie de Weibull, une relation log-linéaire entre la durée de vie et le stress avec le paramètre de forme non constant pour les deux facteurs de stress. Nous considérons ces deux cas lorsque le paramètre dans la relation entre les paramètres de forme et les facteurs de stress sont spécifiés au préalable et lorsque ces paramètres doivent être estimés. Le premier critère primaire d'optimalité est la minimisation de la variance asymptotique de l'estimateur du maximum de vraisemblance (EMV) du centile de durée de vie soumis au stress. Nous aborderons également les critères secon-

1C-C2: Business and Industrial Statistics I Statistique industrielle et de gestion I

naires d'optimalité D et A. Nous illustrerons ces plans par des exemples et présenterons une étude comparative.

BLAKE MACDONALD, PRITAM RANJAN & HUGH CHIPMAN (Acadia University)

[Monday May 27 / lundi 27 mai, 14:00]

GPfit: An R Package for Fitting Gaussian Process Models to Simulator Data

GPfit : un package R pour l'ajustement des modèles de processus gaussien pour données de simulateur

Gaussian process (GP) models are commonly used for emulating expensive computer simulators. Ranjan, Haynes, and Karsten (2011) proposed a computationally stable approach for fitting GP models to deterministic simulator outputs using a robust but intensive genetic algorithm approach for likelihood maximization. Our new R package GPfit implements a modified version of the model proposed by Ranjan et al. (2011). Speed improvements are achieved using a novel parameterization of the spatial correlation and a new multi-start gradient based algorithm over the genetic algorithm method. The main features of GPfit are illustrated along with a comparison against the popular R package mlegp.

Les modèles de processus gaussien sont souvent utilisés pour émuler des simulateurs d'ordinateurs chers. Ranjan, Haynes et Karsten (2011) ont proposé une approche stable au niveau calculatoire pour ajuster les modèles de processus gaussien aux résultats de simulateur déterministes à l'aide d'une approche algorithmique génétique intensive pour la maximisation de la vraisemblance. Notre nouveau package R GPfit offre une version modifiée du modèle proposé par Ranjan et coll. (2011). La vitesse est améliorée grâce à un nouveau paramétrage de la corrélation spatiale et à un nouvel algorithme de gradient à plusieurs points de départ sur la méthode algorithmique génétique. Les caractéristiques principales de GPfit sont illustrées et comparées au populaire package R mlegp.

NATHANIEL PAYNE & TOM LOUGHIN (Simon Fraser University)

[Monday May 27 / lundi 27 mai, 14:15]

Development of a Structural Wood Adhesive Testing Standard Using a Nested Generalized Linear Mixed Model

Développement d'une norme d'essai d'adhésif pour le bois de charpente à l'aide d'un modèle linéaire généralisé mixte emboîté

A new Canadian standard for testing exterior structural wood adhesives (O112.9) has recently been published. This standard requires the analysis of quartiles from percentage wood failure data which are non-normal and are subject to multiple nested sources of variation. In response, we develop a binomial generalized linear mixed model that approximates percentage wood failure. We then use parametric bootstraps to evaluate the fit of the model and approximate the distribution of the required quartiles. Finally, we fit the model and demonstrate the results on data from an experiment conducted by researchers at FPInnovations.

Une nouvelle norme canadienne pour tester les adhésifs de bois de charpente d'extérieur (O112.9) a récemment été publiée. Cette norme nécessite l'analyse de quartiles provenant de données en pourcentage de rupture dans le bois qui sont non-normales et soumises à de nombreuses sources emboîtées de variation. Dans cette optique, nous avons développé un modèle mixte linéaire généralisé binomial qui estime le pourcentage de rupture dans le bois. Nous avons ensuite utilisé des méthodes bootstrap paramétriques pour évaluer l'ajustement du modèle et approximer des quartiles nécessaires. Enfin, nous avons ajusté le modèle et illustré les résultats avec des données provenant d'une expérience menée par les chercheurs à FPInnovations.

1C-C2: Business and Industrial Statistics I Statistique industrielle et de gestion I

JIMMY POON & YI LU (Simon Fraser University)

[Monday May 27 / lundi 27 mai, 14:30]

A Spatial Cross-Sectional Credibility Model with Dependence Among Risks

Modèle de crédibilité transversal spatial avec dépendance entre les risques

A Bühlmann-Straub type credibility model with dependence structure among risk parameters and conditional spatial cross-sectional dependence is studied. Predictors of future losses for the model under both types of dependence are derived by minimizing the expected quadratic loss function and non-parametric estimators of structural parameters are considered in the spatial statistics context. Predictions and estimations made for the proposed model are examined and compared to other models in an application with crop insurance data and a simulation study.

On étudie un modèle de crédibilité de Bühlmann-Straub à l'aide d'une structure de dépendance entre les paramètres de risque et la dépendance transversale spatiale conditionnelle. Des prédicteurs de pertes futures pour le modèle sous les deux types de dépendance sont obtenus en minimisant la fonction quadratique de perte espérée. Des estimateurs non-paramétriques de paramètres structurels sont également pris en compte dans le contexte de statistiques spatiales. On examine les prédictions et les estimations du modèle proposé et on les compare à d'autres modèles dans une application à des données d'assurance-récolte et dans une étude de simulation.

NEIL SPENCER (Acadia University)

[Monday May 27 / lundi 27 mai, 14:45]

Isomorphism Check for Two-Level Star-Based Factorial Designs with Randomization Restrictions

Vérification de l'isomorphisme pour plans factoriels à deux niveaux et en étoile avec restrictions de randomisation

Checking isomorphism of factorial designs with randomization restrictions is a challenging problem in design of experiments. In this talk, I focus on the isomorphism of two balanced stars of $PG(p-1,2)$, or equivalently, a 2^p factorial experiment. A balanced $(t-1)$ -star of $PG(p-1,2)$ is a collection of distinct $(t-1)$ -dimensional subspaces of $PG(p-1,2)$ with a common overlap that forms a cover of $PG(p-1,2)$. I propose a new isomorphism check algorithm. The novelty of my approach lies in the use of the underlying geometry to reduce the search space of all relabellings of $PG(p-1,2)$, and a new method for checking equivalence of stars.

Vérifier l'isomorphisme de plans factoriels avec restrictions de randomisation constitue un problème de taille dans la planification d'expériences. Dans cet article, je vais me concentrer sur l'isomorphisme de deux étoiles symétriques $PG(p-1,2)$, ou de la même façon, sur une expérience 2^p . Une étoile symétrique $(t-1)$ de $PG(p-1,2)$ est constituée de divers sous-espaces dimensionnels distincts $(t-1)$ - de $PG(p-1,2)$ avec un chevauchement fréquent qui forme une couverture de $PG(p-1,2)$. Je propose un nouvel algorithme de vérification d'isomorphisme. La nouveauté dans mon approche réside dans une géométrie sous-jacente pour réduire l'espace de recherche de tous les réétiquetages de $PG(p-1,2)$, ainsi qu'une nouvelle méthode de vérification de l'équivalence des étoiles.

1C-C3: Environmental Modeling: Applications and Case Studies Modélisation en environnement : applications et études de cas

Chair/Président: Jean-François Angers (Université de Montréal)

Room/Salle: EdSouth 164

Abstracts/Résumés

ALISHA ALBERT-GREEN (University of Western Ontario), **DOUGLAS WOOLFORD** (Wilfrid Laurier University), **W. JOHN BRAUN** (University of Western Ontario), **DAVID MARTELL** (University of Toronto)

[Monday May 27 / lundi 27 mai, 13:30]

Visualization Tools for Assessing the Markov Property: Sojourn Times in the Ontario Fire Weather Index

Outils de visualisation pour l'évaluation de la propriété markovienne: durées de séjour de l'indice forêt-météo en Ontario

In Canada, the Fire Weather Index (FWI) represents the potential frontal intensity of a forest fire. Mesoscale weather patterns suggest that the FWI switches between epochs of nil and non-nil behaviour. This research develops a framework for assessing sojourn times in these two phases which are, in general, not well approximated by a Markovian process. Using daily data from a sample of fire-weather stations across Ontario, we model the duration of epochs in each of these two phases using beta-geometric processes which incorporate seasonality of phase-specific run-length behaviour using local likelihood methods.

Au Canada, l'indice forêt-météo (IFM) représente l'intensité frontale potentielle d'un feu de forêt. Les situations météorologiques à moyenne échelle suggèrent que l'IFM oscille entre des époques de comportement nul et non nul. Cette recherche propose un cadre d'évaluation des durées de séjour dans chacune de ces phases qui, règle générale, sont mal approximées par un processus markovien. À l'aide de données quotidiennes provenant d'un échantillon de stations de météorologie forestière de l'Ontario, nous modélisons la durée des époques dans chacune des deux phases en utilisant des processus beta-géométriques qui intègrent la saisonnalité de la durée du cycle spécifique à chaque phase par des méthodes de vraisemblance locale.

WESLEY BURR (Queen's University), **HWASHIN H. SHIN** (Health Canada)

[Monday May 27 / lundi 27 mai, 13:45]

Temperature and Mortality: a Canadian Perspective

Température et mortalité: point de vue canadien

Much work has been completed in the past two decades on the relationship between air pollution, daily temperature and mortality. Temperature is a very strong predictor of mortality, and accounting for its effect before estimating the risk due to air pollution is an important part of the modeling procedure. The Distributed Lag Non-linear Model (dlnm) was recently (2010) developed to allow for inclusion of time structure in the exposure-response relationship between temperature and mortality for these models. We examine these dlnm-based models through an alternative smoother and estimation of phase between the two series produce some interesting results.

Beaucoup de travaux ont été réalisés au cours des deux dernières décennies sur la relation entre la pollution de l'air, la température quotidienne et la mortalité. La température est un excellent prédicteur de la mortalité et il est important de tenir compte de son effet dans le cadre de la modélisation avant d'estimer le risque lié à la pollution de l'air. Le modèle non linéaire à décalage distribué (dlnm) a été développé récemment (2010) pour permettre l'inclusion d'une structure temporelle dans la relation exposition-réponse entre la température et la mortalité. Nous examinons ces modèles par une méthode alternative de lissage et l'estimation de la phase entre les deux séries donne de bons résultats.

1C-C3: Environmental Modeling: Applications and Case Studies Modélisation en environnement : applications et études de cas

LAURA COWEN (University of Victoria), **TAKIS BESBEAS** & **BYRON MORGAN** (University of Kent), **CARL SCHWARZ** (Simon Fraser University)

[Monday May 27 / lundi 27 mai, 14:00]

A Comparison of Abundance Estimates Obtained from Extended Batch Marking and Jolly-Seber Type Experiments
Comparaison d'estimés d'abondance obtenus par marquage de lots étendus et expériences de type Jolly-Seber

Little attention has been paid to the use of multi-sample batch marked studies with the thought that an individual's capture history is necessary for unbiased precise estimates. We developed a model for the extended batch mark study. We use data simulated from a Jolly-Seber type study and convert this to what would have been obtained from an extended batch marking study. We compare abundance estimates obtained from the types of analyses.

Peu d'attention a été portée aux études de marquage de lots à échantillons multiples avec l'idée que l'historique de capture d'un individu est nécessaire pour obtenir des estimés précis non biaisés. Nous proposons un modèle destiné aux études de marquage de lots étendus. Nous utilisons des données simulées provenant d'une étude de type Jolly-Seber que nous convertissons sous la forme de ce que nous aurions obtenu d'une étude de marquage de lots étendus. Nous comparons les estimés d'abondance obtenus de ces analyses.

WENKAI (KEVIN) MA, W. JOHN BRAUN & GREGORY KOPP (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 14:15]

A Simulation Model for Wind-Driven Fire-Spotting
Modèle de simulation pour la dissémination du feu par le vent

Spotting is a dangerous and unpredictable characteristic of many forest wildfires. Firebrands are sent aloft by the convection column and carried by the wind, starting new fires. We develop a simulator for the landing distribution of burning firebrands through the use of a stochastic model which incorporates the physics of flight and burning.

La dissémination du feu est une caractéristique dangereuse et imprévisible de plusieurs feux de forêt. Des tisons s'élèvent dans une colonne de convection et sont emportés par le vent, allumant de nouveaux incendies. Nous présentons un simulateur de la distribution du lieu de dépôt des tisons ardents au moyen d'un modèle stochastique qui intègre la physique du vol et du brûlage.

SHAN QIAO (University of Saskatchewan)

[Monday May 27 / lundi 27 mai, 14:30]

Estimating the Effects of Air Pollutants on Recurrent Hospital Admission for Respiratory Diseases
Estimation des effets des polluants atmosphériques sur les admissions récurrentes à l'hôpital dues aux maladies respiratoires

Air pollution is a well-established risk factor for population health. It is beneficial for us to reveal whether the air quality poses a significant risk on our health system. Through conducting an analysis of recurrent events, we assess the impact of gaseous pollutants (e.g. SO_2 , NO_2) and particulate matters (PM_{10} , $\text{PM}_{2.5}$) on respiratory hospital readmissions. We present a multiplicative model that includes previous event counts as internal covariate components.

La pollution atmosphérique est un facteur de risque bien connu pour la santé de la population. Il est pertinent de vérifier si la qualité de l'air présente un risque pour notre système de santé. En procédant à l'analyse d'événements récurrents, nous évaluons l'impact des polluants gazeux (comme le SO_2 et le NO_2) et les matières particulaires (PM_{10} et $\text{PM}_{2.5}$) sur les réadmissions à l'hôpital liées à des problèmes respiratoires. Nous présentons un modèle multiplicatif qui intègre le dénombrement d'événements antérieurs comme une composante internes des covariables.

1C-C4: Probability I Probabilité I

Chair/Présidente: Gail Ivanoff (University of Ottawa)

Room/Salle: EdSouth 276

Abstracts/Résumés

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

[Monday May 27 / lundi 27 mai, 13:30]

Inference for Discretely Observed Levy-Driven Ornstein-Uhlenbeck Processes

Inférence pour des processus d'Ornstein-Uhlenbeck observés discrètement et gouvernés par des processus de Lévy

Levy-driven Ornstein-Uhlenbeck processes were introduced by Bandorff-Nielsen and Shephard (2001) to model stochastic volatility. The goal is to test the hypothesis that the (unobserved) driving process has uncorrelated increments. The Ornstein-Uhlenbeck process is observed at equally spaced discrete time points and this observed data can be used to approximate the increments of the unobserved driving process. We develop the asymptotic theory of an appropriate test statistic. We support our theoretical findings with simulations.

Les processus d'Ornstein-Uhlenbeck gouvernés par des processus de Lévy ont été introduits par Bandorff-Nielsen et Shephard (2001) pour modéliser la volatilité stochastique. L'objectif est de tester l'hypothèse selon laquelle le processus sous-jacent (non observé) possède des incréments non corrélés. Le processus d'Ornstein-Uhlenbeck est observé à des temps discrets équidistants et ces données observées peuvent être utilisées pour approximer les incréments du processus sous-jacent. Nous développons la théorie asymptotique d'une statistique de test appropriée. Nous appuyons nos résultats théoriques par voie de simulation.

FARID ELKTAIBI & GAIL IVANOFF (University of Ottawa)

[Monday May 27 / lundi 27 mai, 13:45]

Asymptotics for Linear Processes and Change-Point Detection

Comportement asymptotique pour des processus linéaires et la détection de points de rupture

The analysis of the limiting distribution of the sequential empirical process is a classic problem in statistics. However, most of the literature dealing with this problem for dependent observations involves assumptions on mixing or association. This can be avoided using martingale techniques to establish two functional central limit theorems for both a stationary causal linear process and the case considering a change in the marginal distribution. Lastly, we propose a nonparametric approach to detect a change in the marginal distribution function. We consider two consistent tests based on the Kolmogorov-Smirnov type statistic and the Cramér-Von Mises statistic.

L'analyse de la distribution limite du processus empirique séquentiel est un problème classique en statistique. Cependant, la majorité de la littérature traitant de ce problème pour des observations dépendantes comprend des hypothèses sur le mélange ou l'association. Ce problème peut être évité en utilisant des techniques de martingale pour établir deux théorèmes centraux limites fonctionnels à la fois pour un processus linéaire stationnaire de causalité et pour le cas considérant un changement dans la distribution marginale. Enfin, nous proposons une approche non paramétrique pour détecter un changement dans la fonction de distribution marginale. Nous considérons deux tests convergents basés sur la statistique de type Kolmogorov-Smirnov et la statistique de Cramér-Von Mises.

1C-C4: Probability I Probabilité I

CRAIG KAINULAINEN (Lakehead University)

[Monday May 27 / lundi 27 mai, 14:00]

Average Distance Results in Geometric Figures and Compact Connected Spaces

Distances moyennes dans des figures géométriques et des espaces compacts connexes

In this talk we look at the problem of finding the expected average distance within compact convex regions when two independent and uniformly distributed points are chosen at random. This distance is first calculated for some basic geometrical figures under the Euclidean metric, with the results given in terms of the diameter of each region. For L_p spaces, the average distance between a fixed point and any other random point is derived when $p = 1$ and $p = \infty$, and then used to calculate the average distance between any two random points in each case.

Dans cet exposé, nous examinons le problème de la détermination de la distance moyenne espérée dans des régions convexes compactes lorsque deux points indépendants et uniformément répartis sont choisis au hasard. Cette distance est d'abord calculée pour certaines figures géométriques de base selon la métrique euclidienne et les résultats sont donnés en termes de diamètre de chaque région. Pour les espaces L_p , la distance moyenne entre un point fixe et tout autre point aléatoire est dérivée lorsque $p = 1$ et $p = \infty$, puis utilisée pour calculer la distance moyenne entre deux points aléatoires dans chaque cas.

DELI LI (Lakehead University), **OLEG KLESOV** (National Technical University of Ukraine Kiev Polytechnic Institute), **GEORGE STOICA** (University of New Brunswick)

[Monday May 27 / lundi 27 mai, 14:15]

On the Central Limit Theorem along Subsequences of Sums of I.I.D. Random Variables

À propos du théorème central limite pour des sous-suites de sommes de variables aléatoires i.i.d.

In this paper we show that, for given positive non-decreasing sequence $\{a_n; n \geq 1\}$ with $\lim_{n \rightarrow \infty} a_n = \infty$ and $\lim_{n \rightarrow \infty} a_{n+1}/a_n = 1$ and given nondecreasing function $h(\cdot) : (0, \infty) \rightarrow (0, \infty)$ with $\lim_{x \rightarrow \infty} h(x) = \infty$, there exists a sequence $\{X, X_n; n \geq 1\}$ of symmetric i.i.d. random variables such that $\mathbb{E}h(|X|) = \infty$ and, for some subsequence $\{n_k; k \geq 1\}$ of $\{1, 2, \dots\}$, $S_{n_k}/\sqrt{n_k a_{n_k}} \Rightarrow N(0, 1)$, where $S_n = \sum_{i=1}^n X_i, n \geq 1$.

Dans cet article, nous montrons que, pour la suite positive non décroissante donnée $\{a_n; n \geq 1\}$ avec $\lim_{n \rightarrow \infty} a_n = \infty$ et $\lim_{n \rightarrow \infty} a_{n+1}/a_n = 1$ et la fonction non décroissante donnée $h(\cdot) : (0, \infty) \rightarrow (0, \infty)$ avec $\lim_{x \rightarrow \infty} h(x) = \infty$, il existe une suite $\{X, X_n; n \geq 1\}$ de variables aléatoires symétriques i.i.d. telle que $\mathbb{E}h(|X|) = \infty$ et, pour certaines sous-suites $\{n_k; k \geq 1\}$ de $\{1, 2, \dots\}$, $S_{n_k}/\sqrt{n_k a_{n_k}} \Rightarrow N(0, 1)$, où $S_n = \sum_{i=1}^n X_i, n \geq 1$.

WEIWEI LIU, REG KULPERGER & HAO YU (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 14:30]

Asymptotic Theory for GARCH-in-Mean Processes

Théorie asymptotique pour des processus GARCH en moyenne

The GARCH-in-mean process is an important extension of the standard GARCH (generalized autoregressive conditional heteroscedastic) process and has wide applications in finance and economics. Estimation of GARCH type models usually involves quasi-maximum likelihood (QML) estimation as it produces consistent and asymptotically normally distributed estimators under certain regu-

Le processus GARCH en moyenne constitue une extension importante du processus GARCH (generalized autoregressive conditional heteroscedastic) classique et présente un vaste éventail d'applications en finance et en économie. L'estimation de modèles de type GARCH implique généralement l'estimation du quasi-maximum de vraisemblance (QMV) car il produit des estimateurs convergents et dont la distribution asymptotique est normale sous certaines condi-

1C-C4: Probability I Probabilité I

larity conditions. The focus of this work is to establish a set of conditions under which the QML estimator of GARCH-in-mean process is strongly consistent with a Gaussian limiting distribution. Some general Markov model tools are utilized to derive the result.

tions de régularité. Le but de cette recherche est d'établir un ensemble de conditions en vertu desquelles l'estimateur QML du processus GARCH en moyenne présente une convergence forte et une distribution limite gaussienne. Certains outils généraux pour les modèles de Markov sont utilisés pour obtenir le résultat.

SAMIRA SADEGHI (University of Alberta)

[Monday May 27 / lundi 27 mai, 14:45]

On The Almost Sure Rates of Convergence of Linear Algorithms with Slowly Decreasing Gain

À propos des taux quasi-certains de convergence d'algorithmes linéaires avec gain en décroissance lente

Herein, we study almost sure rates of convergence of linear algorithms $h_{k+1} = h_k + \frac{1}{k^\alpha}(b_k - A_k h_k)$, where $\alpha \in (0, 1)$, $\{A_k\}_{k=1}^\infty$ are symmetric, positive-semidefinite random matrices and $\{b_k\}_{k=1}^\infty$ are random vectors. We show $|h_k - A^{-1}b| = o(k^{-\gamma})$ a.s. $\forall \gamma \in [0, \alpha)$ such that $\frac{1}{n^{\alpha-\gamma}} \sum_{k=1}^n (A_k - A) \rightarrow 0$ and $\frac{1}{n^{\alpha-\gamma}} \sum_{k=1}^n (b_k - b) \rightarrow 0$ a.s. for some positive-definite A and vector b . When $\alpha - \gamma \in (\frac{1}{2}, 1)$, we verify our assumptions using the Marcinkiewicz strong law of large numbers, which allows $\{A_k\}$ and $\{b_k\}$ to have heavy-tails, provided $E\|A_k\|^{\frac{1}{\alpha-\gamma}} < \infty$ and $E|b_k|^{\frac{1}{\alpha-\gamma}} < \infty$. Finally, we provide experimental results that confirm our convergence results and give some decreasing-gain design based upon our theoretical results.

Ici, nous étudions les taux quasi-certains de convergence des algorithmes linéaires $h_{k+1} = h_k + \frac{1}{k^\alpha}(b_k - A_k h_k)$, où $\alpha \in (0, 1)$, $\{A_k\}_{k=1}^\infty$ sont des matrices symétriques aléatoires semi-définies positives et $\{b_k\}_{k=1}^\infty$ sont des vecteurs aléatoires. Nous montrons que $|h_k - A^{-1}b| = o(k^{-\gamma})$ p. s. $\forall \gamma \in [0, \alpha)$ tel que $\frac{1}{n^{\alpha-\gamma}} \sum_{k=1}^n (A_k - A) \rightarrow 0$ et $\frac{1}{n^{\alpha-\gamma}} \sum_{k=1}^n (b_k - b) \rightarrow 0$ p. s. pour une certaine matrice définie positive A et un vecteur b . Lorsque $\alpha - \gamma \in (\frac{1}{2}, 1)$, nous vérifions nos hypothèses à l'aide de la loi forte des grands nombres de Marcinkiewicz, laquelle admet que $\{A_k\}$ et $\{b_k\}$ aient des queues lourdes, en autant que $E\|A_k\|^{\frac{1}{\alpha-\gamma}} < \infty$ et $E|b_k|^{\frac{1}{\alpha-\gamma}} < \infty$. Finalement, nous présentons des résultats expérimentaux qui confirment le taux de convergence et présentent un gain décroissant d'après nos résultats théoriques.

1C-C5: Survey: Theory, Methods and Applications Échantillonnage : théorie, méthodes et applications

Chair/Président: Changbao Wu (University of Waterloo)

Room/Salle: EdSouth 228

Abstracts/Résumés

MIN CHEN, CHANGBAO WU & MARY THOMPSON (University of Waterloo)

[Monday May 27 / lundi 27 mai, 13:30]

Pseudo Empirical Likelihood Method for Pretest-Posttest Studies Using Survey Data

Méthode de pseudo-vraisemblance empirique pour les études prétest/post-test avec données d'enquête

Pretest-posttest studies are an important and popular method for evaluating a treatment effect. The empirical likelihood method is a useful inference tool for such studies wherein information on a set of baseline variables can be effectively incorporated. We extend the methodology to apply to social or psychological studies with data collected through complex longitudinal surveys. Our proposed pseudo empirical likelihood method can be used to assess change of people's behaviour or response patterns due to the implementation of new public policies or a change in data collection protocol. The method is illustrated using data from the International Tobacco Control Project.

Les études prétest/post-test représentent une méthode importante et populaire pour évaluer l'effet d'un traitement. La méthode de vraisemblance empirique est un outil d'inférence utile pour de telles études dans lesquelles des informations sur un ensemble de variables initiales peuvent être intégrées de façon efficace. Nous étendons l'application de la méthodologie aux études sociales ou psychologiques avec des données recueillies par le biais d'enquêtes longitudinales complexes. Notre méthode basée sur la pseudo-vraisemblance empirique peut être utilisée pour évaluer le changement de comportement des gens ou les schémas de réponse découlant de la mise en œuvre de nouvelles politiques publiques ou d'un changement dans le protocole de collecte des données. La méthode est illustrée à l'aide des données du Projet international d'évaluation de la lutte anti-tabac.

GAURI SANKAR DATTA (University of Georgia, U.S. Bureau of the Census), **WILLIAM BELL & JERZY WIECZOREK** (U.S. Bureau of the Census)

[Monday May 27 / lundi 27 mai, 13:45]

A Weighted Likelihood Approach to Model-Based Small Area Estimation with Unit-Level Data

Approche de vraisemblance pondérée pour l'estimation de modèles pour petits domaines avec données au niveau des unités

Small area estimation uses area-level or unit-level models. Area-level models apply to direct survey estimates that are typically design consistent, leading to design consistent model predictions. Unit-level models, however, typically apply to survey microdata ignoring the sampling weights, not leading to design consistent model predictions. Kott, Rao and others developed methods to incorporate sampling weights in the unit-level normal nested error regression model to achieve design consis-

L'estimation pour petits domaines fait appel à des modèles au niveau des régions ou des unités. Les modèles au niveau des régions s'appliquent aux estimés d'enquêtes directs généralement convergents sous le plan, donnant ainsi des prévisions du modèle convergents également sous le plan. Cependant, les modèles au niveau des unités s'appliquent en général aux microdonnées d'enquête en ignorant les poids d'échantillonnage, ne conduisant pas à des prévisions convergentes sous le plan. Kott, Rao et d'autres ont développé des méthodes pour intégrer les poids d'échan-

1C-C5: Survey: Theory, Methods and Applications Échantillonnage : théorie, méthodes et applications

tency. We propose a pseudo-likelihood incorporating sampling weights that applies to normal or non-normal data, including binary and count data. Using a Bayesian approach, we apply our method to American Community Survey data.

tillonnage dans le modèle de régression avec erreurs normales imbriquées au niveau de l'unité afin d'atteindre la convergence sous le plan. Nous proposons une pseudo-vraisemblance incorporant les poids d'échantillonnage qui s'applique à des données normales ou non normales, y compris les données binaires et de dénombrement. En utilisant une approche bayésienne, nous appliquons notre méthode aux données de l'American Community Survey.

DARREN GRAY (Statistics Canada)

[Monday May 27 / lundi 27 mai, 14:00]

A Multi-Constraint Sample Allocation for Detailed National Estimates Made Trickier by Provincial Buy-Ins

Une répartition d'échantillon sous contraintes multiples pour des estimations nationales détaillées complexifiée par l'ajout de demandes de partenaires provinciaux

For the Survey of Innovation and Business Strategies (SIBS), the 2009 sample allocation was designed to satisfy precision requirements for three modules of questions, each of which targeted different industry and employment size group combinations at the national level. For the 2012 iteration of the survey, this already complex system of constraints was made more complicated when provincial specifications were added to the equation. A non-linear programming method was proposed to solve this complex problem. This paper will go over the methods used to ensure that all precision requirements were satisfied, while minimizing sample size and avoiding small stratum sizes.

Pour l'enquête sur l'innovation et les stratégies d'entreprise, l'allocation de l'échantillon en 2009 devait satisfaire les exigences de précision pour trois modules de questions ciblant chacun des combinaisons d'industries et groupes de taille d'emploi différents au niveau national. Pour l'enquête de 2012, ce système de contraintes déjà complexe s'est complexifié davantage par l'ajout de demandes provinciales. Une méthode de programmation non-linéaire a été proposée pour résoudre ce problème. Cet article discutera des méthodes employées pour s'assurer que toutes les exigences soient satisfaites tout en minimisant la taille d'échantillon et en évitant les strates de petite taille.

PIERRE-OLIVIER JULIEN, CHRISTIAN GAGNÉ & CARLOS A. LEON (Statistique Canada)

[Monday May 27 / lundi 27 mai, 14:15]

Negative Sample Coordination: An Overview of Some Methods

Coordination négative d'échantillons: un survol de quelques méthodes

A considerable amount of data is produced by statistical agencies, which can add a significant workload on respondents. To reduce respondent's burden, a number of methods for sample coordination of surveys based on a common frame have been developed. Statistics Canada is currently putting in place a central project consisting of more than 100 surveys based on its Business Register. Coordination of these samples would be possible from a common informatics tool. We present here an overview of some sample coordination methods which could be implemented in such a tool.

Une quantité considérable de données est produite par les agences statistiques. Ceci ajoute une charge de travail parfois importante aux répondants. Plusieurs méthodes de coordination d'échantillons tirés à même une base de sondage unique furent développées pour réduire leur fardeau de réponse. On échauffe présentement à Statistique Canada un projet central qui couvrira plus de 100 enquêtes basées sur le Registre des entreprises. La coordination de tous ces échantillons pourrait être possible à partir d'un outil informatique commun. Cette présentation se veut un survol de certaines méthodes qui pourraient être implantées dans un tel outil.

1C-C5: Survey: Theory, Methods and Applications Échantillonnage : théorie, méthodes et applications

CHRISTIAN OLIVIER NAMBEU & MIKE HIDIROGLOU (Statistique Canada)

[Monday May 27 / lundi 27 mai, 14:30]

Calibration Using Estimated Constraints

Calage utilisant des contraintes estimées

Calibration is an estimation approach intensively used in practice because of its many advantages. Very often, calibration totals are known quantities in the population. In this paper, we study the case when some known totals are replaced by estimated totals. The corresponding estimator is analysed and compared to standard calibration estimators in terms of bias and mean squared error both theoretically and by the means of a simulation.

Le calage est une approche d'estimation intensément utilisée en pratique à cause de ses nombreux avantages. Très souvent, les totaux de calage sont des quantités connues au niveau de la population. Dans cet article, nous étudions le cas où certains totaux sont remplacés par des totaux estimés. L'estimateur correspondant est analysé et comparé à des estimateurs par calage habituels en termes de biais et d'erreur quadratique moyenne théoriquement et via une étude par simulations.

1C-C6: Variable Selection and Statistical Learning Sélection de variables et apprentissage statistique

Chair/Président: John Koval (University of Western Ontario)

Room/Salle: EdSouth 177

Abstracts/Résumés

NASRIN AFZAL & XUEWEN LU (University of Calgary)

[Monday May 27 / lundi 27 mai, 13:30]

Variable Selection in a Partially Linear Additive Hazard Model with the Elastic Net Penalty

Sélection de variables pour un modèle partiellement linéaire à risques additifs avec pénalité en filet élastique

We consider the problem of simultaneous variable selection and estimation in a partially linear additive hazards model for survival data with a large number of covariates in the linear part. We utilize B-Splines to approximate the nonparametric function and the elastic net penalty as a regularization strategy. A cyclic coordinate descent algorithm is developed to fit the model. Simulation studies are conducted to assess the performance of the algorithm and a real data set is used for illustration.

Nous étudions les problèmes de sélection simultanée de variables et d'estimation pour les modèles partiellement linéaire à risques additifs pour des données de survie avec un nombre élevé de covariables dans la partie linéaire. Nous utilisons les B-splines pour approximer la fonction non paramétrique et la pénalité en filet élastique en tant que stratégie de régulation. Un algorithme de descente cyclique des coordonnées est développé pour ajuster le modèle. Des études de simulation permettent d'établir la performance de l'algorithme et un jeu de données réelles est utilisé pour illustrer la méthode.

BRIAN FRAN CZAK, RYAN BROWNE & PAUL MCNICHOLAS (University of Guelph)

[Monday May 27 / lundi 27 mai, 13:45]

Clustering and Classification of Asymmetric Data

Regroupement et classification de données asymétriques

A flexible model-based clustering and classification method is introduced as an alternative to the standard symmetric approaches. Capable of handling skewed or asymmetric data, we present an algorithm that uses a semi-parametric transform of the original data. A variant of the expectation-maximization algorithm is used to estimate the model parameters. The bayesian information criterion (BIC) and integrated complete likelihood (ICL) are compared for model selection. The capabilities of our models are demonstrated on both real and simulated data and the results compared to the popular symmetric (Gaussian and student- t) alternatives.

Une méthode flexible de regroupement et classification basée sur un modèle est présentée comme alternative aux approches symétriques standards. Nous présentons un algorithme utilisant une transformation semi-paramétrique des données qui peut accommoder des données asymétriques. Une variante de l'algorithme EM est utilisée pour estimer les paramètres. Le critère d'information de Bayes (BIC) et la vraisemblance complète intégrée (ICL) sont comparées pour la sélection de modèle. Les caractéristiques de nos modèles sont illustrées à l'aide de données réelles et simulées et les résultats sont comparés aux alternatives symétriques (gaussiennes et t de student) populaires.

1C-C6: Variable Selection and Statistical Learning Sélection de variables et apprentissage statistique

LONGLONG HUANG & XUEWEN LU (University of Calgary), **KAREN A. KOPCIUK** (Alberta Health Services)

[Monday May 27 / lundi 27 mai, 14:00]

Jackknife Empirical Likelihood Based Variable Selection for Accelerated Failure Time Model

Sélection de variables basée sur la vraisemblance empirique jackknife pour un modèle de temps de défaillance accéléré

We consider variable selection in the semiparametric accelerated failure time model. Due to a lack of a well defined likelihood function, we introduce the information criteria under jackknife empirical likelihood. Our method is based on a smoothed rank estimating function and on its representation as a U-statistic. We establish the asymptotic properties of the proposed method. We conduct simulation studies to demonstrate that the new method performs well compared to the parametric information criteria and is also robust to model misspecification. We analyze a real data set to illustrate the advantages of the new method in a practical setting.

Nous étudions la sélection de variables dans un modèle semi-paramétrique de temps de défaillance accéléré. À cause de l'absence d'une fonction de vraisemblance bien définie, nous proposons un critère d'information basé sur la vraisemblance empirique jackknife. Notre méthode s'appuie sur le lissage d'une fonction d'estimation basée sur les rangs et sa représentation en tant que U-statistique. Nous établissons les propriétés asymptotiques de la méthode proposée. Nous procédons à des études de simulation afin de démontrer que la nouvelle méthode donne de bons résultats lorsqu'elle est comparée aux critères d'information paramétriques et aussi, qu'elle est robuste à une mauvaise spécification du modèle. Nous analysons un jeu de données réelles afin d'illustrer les avantages pratiques de la nouvelle méthode.

ZI ZHEN LIU, REG KULPERGER & HAO YU (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 14:15]

The Adaptive Positive LASSO Procedure for Modelling ARCH Processes

Modélisation des processus ARCH par la méthode du LASSO positif adaptatif

The LASSO (Tibshirani, 1996) are popular in data analysis such as linear regression and Cox PH models for simultaneous variable selection and parameter estimation. Recently, some researchers have successfully applied the LASSO to autoregressive (AR) time series models. However, the results pertaining to applicability of the (adaptive) LASSO methodology to autoregressive heteroscedastic (ARCH) (Engle, 1982) time series models are not yet available in literature. We propose the Adaptive Positive LASSO procedure for modelling the pure ARCH processes that have some parameters being equal to zero. The proposed methodology shows promising results for ARCH models.

Le LASSO (Tibshirani, 1996) est populaire pour procéder simultanément à l'ajustement d'un modèle et à la sélection de variables dans des analyses de données telles que la régression et le modèle à risques proportionnels de Cox. Récemment, des chercheurs ont réussi à utiliser le LASSO pour des modèles de séries chronologiques autorégressives (AR). Cependant, la littérature ne présente pas encore de résultats indiquant que la méthodologie (adaptative) LASSO peut aussi s'appliquer aux modèles de séries chronologiques autorégressives hétéroscédastiques (ARCH) (Engle, 1982). Nous proposons la méthode du LASSO positif adaptatif pour modéliser les processus ARCH purs dont certains paramètres sont nuls. La méthodologie proposée montrent des résultats prometteurs pour les modèles ARCH.

1C-C6: Variable Selection and Statistical Learning Sélection de variables et apprentissage statistique

JABED HOSSAIN TOMAL, WILLIAM WELCH & RUBEN H. ZAMAR (University of British Columbia)

[Monday May 27 / lundi 27 mai, 14:30]

Ensembling Classification Models Based on Phalanxes of Variables with Applications in Drug Discovery
Apprentissage par ensemble pour des modèles de classification basés sur les phalanxes des variables avec application à la découverte de médicaments

We have proposed an ensemble which aggregates over groups of predictor variables. We form the groups (we call phalanxes) by joining variables together. The variables in a phalanx are good to put together, and the variables in different phalanxes are good to ensemble. We then build our ensemble by growing a random forest in each phalanx and aggregating them over the phalanxes. We have applied our ensemble to rank rare active compounds in four drug discovery datasets, and compared its performances with random forest. Our ensemble was found superior to random forest in most of the times.

Nous proposons un apprentissage par ensemble qui agrège en se basant sur des groupes de prédicteurs. Nous formons les groupes (nommés phalanxes) en joignant les variables entre elles. Les variables d'un phalanx vont bien ensemble et les variables de différents phalanxes peuvent être utilisées pour créer des ensembles. Nous bâtissons notre ensemble en ajustant une forêt aléatoire sur chaque phalanx, puis en les agrégeant sur tous les phalanxes. Nous avons appliqué nos ensembles pour classer les composés actifs rares dans quatre jeux de données sur la découverte de médicaments et avons comparé les résultats à ceux obtenus à l'aide d'une forêt aléatoire. Notre méthode d'ensemble s'est montrée supérieure à la forêt aléatoire la plupart du temps.

YUKUN ZHANG & XUEWEN LU (University of Calgary)

[Monday May 27 / lundi 27 mai, 14:45]

Coordinate Descent Methods for the Birnbaum-Saunders Regression Model with High-Dimensional Survival Data
Méthodes de descente des coordonnées pour le modèle de régression de Birnbaum-Saunders avec des données de survie de grande dimension

We consider variable selection in the Birnbaum-Saunders regression model for survival data with a large number of covariates. We introduce a pathwise algorithm via cyclical coordinate descent based on the elastic net penalty. The methods of stochastic EM and iterative least squares are incorporated in the algorithm for easy implementation. We evaluate the proposed method by simulation studies and use a real data analysis for illustration.

Nous étudions la sélection de variables dans le modèle de régression de Birnbaum-Saunders pour les données de survie avec un grand nombre de covariables. Nous proposons un algorithme par chemin de descente cyclique des coordonnées basé sur la pénalité en filet élastique. Les méthodes d'EM stochastique et de moindre carré itératif sont incorporés dans l'algorithme pour en faciliter la mise en oeuvre. Nous évaluons la méthode proposée par des études de simulation et l'illustrons à l'aide d'un jeu de données réelles.

1C-P: Poster Session Séance d'affichage

Room/Salle: Main floor atrium, Katz Group Centre

Abstracts/Résumés

TARYN ATHEY, PAUL MCNICHOLAS & ROBERT HANNER (University of Guelph)

[Monday May 27 / lundi 27 mai, 13:30]

Using Frequency Matrices for a Cross-Taxa Analysis of Errors in DNA Barcodes

Utilisation des matrices de fréquence pour une analyse des erreurs entre taxons dans le codage à barres moléculaire

DNA barcoding is a method of species identification utilizing a standardized gene region. Reference barcode sequences are uploaded into the Barcode of Life Data System (BOLD). Because BOLD is a reference library, accuracy of the DNA barcodes within BOLD is imperative to the application of DNA barcodes. Frequency matrices have been successfully used to assess the occurrence of errors in avian barcodes and to find pseudogenes within BOLD. This frequency matrix approach was used to assess rates and locations of errors within barcodes of different taxa. Assessing errors within the BOLD system is necessary to ensure an accurate reference library.

Le codage à barres moléculaire est une méthode servant à identifier les espèces en se basant sur les gènes normalisés d'une région. Les séquences de code à barres sont téléchargées dans le Barcode of Life Data System (BOLD). Puisque le BOLD est une bibliothèque de référence, l'exactitude des codes à barres moléculaires s'y trouvant est impérative aux applications de codage à barres moléculaire. Des matrices de fréquences ont été utilisées avec succès afin d'évaluer la fréquence d'erreurs dans les codes à barres moléculaires aviaires afin de trouver des pseudogènes dans le BOLD. Cette approche par matrice de fréquence a été utilisée afin d'évaluer les taux et les localisations des erreurs dans les codes à barres de différents taxons. Il est nécessaire d'évaluer les erreurs dans le BOLD afin d'assurer l'exactitude de cette bibliothèque de référence.

MAGGIE CHU & WENQING HE (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 13:30]

Cluster Analysis of Time-Course RNA-seq Data: Evaluation of Initialization

Analyse de regroupement pour la progression temporelle de données seq-ARN: évaluation et initialisation

Finite mixture models are commonly used in cluster analyses of genomic data and the expectation maximization (EM) algorithm is often used as a method for the maximum likelihood estimation. The performance of the EM algorithm can be improved by using good starting values. We employed the finite mixture model in clusterwise regression specifically for time-course discrete count data from RNA-seq experiments to partition genes, and present the evaluation of different initialization procedures for mixtures of negative binomial models with time-dependence effects in a simulation study. These procedures are illustrated using a real RNA-seq experiment on embryonic stages of fruit flies.

Les modèles de mélange finis sont couramment utilisés dans les analyses de regroupement de données génomiques et l'algorithme EM est souvent utilisé pour l'estimation par la méthode du maximum de vraisemblance. La performance de l'algorithme EM peut être améliorée en utilisant de bonnes valeurs de départ. Nous utilisons les modèles de mélange finis pour réaliser une régression par groupe pour la progression temporelle de données discrètes de dénombrement provenant d'expériences seq-ARN afin de partitionner les gènes, et nous présentons une simulation évaluant différentes procédures d'initialisation pour un mélange de modèles binomiaux négatifs avec un effet dépendant du temps. Ces procédures sont illustrées à l'aide d'une véritable expérience seq-ARN portant sur les stades embryonnaires de la mouche à fruits.

1C-P: Poster Session Séance d'affichage

BOLANLE DANSU & LISA M. LIX (University of Manitoba), **TOLULOPE SAJOBI** (University of Calgary)

[Monday May 27 / lundi 27 mai, 13:30]

Statistical Inference for Discriminant Function Coefficients in Doubly Multivariate Data

Inférence statistique pour les coefficients de la fonction discriminante avec des données doublement multivariées

Discriminant function coefficients (DFCs) from linear discriminant analysis are used to predict group membership and evaluate the relative importance of correlated variables for group separation. This study develops new statistical procedures to test for change in the relative importance of variables in doubly multivariate data, which are collected repeatedly over time. These DFC procedures use an asymptotic z-test, theoretical t-test, and empirical t-test based on a bootstrap critical value. A Monte Carlo study compares the Type I error and statistical power rates of the new procedures for two time points under different configurations of the group mean and covariance structures.

Les coefficients de la fonction discriminante (CFD) issus d'une analyse discriminante linéaire sont utilisés pour prédire l'appartenance à un groupe et évaluer l'importance relative des variables corrélées pour la séparation des groupes. Cette étude développe de nouvelles procédures statistiques pour tester un changement dans l'importance relative des variables pour des données doublement multivariées qui sont observées à répétition au cours du temps. Ces procédures pour les CFD utilisent un test z asymptotique, un test t théorique et un test t empirique basé sur des valeurs critiques issues du bootstrap. Une étude de Monte Carlo compare l'erreur de type I et la puissance des nouvelles procédures pour deux temps donnés sous différentes configurations de la moyenne et de la covariance des groupes.

SPECTRUM HAN (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 13:30]

Models for Crown Fire Rate of Spread

Modèles pour la vitesse de propagation des feux de cimes

A crownfire rate of spread (ROS) model has been proposed (Cruz et al, 2005) using windspeed, canopy bulk density (CBD) and moisture data. The data exhibit clustering and nonlinearity that was not accounted for. We re-analyze these data using a mixed-effects logistic model which successfully predicts whether a given crownfire is active or passive, and then predicts rate of spread using a nonlinear regression model. Data visualization helps identify the features of the new model. However, the ideal method would be to model wildfires with mixed effects at all stages. Because of limited data, this is not possible.

Un modèle pour la vitesse de propagation des feux de cimes (Cruz et al, 2005) est basé sur des données de vitesse du vent, de densité du couvert forestier et d'humidité. Les données présentent des regroupements et un manque de linéarité qui ont été ignorés. Nous analysons les données de nouveau en utilisant un modèle logistique à effets mixtes qui réussit à prédire si un feu de cime est actif ou passif, puis à prédire sa vitesse de propagation à l'aide d'un modèle de régression non linéaire. La visualisation des données aide à identifier les caractéristiques du nouveau modèle. Cependant, la méthode idéale consisterait à modéliser les feux de forêt avec des effets mixtes à tous les stades, ce qui est impossible à cause des données limitées.

ZHIYONG JIN (University of Manitoba)

[Monday May 27 / lundi 27 mai, 13:30]

A Fast Numerical Algorithm for Boundary Crossing Probabilities for Jump Diffusion Processes

Un algorithme numérique rapide pour les probabilités de traverser des frontières dans un processus de diffusion avec sauts

Based on the explicit formula derived by Wang and Shao (2012) for Boundary crossing probabilities

En nous basant sur la formule explicite de Wang et Shao (2012) pour les probabilités de traverser des frontières

1C-P: Poster Session Séance d'affichage

(BCP) of Brownian motion with random jumps, we implement a straightforward and fast numerical algorithm. Using Laplace transform, the formula and algorithm can be used to approximate more general diffusion processes such as Geometric Brownian Motion and Ornstein Uhlenbeck Process. Jump sizes can follow general distributions and even be correlated, and the boundary can be one side or two side. Using the formula and algorithm, we first compute the BCP for piecewise linear boundaries and then approximate the BCP for general nonlinear boundaries.

(PTF) d'un mouvement brownien avec des sauts aléatoires, nous mettons en oeuvre un algorithme numérique rapide et direct. En se servant de la transformée de Laplace, la formule et l'algorithme peuvent servir à approximer des processus de diffusion plus généraux tels qu'un mouvement brownien géométrique et le processus d'Ornstein-Uhlenbeck. La taille des sauts peut suivre une distribution arbitraire et même être corrélée tandis que les frontières peuvent être unilatérales ou bilatérales. À l'aide de la formule et de l'algorithme, nous calculons d'abord les PTF pour des frontières linéaires par morceaux, puis nous approximations les PTF pour des frontières non linéaires générales.

YANG LIU, MATÍAS SALIBIÁN-BARRERA, RUBEN H. ZAMAR & JAMES V. ZIDEK (University of British Columbia)
[Monday May 27 / lundi 27 mai, 13:30]

Lower Quantile Estimation via Subjectively Censored Weibull MLE with Bootstrap Threshold Selection
Estimation des quantiles inférieurs à l'aide d'un estimateur du maximum de vraisemblance pour le modèle Weibull censuré subjectivement dont la sélection des seuils est effectuée par le bootstrap

In the lower quantile estimation of the wood strength data, the current industrial standard applies a censored Weibull MLE approach instead of the intuitive parametric or non-parametric approaches, where the lower quantile is estimated with 90% of the data subjectively censored. Our study shows that this subjectively censored MLE can be viewed as a semi-parametric approach and achieves better balance between the bias and variance than the parametric or non-parametric quantile estimates. We also propose to further improve this approach by selecting the censoring threshold via a bootstrap estimate of the mean square error.

Pour l'estimation des quantiles inférieurs dans les données sur la résistance du bois, les normes industrielles courantes consistent à calculer un estimateur du maximum de vraisemblance pour le modèle Weibull censuré plutôt que d'adopter les approches intuitives paramétriques ou non paramétriques, où les quantiles inférieurs sont estimés avec 90 % des données censurées subjectivement. Notre étude montre que cet estimateur du maximum de vraisemblance censuré peut être vu comme une approche semi-paramétrique et obtenir un meilleur compromis entre le biais et la variance que les estimateurs paramétriques ou non paramétriques des quantiles. Nous proposons aussi d'améliorer davantage cette approche en choisissant le seuil de censure à l'aide d'un estimé bootstrap de l'erreur quadratique moyenne.

NICKCHI PAYMAN, PEZESHK HAMID & SHEMEHSAVAR SOUDABEH (University of Tehran)
[Monday May 27 / lundi 27 mai, 13:30]

A Bidirectional Hidden Markov Model Together with External Hints for Gene Prediction
Un modèle de Markov caché bidirectionnel avec indices extérieurs pour la prévision de gènes

Annotation of exponentially growing number of genomic sequences requires fast and reliable computing methods. Hidden Markov models are statistical tools applied to annotate genomic sequences. Using external hints from other sources than just the query sequence is a reliable job which could increase the accuracy of gene finding. We discuss a

L'annotation de séquences de gènes dont le nombre croît exponentiellement requiert des méthodes de calcul rapides et fiables. Les modèles de Markov cachés sont un des outils statistiques utilisés pour annoter les séquences génomiques. L'utilisation d'indices extérieurs à la séquence peut améliorer l'exactitude des découvertes de gènes. Nous étudions un modèle de Markov caché généralisé basé sur les

1C-P: Poster Session Séance d'affichage

generalized hidden Markov model based on external hints to detect genes in DNA sequences. The hints come from different sources such as EST or protein databases. In this work we introduce a bidirectional hidden Markov model which undertakes external hints to increase the precision of gene prediction in eukaryotes.

indices extérieurs afin de détecter des gènes dans les séquences d'ADN. Les indices proviennent de sources telles que les bases de données EST et protéiniques. Dans ce travail, nous proposons un modèle de Markov caché bidirectionnel qui utilise des indices extérieurs pour améliorer la précision des prévisions géniques chez les eucaryotes.

ADAM RAHMAN (Western University), **DUNCAN MURDOCH** (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 13:30]

Sweavetest - Automatic Multiple Choice Test Analysis

Sweavetest - Analyse automatique des tests à choix multiples

The Sweavetest package for R contains support functions to format, grade, and analyse multiple choice tests. In this poster, we will show the report that is automatically generated using the Sweavetest package, showing the descriptive statistics and exam reliability statistics for the exam as a whole, and presenting item analysis for each individual question. Errors in answer keys and poorly worded questions are easily recognized through the information generated. Several plots are also created to visually show the instructor how well the students are performing, to help identify possible problem questions, and to detect answer patterns that may suggest cheating.

La bibliothèque de fonctions R Sweavetest contient des fonctions pour formater, corriger et analyser des tests à choix multiples. Sur cette affiche, nous montrons le rapport qui est généré automatiquement par Sweavetest, présentant les statistiques descriptives et les statistiques de fiabilité pour l'examen au complet, puis une analyse par item pour les questions individuelles. Les erreurs dans les choix de réponses proposés et les questions mal conçues sont facilement reconnues grâce à l'information générée. Plusieurs graphiques sont aussi créés afin de montrer à l'enseignant le niveau de succès des étudiants, de l'aider à identifier des problèmes avec certaines questions et à détecter des patrons causés par la tricherie.

TIANJI SHI (University of British Columbia), **CIPRIAN PIRVU** (FPInnovations), **WILLIAM WELCH & JAMES V. ZIDEK** (University of British Columbia)

[Monday May 27 / lundi 27 mai, 13:30]

Tension Proofloading on Fingerjointed Lumber: Benefits and Optimization

Essais de charge pour la tension sur les pièces de bois à entures multiples: bénéfiques et optimisation

Tension proofloading is a quality control procedure designed to filter out weak and/or badly produced fingerjointed lumber pieces. It is carried out under combinations of stress level and rate. Although proofloading indeed destroys weaker lumber pieces, it may also affect the strength of lumber pieces that survive proofloading. One objective of our study is to statistically analyze the trade-off between the removal of weaker lumber and the possible damage inflicted on the survivors. Our other objective is to optimize the proofload conditions by modelling the theoretical tensile stress distribution of survivors under a range of *untried* proofload con-

Les essais de charge pour la tension sont une procédure de contrôle de la qualité conçue pour éliminer les pièces de bois à entures multiples faibles et/ou défectueuses. Ils sont effectués sous une combinaison de niveaux de stress et de taux. Même si les essais détruisent les pièces les plus faibles, ils peuvent aussi affaiblir la résistance des pièces restantes. Un objectif de notre étude consiste à analyser statistiquement le compromis à faire entre le retrait des pièces plus faibles et le dommage possiblement infligé aux pièces restantes. Notre autre objectif consiste à optimiser les conditions des essais en modélisation la distribution théorique des niveaux de stress des survivants sous un ensemble de conditions *non essayées*.

1C-P: Poster Session Séance d'affichage

ditions.

ZHENG SUN, RICHARD LOCKHART & MICHAEL STEPHENS (Simon Fraser University)

[Monday May 27 / lundi 27 mai, 13:30]

Testing Fit for Latent Variables

Tests d'ajustement pour des variables latentes

Many years ago Krumbein (1935) developed a method for estimating the diameters of rocks embedded in clay or similar materials, based on slices through the material and the rocks. This work is extended with the help of modern numerical techniques such as MCMC and the EM algorithm. A procedure is given to test the fit of latent (unobserved) variables to a given distribution, based on related observed variables. An application is to test if the unobserved errors in logistic regression are, for example, normally distributed.

Il y a plusieurs années, Krumbein (1935) a développé une méthode pour estimer le diamètre de roches encastrées dans l'argile ou dans d'autres matériaux similaires basée sur des tranches du substrat et des roches. Ce travail est généralisé avec l'aide de techniques numériques modernes telles que les MCMC et l'algorithme EM. Une procédure est développée afin de tester l'ajustement de variables latentes (non observées) à une distribution en se basant sur des variables observées reliées. Par exemple, une application consiste à tester si les erreurs non observées de la régression logistique sont normalement distribuées.

YANG TANG, PAUL MCNICHOLAS & RYAN BROWNE (University of Guelph)

[Monday May 27 / lundi 27 mai, 13:30]

Common Factors For Clustering Data of Mixed Type

Facteurs communs pour le regroupement de données de type mixte

Recent work on clustering and classification of data with mixed type, based on underlying Gaussian latent variables, is extended by drawing on ideas from common factor analyzers. Parameter estimates are derived using variational approximations and the BIC is used to select the number of components as well as the number of factors. Our approach is illustrated on real and simulated data.

Des travaux récents sur le regroupement et la classification de données de types mixtes basés sur l'hypothèse de variables latentes gaussiennes sous-jacentes sont généralisés en se servant de modèles dont la structure de covariance correspond à l'analyse factorielle. Les estimés des paramètres sont trouvés par une approximation variationnelle et le critère du BIC est utilisé pour choisir le nombre de composantes et le nombre de facteurs. Notre approche est illustrée à l'aide de données réelles et simulées.

JINKUN XIAO & IAN MCLEOD (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 13:30]

Improved Portmanteau Diagnostic Check for ARFIMA Time Series Models

Diagnostic de « Portmanteau » amélioré pour les modèles de séries chronologiques ARFIMA

The multivariate ARMA diagnostic test based on the generalized variance of the residual autocorrelation matrix was introduced by Peña and Rodriguez (2002) and developed by Mahdi, E. and McLeod, A.I. (2012). In this work we suggest an approximate method for computing the p-value for this test and we investigate the accuracy and the

Le test de diagnostic ARMA multivarié basé sur la variance généralisée des résidus de la matrice d'autocorrélation a été proposé par Peña et Rodriguez (2002) et développé par Mahdi, E. et McLeod, A.I. (2012). Dans ce travail, nous proposons une méthode de calcul approximative pour les p-values de ce test et nous étudions la précision et la puissance de ce test par simulation. Nos simulations comparent

1C-P: Poster Session Séance d'affichage

power of this test by simulation. Our simulations compare the approximate p-value method with the exact Monte-Carlo test. We also compare this new test with Box-Ljung portmanteau test for ARFIMA derived in Li and McLeod (1986). Since the simulations were extensive, serial farming on SharcNet was utilized.

la méthode approximative pour les p-values avec le test exact de Monte Carlo. Nous comparons aussi ce nouveau test avec le test Box-Ljung « portmanteau » pour les ARFIMA présenté par Li et McLeod (1986). Puisque le nombre de simulations était considérable, elles ont été effectuées en parallèle sur SharcNet.

LU XIN & MU ZHU (University of Waterloo), **HUGH CHIPMAN** (Acadia University)

[Monday May 27 / lundi 27 mai, 13:30]

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 real social network data.

Les données de réseaux transactionnels décrivent les communications entre les acteurs d'un réseau social. Plusieurs 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 de la modélisation par mélanges et de l'analyse d'événements récurrents, notre modèle est non seulement capable de détecter des communautés, mais il permet aussi de décrire comment les schémas de communications évoluent au sein de ces communautés. Nous illustrons les avantages de notre modèle à partir de données réelles sur les réseaux sociaux.

LIN XUE & DEPING JIANG (University of Manitoba)

[Monday May 27 / lundi 27 mai, 13:30]

Longitudinal Analysis of Development and Well-Being of Canadian Children and Youth

Analyse longitudinale du développement et du bien-être des enfants et de la jeunesse canadienne

This paper will use the data from the National Longitudinal Survey of Children and Youth (NLSCY), which follows the development and well-being of Canadian children. The study will focus on the prevalence of mental health problems among adolescence and youth over past decade. Random effect logistic model for longitudinal data will be conducted with or without using survey weights. Simulation studies will then be used to illustrate the differences in estimates of prevalence rate and change rate between using or without using the survey weight.

Cet article utilise les données de l'Enquête longitudinale nationale sur les enfants et les jeunes qui suit le développement et le bien-être des enfants canadiens. L'étude se penchera sur la prévalence de problèmes mentaux à l'adolescence et chez les jeunes au cours de la dernière décennie. Un modèle logistique à effets aléatoires pour données longitudinales est ajusté en utilisant ou non les poids de sondage. Des études de simulations illustrent la différence dans les estimés de la prévalence et de la vitesse des changements lorsque les poids de sondage sont utilisés ou non.

1D-I1: Advances in High-Dimensional Data Analysis Avancées en matière d'analyse de données de grande dimension

Chair/Présidente: Xiaojian Xu (Brock University)
Organizer/Responsable: Sanjoy Sinha (Carleton University)

Room/Salle: ECHA L1-140

Abstracts/Résumés

CHRISTINA KENDZIORSKI & NING LENG (University of Wisconsin - Madison)

[Monday May 27 / lundi 27 mai, 15:30]

EBSeq: An Empirical Bayesian Approach for Inference in RNA-seq Experiments

EBSeq : une approche bayésienne empirique pour l'inférence dans les expériences de séquençage d'ARN

High-throughput cDNA sequencing (RNA-seq) experiments allow for the identification of differentially expressed (DE) genes and their corresponding isoforms on a genome-wide scale. A number of methods have been developed for identifying DE genes in an RNA-seq experiment, but they are deficient for identifying DE isoforms, and most do not accommodate experiments measuring expression over time. We have developed an empirical Bayesian modeling approach for identifying differential expression in an RNA-seq experiment (EBSeq) comparing multiple biological conditions at a single time point, as well as over time. Evaluation via simulation and case studies demonstrates advantages in numerous settings.

Les expériences de séquençage d'ADNc à très haut débit (RNA-seq) permettent d'identifier les gènes exprimés différemment (ED) et leurs isoformes correspondantes à l'échelle du génome. Diverses méthodes existent pour identifier les gènes ED dans une expérience RNA-seq, mais elles n'identifient pas les isoformes ED et la plupart ne sont pas adaptées aux expériences mesurant l'expression au fil du temps. Nous avons mis au point une approche de modélisation bayésienne empirique pour identifier l'expression différentielle dans une expérience RNA-seq (EBSeq) comparant diverses conditions biologiques à un moment donné et au fil du temps. Une évaluation par le biais de simulations et d'études de cas démontre les avantages dans de nombreuses situations.

ABDUS SATTAR (Case Western Reserve University), **SHILI LIN & TASLIM CENNY** (Ohio State University), **SHELDON BAI & SHAROON HANOOK** (Case Western Reserve University)

[Monday May 27 / lundi 27 mai, 16:00]

Prediction of Genome-Wide Promoters and Enhancers Using a Flexible Semiparametric Mixed Effects Model

Prévision de promoteurs et amplificateurs génétiques via un modèle semi-paramétrique souple à effets mixtes

Sequencing data have emerged as the preferred type of molecular level data for biomedical research. Rapidly decreasing costs of next-generation sequencing (NGS) data, NGS data become a standard molecular data in studying genetics, genomics, and epigenomics. Our goal is to predict regulatory elements of gene expression such as promoters and enhancers using NGS data. We have predicted genome-wide common pattern of promoters and histone modification using a semi-parametric mixed effects model. NGS data from

Les données de séquençage sont devenues les données moléculaires de choix en recherche biomédicale. Avec la baisse du coût des données de séquençage de nouvelle génération (NGS), celles-ci sont aujourd'hui la norme en génétique, génomique et épigénomique. Notre objectif est de prévoir les éléments régulateurs de l'expression génétique (promoteurs et amplificateurs) à l'aide de données NGS. Nous avons prédit une tendance de promoteurs et de modification d'histone à l'échelle du génome à l'aide d'un modèle semi-paramétrique à effets mixtes. Nous utilisons des données NGS de cancers du sein et de la prostate pour nos

1D-I1: Advances in High-Dimensional Data Analysis Avancées en matière d'analyse de données de grande dimension

breast and prostate cancers are used in predicting prévisions.
promoters and enhancers.

SANJOY SINHA (Carleton University), **ABDUS SATTAR** (Case Western Reserve University)

[Monday May 27 / lundi 27 mai, 16:30]

Robust Methods for Selecting Differential miRNAs in Breast Cancer

Méthodes robustes de sélection de micro-ARN différentiels dans le cancer du sein

Genome-wide studies have been considered in recent years to profile microRNA (miRNA) expression levels derived from blood cells in search of miRNA signatures that can distinguish individuals with breast cancer from healthy controls. In this talk, innovative robust methods will be introduced for identifying differentially expressed miRNAs in breast cancer cells. The performance of the proposed methods will be investigated empirically using Monte Carlo simulations. An application will be presented using miRNA expression data obtained from breast cancer patients at the University Hospitals Case Medical Center along with healthy controls.

Récemment, des études pangénomiques ont été conçues pour profiler le niveau d'expression de micro-ARN dans les cellules sanguines, en vue de découvrir des signatures micro-ARN permettant de distinguer entre les individus souffrant de cancer du sein et les groupes témoins en santé. Dans cette présentation, nous introduisons des méthodes robustes novatrices permettant d'identifier des micro-ARN exprimés différemment dans les cellules de cancer du sein. Nous évaluons ces méthodes de façon empirique par des simulations de Monte Carlo. Nous en présentons une application avec des données d'expression micro-ARN obtenues de patientes atteintes d'un cancer du sein au University Hospitals Case Medical Center et des groupes témoins en santé.

1D-I2: Recent Developments in R packages Développements récents dans les packages R

Organizer and Chair / Responsable et président: Duncan Murdoch (University of Western Ontario)

Room/Salle: EdSouth 176

Abstracts/Résumés

PETER SOLYMOS, KHURRAM NADEEM & SUBHASH LELE (University of Alberta)

[Monday May 27 / lundi 27 mai, 15:30]

Complex Models in Ecology: Challenges and Solutions

Modèles complexes en écologie : difficultés et solutions

Ecological data and models are becoming increasingly complex due to methodological advances and data sharing. Dealing with missingness, observation and measurement error models are prime examples for creating challenging situations for modelers. I am going to present examples for such cases and walk through the solutions with special emphasis on software implementation and parameter identifiability issues. I am going to briefly introduce several R packages (dclone, dcmlc, PVA-Clone) which are the products of our efforts in making complex models more accessible to R users.

Dus aux progrès en méthodologie et partage des données, les données et les modèles écologiques deviennent de plus en plus complexes. Les données manquantes, l'observation et les modèles d'erreurs de mesures créent des situations difficiles pour les modélisateurs. Je présente des exemples de tels cas et les solutions possibles, en insistant sur les problèmes d'utilisation de logiciels et d'identification des paramètres. J'introduis brièvement plusieurs progiciels R (dclone, dcmlc, PVAClone) issus de nos efforts pour rendre les modèles complexes plus accessibles aux utilisateurs de R.

MICHAEL FRIENDLY & MATTHEW SIGAL (York University)

[Monday May 27 / lundi 27 mai, 16:00]

Recent Advances in Visualization for Multivariate Linear Models

Progrès récents en visualisation pour les modèles linéaires multivariés

This talk surveys our work developing visualization methods in R for understanding the effects of predictors in multivariate linear models (MLMs). We begin with a description of the HE plot framework implemented in the heplots package, wherein multivariate tests can be visualized via Hypothesis-Error ellipsoids. These ideas extend to multivariate repeated measures designs. When the rank of the hypothesis matrix for a term exceeds 2, these effects can be visualized in a reduced-rank canonical space via the candisc package. Finally, we discuss the extension of these methods to robust MLMs, incorporating generalizations of influence measures and diagnostic plots in mvinfluence.

Nous présentons nos travaux visant à développer des méthodes de visualisation en R permettant de mieux comprendre les effets des indicateurs dans les modèles linéaires multivariés (MLM). Nous commençons par décrire le cadre du tracé HE utilisé dans le progiciel heplots, qui permet de visualiser des tests multivariés par le biais des ellipsoïdes hypothèse/erreur. Ces idées s'étendent aux plans de mesures répétées à multiples variables. Si le rang de la matrice d'hypothèses d'un terme donné est supérieur à 2, ces effets peuvent être visualisés dans un espace canonique de rang réduit via le progiciel candisc. Enfin, nous discutons de l'extension de ces méthodes aux MLM robustes, en intégrant la généralisation des mesures d'influence et les tracés de diagnostic dans mvinfluence.

1D-I2: Recent Developments in R packages Développements récents dans les packages R

PAUL GILBERT (retired)

[Monday May 27 / lundi 27 mai, 16:30]

2piQA: Crowd-Sourcing R Package Quality Assessment

2piQA : Évaluation de la qualité du package R de « crowd-sourcing »

One often hears users complain about the difficulty of choosing which R packages to use. 2piQA provides a mechanism to leverage motivations producing exponential R package growth to also improve package quality. It provides a way to “vote”, by contributing work, rather than clicking “like”. User written test snippets are used to show that a package works, or does not work, as expected. Contributing snippets will be interesting not only for fun, and possibly to write articles about, but also for those who rely on a package, such as other package developers, or institutions that use packages for their daily work.

On entend souvent les utilisateurs se plaindre qu’il est difficile de choisir quel package R utiliser. 2piQA offre un mécanisme pour exploiter ces motivations, produire une croissance exponentielle des packages R et en améliorer la qualité. Il permet de « voter » en contribuant des travaux plutôt que de cliquer sur « J’aime ». Des extraits de code rédigés par les utilisateurs sont utilisés pour montrer que le package fonctionne ou ne fonctionne pas comme prévu. Ces extraits sont intéressants et utiles non seulement à des fins de divertissement ou de recherche, mais aussi pour les développeurs de packages et les institutions qui utilisent ces packages dans leur travail quotidien.

1D-I3: Some Thoughts on How to be More Successful Quelques pistes pour connaître un plus grand succès

Organizer and Chair / Responsable et présidente: Dena Schanzer (Public Health Agency of Canada)

Session sponsored by the Accreditation Committee
Séance parrainée par le Comité d'accréditation

Room/Salle: EdSouth 254

Description

In 2004, the SSC formally established the SSC Accreditation Program for professional statisticians who practice in Canada. Our 1st Accreditation sponsored session aims to engage our accredited members who apply statistical methods to solve real-world problems as well as our students and those who teach them in a discussion of soft skills for a successful career in the practice of statistics. Based on his 32-year career in industry and extensive involvement in professional societies, the speaker will provide a variety of career learnings and recommendations on how to be more successful in one's career. The many benefits and advantages of professional society involvement will be presented, and the recent emphasis of the American Statistical Association on non-technical skills will be highlighted.

En 2004, la SSC a formellement instauré son Programme d'accréditation destinés aux statisticiens professionnels pratiquant au Canada. Notre première séance parrainée sur l'accréditation vise à inviter nos membres accrédités, qui appliquent des méthodes statistiques à des problèmes réels, ainsi que nos étudiants et leurs professeurs à prendre part à une discussion axée sur les compétences générales contribuant à la réussite d'une carrière en statistique. En se fondant sur sa carrière de 32 ans dans l'industrie et sa vaste implication dans les associations professionnelles, le conférencier présente un éventail d'apprentissages et de recommandations permettant de connaître un plus grand succès professionnel. Il démontre les nombreux bienfaits et avantages de l'implication dans des associations professionnelles et expose l'accent mis récemment par l'American Statistical Association sur les compétences non-techniques.

1D-I4: Statistics and Ecology La statistique et l'écologie

Organizer and Chair / Responsable et président: Louis-Paul Rivest (Université Laval)

Room/Salle: ECHA 1-498

Abstracts/Résumés

JOANNA MILLS FLEMMING (Dalhousie University), **EVA CANTONI** (University of Geneva), **ALAN WELSH** (Australian National University)

[Monday May 27 / lundi 27 mai, 15:30]

Modelling and Prediction for Clustered Count Data with Excess Zeros

Modélisation et prévision pour les données de comptage regroupé avec zéros excessifs

Clustered counts with many zeros are typical of data on endangered species, particularly in marine environments, but are difficult to model with available methods. Here we present a general formulation of a mixed effects hurdle model where a novel approach to introducing random effects allows extensions beyond the usual multivariate normality assumption and facilitates inference about dependence between the two parts of the model. We develop empirical best predictors of the random effects and other cluster specific targets and assess their variability via a fast bootstrap. Our research is illustrated using data on critically endangered sharks in the North Atlantic.

Les dénombrements regroupés avec de nombreux zéros sont courants dans les données sur les espèces menacées, notamment en milieu marin, mais ces dénombrements sont difficiles à modéliser avec les méthodes disponibles. Nous présentons une formulation générale d'un modèle « hurdle » à effets mixtes : une façon inédite d'introduire les effets aléatoires permet de l'étendre au-delà de l'hypothèse de normalité multivariée habituelle et facilite l'inférence concernant la dépendance entre les deux parties du modèle. Nous déterminons les meilleurs prédicteurs empiriques des effets aléatoires et autres cibles spécifiques au groupe et évaluons la variabilité par le biais d'un bootstrap rapide. Nous illustrons nos recherches par des données sur les requins de l'Atlantique nord menacés d'extinction.

SUBHASH LELE & KHURRAM NADEEM (University of Alberta)

[Monday May 27 / lundi 27 mai, 16:00]

Population Viability Analysis in the Presence of Observation Error: Consequences and Statistical Inference

Analyse de viabilité de la population en présence d'erreur d'observation : conséquences et inférence statistique

Population Viability Analysis (PVA) is used to quantify extinction risk to study populations. These risk calculations strongly depend on the choice of the population growth model and the inclusion of demographic and environmental stochasticity. Another component that is critical is the observation error in the data. We show that inclusion of observation error can be critical in order to conduct proper model selection and PVA. We use data cloning to conduct likelihood based PVA in the presence of observation error, demographic and environmental stochasticity. Extinction risks predicted by with and without observation error models are quite different.

L'Analyse de viabilité de la population (AVP) permet de quantifier le risque d'extinction des populations étudiées. Ces calculs de risques dépendent fortement du choix du modèle de croissance de la population et de l'inclusion éventuelle du caractère stochastique démographique et environnementale. L'erreur d'observation est un autre élément clé. Nous montrons qu'il est essentiel d'inclure cette erreur d'observation pour garantir une sélection de modèle et une AVP appropriées. Nous clonons des données pour effectuer une AVP basée sur la probabilité en présence d'erreur d'observation et d'un caractère stochastique démographique et environnementale. Le risque d'extinction résultant des modèles avec et sans erreur d'observation est très différent.

1D-I4: Statistics and Ecology La statistique et l'écologie

THIERRY DUCHESNE, LOUIS-PAUL RIVEST & DANIEL FORTIN (Université Laval)

[Monday May 27 / lundi 27 mai, 16:30]

Statistical Analysis of Animal Movement

Analyse statistique du déplacement animal

A good understanding of the determinants of animal movement is crucial for predicting the spatial organization of animals, for instance to plan animal species reinsertion programs or to steer vaccine drop campaigns to counter disease spread. With the development of tools associated with geographic information systems, substantial information on animal movement has become available and statistical methods to extract this information must be fine-tuned. In this talk we compare two statistical approaches for inference on the determinants of animal movement, one based on conditional logistic regression and the other based on regression models for angular data.

Une bonne compréhension des déterminants du déplacement animal est incontournable pour prédire l'organisation spatiale des animaux, par exemple pour planifier les programmes de réintroduction d'espèces animales ou pour orienter les épandages vaccinaux visant à contrer la propagation d'une maladie. Avec le développement d'outils liés aux systèmes d'information géographique, beaucoup d'information sur le déplacement animal est disponible, et les méthodes statistiques pour l'extraire doivent être mises au point. Nous comparons dans cet exposé deux approches statistiques pour inférer sur les déterminants du déplacement animal, soit une basée sur la régression logistique conditionnelle et l'autre sur la régression pour données angulaires.

1D-15: Stochastic Processes Arising in Environmental Sciences Processus stochastiques dans les sciences environnementales

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: ECHA L1-430

Abstracts/Résumés

DON DAWSON (Carleton University)

[Monday May 27 / lundi 27 mai, 15:30]

Some Probabilistic Objects Motivated by Evolutionary Biology and Ecology

Quelques objets probabilistes motivés par la biologie évolutive et l'écologie

We investigate the roles of rare mutations and spatial migration in the emergence and spatial distribution of new types in evolutionary theory in the framework of a class of spatially structured stochastic population models which incorporate the effects of migration, selection and mutation. The objective is to investigate the evolution of the system in a hierarchy of space and time scales and to describe the emergence and spatial distribution of selectively advantageous mutants. The basic tool is a class of set-valued processes which provides a dual representation of the system of interacting Fleming-Viot processes. (Joint work with Andreas Greven.)

Nous explorons les rôles des mutations rares et des migrations spatiales sur l'émergence et la distribution spatiale de nouveaux types dans la théorie de l'évolution sous le cadre d'une classe de modèles de population aléatoire structurée dans l'espace qui englobent les effets de migration, sélection et mutation. L'objectif est d'évaluer l'évolution du système dans une hiérarchie d'espaces et d'échelles de temps et de décrire l'émergence et la distribution spatiale des mutants sélectivement avantageux. L'outil de base est une classe de processus à valeur fixe qui fournit une double représentation du système de processus interagissants de Fleming-Viot. (Travail réalisé conjointement avec Andreas Greven.)

GAIL IVANOFF (University of Ottawa)

[Monday May 27 / lundi 27 mai, 16:00]

Renewal Processes in Two Dimensions

Processus de renouvellement en deux dimensions

We define a renewal property for point processes in two dimensions and show that the law of the renewal process is determined by a so-called avoidance probability function. The avoidance probability is the two-dimensional analogue of the survival function of the interarrival distribution of a renewal process in one dimension. We introduce nonparametric methods for estimation of the avoidance probability. We see that martingale methods yield a unified approach for renewal processes in both one and two dimensions, and can be used

Nous définissons une propriété de renouvellement des processus ponctuels en deux dimensions et montrons que la loi du processus de renouvellement est déterminée par la fonction de probabilité d'évitement. La probabilité d'évitement est analogue en deux dimensions à la fonction de survie de la distribution inter arrivée d'un processus de renouvellement en une dimension. Nous introduisons les méthodes non-paramétriques pour estimer la probabilité d'évitement. Nous voyons que les méthodes de martingales entraînent une approche unifiée pour les processus de renouvellement en une et deux dimensions et qu'elles peuvent être utilisés

1D-I5: Stochastic Processes Arising in Environmental Sciences Processus stochastiques dans les sciences environnementales

for both synchronous and asynchronous data. This talk is based partially on joint work with Katherine Davies.

avec des données synchrones et asynchrones. Cet exposé se base en partie sur un travail effectué conjointement avec Katherine Davies.

REG KULPERGER & IRENA CREED (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 16:30]

Trend Analysis of Water Yields in Northern Catchments

Analyse de la tendance des apports d'eau dans les bassins versants nordiques

Determining catchment responses to climate signals gives insight into the potential effects of climate change. This talk describes methods used to study the water yield over time from four headwater catchments in the Turkey Lakes Watershed in northern Ontario. The relatively simple statistical ideas of regression are used to remove signals thereby yielding some of the more interesting and subtler features which are ecologically interesting. This first analysis is based on annual data. Daily data are needed to study shorter lag effects, in terms of days or weeks.

Déterminer les réponses des bassins versants aux signaux climatiques nous fournit des indications des effets potentiels du changement climatique. Cet exposé décrit les méthodes employées pour étudier l'apport d'eau au fil du temps en amont de bassins versants dans les lignes de partage des eaux des lacs Turkey dans le nord de l'Ontario. Les idées statistiques relativement simples de la régression sont utilisées pour supprimer les signaux et ainsi obtenir quelques-unes des caractéristiques les plus intéressantes et subtiles qui sont d'intérêt écologique. La première analyse se base sur des données annuelles. Des données journalières sont requises pour étudier les effets de plus courts décalages selon les jours ou les semaines.

1D-C1: Analysis of Survival and Clinical Trials Data Analyse de données de survie et d'essais cliniques

Chair/Président: Richard Cook (University of Waterloo)

Room/Salle: EdSouth 228

Abstracts/Résumés

JEFF BAKAL, YINNGAN ZHENG, CYNTHIA M. WESTERHOUT & PAUL W. ARMSTRONG (University of Alberta)

[Monday May 27 / lundi 27 mai, 15:30]

Refining the Analysis of Composite Endpoints in Clinical Trials

Raffinage de l'analyse des critères d'évaluation des essais cliniques

Controversy exists in the interpretation of clinical trials using composite endpoints. These trials may combine several endpoints of varying severity to derive a single measure of the difference in efficacy between treatments. The interpretation of trial information becomes challenging when the treatment effect varies in magnitude and direction across the selected endpoints and when there are multiple endpoints per patient. Recently several novel methods have been developed to improve the analysis of these trials to better utilize the information and reflect the patient experience. We explore several of these methods in the context of a series of recent clinical trials.

Des controverses existent dans l'interprétation des essais cliniques utilisant des critères composites. Ces essais peuvent combiner plusieurs paramètres de sévérité pour obtenir une mesure de la différence d'efficacité entre les traitements. L'interprétation des essais devient difficile quand l'effet thérapeutique varie en grandeur et direction entre les paramètres choisis et lorsqu'il y a plusieurs résultats par patient. Récemment, plusieurs méthodes nouvelles ont été développées pour améliorer l'analyse de ces essais afin de mieux utiliser les informations et tracer l'expérience du patient. Nous explorons plusieurs de ces méthodes dans le cadre d'une série d'essais cliniques récents.

PIERRE-JÉRÔME BERGERON (University of Ottawa), **VAHID PARTOVI NIA** (École Polytechnique de Montréal), **JAIME YOUNGER** (Institute for Clinical Evaluative Sciences), **EWA MAKVANDI** (University of Ottawa)

[Monday May 27 / lundi 27 mai, 15:45]

Analyzing Length-Biased Survival Data Using the R Package lbiassurv

L'analyse de données censurées à droite avec biais de longueur en R avec le package lbiassurv

A how-to guide for using the R package lbiassurv, for parametric and nonparametric analysis of length-biased right-censored data. It includes simulation functions, nonparametric maximum likelihood estimator (NPMLE) of the bias-corrected survival curve using EM algorithm by Vardi with bootstrap confidence bands, parametric analysis using a variety of standard survival models (Weibull, log-normal, log-logistic and more), one-sample goodness-of-fit test, some graphical methods and tests for stationarity of incidence for left-truncated data.

Un guide d'utilisation du package lbiassurv en R, pour l'analyse paramétrique et nonparamétrique de données à biais de longueur censurées à droite. Sont présentés : les fonctions de simulation, l'estimation nonparamétrique de maximum de vraisemblance (ENPMV) de la courbe de survie corrigée pour le biais avec l'algorithme EM de Vardi et bandes de confiance par la méthode de Cyrano, estimation par maximum de vraisemblance à l'aide de modèles communs (Weibull, log-normal, log-logistique et autres), test de validité d'ajustement, quelques méthodes graphiques ainsi que des tests pour le biais de longueur des données tronquées à gauche.

1D-C1: Analysis of Survival and Clinical Trials Data Analyse de données de survie et d'essais cliniques

KEYUE DING (Queen's University), **ZHOUSHANYUE HE** (University of Waterloo), **JIANFENG XU** (Queen's University)

[Monday May 27 / lundi 27 mai, 16:00]

Discovery of Pre-Post-Treatment Predictive Biomarker

Découverte de bio-marqueurs de prédiction avant et après traitement

Serum samples were collected in some cancer trials. Based on a drug's mechanism of action, only patients with over expressed level, or significantly changed in certain markers after treatment could benefit from the study drug. Estimation of treatment effect as a function of biomarker is derived using the local partial likelihood method. The cut-off point of the marker's level or levels of change after treatment was determined according to certain pre-specified threshold effects using Cross-validation methods. We applied the method to a NCIC clinical trials group trial and identified that certain interested serum markers could predict differential treatment effect.

Des échantillons de sérum ont été recueillis dans certains essais sur le cancer. En fonction du mécanisme d'action des médicaments, seulement les patients avec un niveau excédentaire, ou avec un niveau considérablement modifié dans certains marqueurs après le traitement pouvaient bénéficier de ce médicament en étude. L'estimation de l'effet de traitement comme fonction du bio-marqueur est obtenue avec une méthode de vraisemblance partielle locale. Le point critique du niveau ou des niveaux de changement du marqueur après le traitement a été déterminé en fonction de certains effets de seuil avec des méthodes de validation croisée. Nous avons appliqué la méthode à des essais de groupe cliniques de l'Institut national du cancer du Canada et avons déterminé que certains marqueurs de sérum intéressants pouvaient prédire l'effet de traitement différentiel.

ADEL ELGHAFGHUF & HENRIK STRYHN (University of Prince Edward Island)

[Monday May 27 / lundi 27 mai, 16:15]

A Cross-Classification and Multiple Membership Model Applied to Time-to-Event Data

Un modèle de classification croisée et d'appartenance multiple appliqué à des données de temps avant événement

A cross-classified and multiple membership frailty model was applied to calf mortality data from western Canada, where 23480 calves from 175 farms were followed for 180 days after calving. The farms were cross-classified by 49 veterinary clinics and 9 ecological regions and in a multiple member relation to the veterinary clinics resulting in 3-level cross-classified and multiple membership data structures. The analysis demonstrated a greater variation among farms than among both ecological regions and veterinary clinics. Further, a small simulation study for a setting similar to the real data gave evidence that the used approach provides valid estimates.

Un modèle de fragilité de classification croisée et d'appartenance multiple a été appliqué à des données sur la mortalité des veaux provenant de l'Ouest canadien : un suivi de 23 480 veaux provenant de 175 fermes a été fait pendant 180 jours après le vêlage. Les fermes ont été classifiées en croisée par 49 cliniques vétérinaires et 9 régions écologiques et dans de nombreuses relations d'appartenance aux cliniques vétérinaires. Ceci crée une structure de données à classification croisée et d'appartenance multiple à 3 niveaux. L'analyse démontre une plus grande variation entre les fermes que parmi les régions écologiques et les cliniques vétérinaires. De plus, une petite étude de simulation ayant des caractéristiques similaires aux données réelles a mis en évidence que l'approche utilisée offre des estimations valides.

GABRIEL ESCARELA (Universidad Autónoma Metropolitana - Iztapalapa)

[Monday May 27 / lundi 27 mai, 16:30]

Regression Analysis of Competing Risks with Missing Covariate Data

Analyse de régression pour les risques concurrents avec des données manquantes pour covariables

1D-C1: Analysis of Survival and Clinical Trials Data Analyse de données de survie et d'essais cliniques

A likelihood-based technique for the regression analysis of right-censored competing risks data when two of the covariates are partially missing was developed. The approach envisaged here comprises two models: one describes the covariate effects on both long-term incidence and conditional latencies for each cause of death, whilst the other deals with the observation process by which the covariates are missing. The former is formulated with a well-established mixture model and the latter is characterized by a probability function which is constructed using both a copula model and multinomial margins. The two methods are illustrated on a prostate cancer cohort.

On a développé une technique basée sur la vraisemblance pour l'analyse des données par la régression pour des risques antagonistes censurées à droite lorsque les données sont partiellement manquantes pour deux des covariables. L'approche envisagée ici comprend deux modèles : l'un décrit les effets de covariable sur l'incidence à long terme et sur les temps d'arrêt conditionnels pour chaque cause de mortalité, alors que l'autre traite du processus d'observation par lequel les covariables sont manquantes. Le premier est formulé avec un modèle de mélange bien établi et le deuxième se caractérise par une fonction de probabilité construite à l'aide d'un modèle de copules et de marges multinomiales. Ces méthodes sont illustrées avec une cohorte de cancer de la prostate.

MICHAEL MCISAAC & RICHARD COOK (University of Waterloo), **MELANIE POULIN-COSTELLO** (Amgen Inc)
[Monday May 27 / lundi 27 mai, 16:45]

Misspecified Models in Analyses of Incomplete Clinical Trial Data

Modèles mal spécifiés dans l'analyse de données incomplètes provenant d'essais cliniques

Participants in clinical trials often withdraw before completion of follow-up, leading to incomplete data. Individuals who remain in the study may no longer form groups with similar distributions of demographic features, which compromises the validity of causal inferences. The purpose of this talk is to discuss mechanisms by which incomplete data can arise and methods which can be employed to minimise their effect. The impact of model misspecification in methods for dealing with incomplete data is also studied based on the limiting behaviour of resulting estimators. These considerations are illustrated using data from clinical trials in dermatology.

Il arrive souvent que les participants se retirent des essais cliniques avant la fin de la période de suivi, ce qui engendre des données incomplètes. Les personnes qui restent dans l'étude risquent de ne plus former de groupes avec les mêmes distributions de caractéristiques démographiques, ce qui compromet la validité des inférences causales. L'objectif de cet article est de discuter des mécanismes par lesquels des données incomplètes peuvent survenir et des méthodes à employer pour minimiser leur effet. On étudie également les conséquences de l'erreur de spécification des méthodes pour traiter des données incomplètes en fonction du comportement limite des estimateurs résultants. Les considérations sont illustrées par des données provenant d'essais cliniques en dermatologie.

1D-C2: Biostatistics - Applications and Case Studies Biostatistique - applications et études de cas

Chair/Présidente: Wendy Lou (University of Toronto)

Room/Salle: EdSouth 265

Abstracts/Résumés

TANJA HOEGG, SYLVIA ESTERBY & PARAMJIT GILL (University of British Columbia - Okanagan), **CYNTHIA ARAUJO & RASIKA RAJAPAKSHE** (BC Cancer Agency)

[Monday May 27 / lundi 27 mai, 15:30]

Breast Cancer in British Columbia - Identification of High Risk Women Based on Breast Cancer Risk Modelling
Cancer du sein en Colombie-Britannique - identification des femmes à haut risque par modélisation du risque de cancer du sein

Although there are many known factors associated with an increased risk of breast cancer, currently, age remains the only eligibility criterion for the Screening Mammography Program of BC. Using five year follow-up data of 220,000 SMP BC participants, we aim to identify women at high risk of breast cancer and create a personalized access criterion to earlier breast screening. We evaluated the discriminatory performance of the existing Gail et al. breast cancer risk model for the BC population. Accounting for well-established risk factors, we estimate survival probabilities as a basis for improved outcome stratification of the population.

Bien que de nombreux facteurs contribuent à augmenter le risque de cancer du sein, l'âge est aujourd'hui encore le seul critère d'éligibilité au Programme de dépistage de cancer du sein en Colombie-Britannique (SMPBC). Sur la base de données de suivi sur cinq ans concernant 220 000 participantes au SMPBC, nous cherchons à identifier les femmes qui présentent un risque élevé de cancer du sein et à créer un critère d'accès personnalisé au dépistage précoce. Nous évaluons l'efficacité de la discrimination du modèle existant actuel de Gail et al. pour la population de la province. En tenant compte de facteurs de risque bien établis, nous estimons les probabilités de survie afin d'améliorer la stratification des résultats de la population.

MOHAMMAD EHSANUL KARIM, PAUL GUSTAFSON, A. JOHN PETKAU, AFSANEH SHIRANI, YINSHAN ZHAO, ELAINE KINGWELL & HELEN TREMLETT (University of British Columbia)

[Monday May 27 / lundi 27 mai, 15:45]

A Simulation Study of Methods Used to Reduce Variability in the Inverse Probability of Treatment Weights for Marginal Structural Cox Models

Étude par simulation des méthodes utilisées pour réduire la variabilité de la probabilité inverse des pondérations de traitement dans les modèles de Cox structurels marginaux

Extreme values among the inverse probability of treatment weights can unduly affect the results of marginal structural Cox models. Besides standardization, methods including truncation, trimming and normalization are suggested in the literature to reduce the variability of these weights. To compare the performance of these methods, we used simulated data that mimicked survival data in which both treatment and confounders were time-dependent. The findings from this simulation study

Les valeurs extrêmes dans la probabilité inverse des pondérations de traitement peuvent affecter outre mesure les résultats des modèles de Cox structurels marginaux. Outre la standardisation, la littérature suggère diverses méthodes comme la troncature, l'ajustement et la normalisation pour réduire la variabilité de ces pondérations. Pour comparer la performance de ces méthodes, nous avons étudié des données simulées qui imitent des données de survie dans lesquelles le traitement et les variables confusionnelles dépendent du temps. Nous avons utilisé les résultats de cette

1D-C2: Biostatistics - Applications and Case Studies Biostatistique - applications et études de cas

were used to guide the weighted adjustment in an application where we investigated the impact of beta-interferon treatment in delaying disability progression in patients from the British Columbia Multiple Sclerosis database (1995-2008).

étude par simulation pour la correction pondérée dans une application étudiant l'impact d'un traitement par l'interféron bêta sur la progression du handicap chez les patients de la base de données sur la sclérose en plaques de la Colombie-Britannique (1995-2008).

PROSANTA MONDAL, LILIAN THORPE & HYUN LIM (University of Saskatchewan)

[Monday May 27 / lundi 27 mai, 16:00]

Predictors for Nursing Home Placement (NHP) and Mortality Among Intellectually Disabled Adults With and Without Down Syndrome (DS): Application of Competing Risks Models

Prédicteurs du placement en maison de soins (PMS) et de la mortalité chez les adultes présentant des déficiences intellectuelles avec et sans syndrome de Down (SD) : application de modèles de risques concurrents

Objective: To identify predictors for NHP and mortality. Methods: Cox cause-specific and sub-distribution hazards models were used. Results: Among 343 participants, 14.6% had NHP, 21% died without NHP. Interaction of DS and age, higher physical-depression had a higher hazard of NHP in cause-specific hazards model, only interaction of DS and age was significant in sub-distribution hazards model of NHP. Older age, higher physical-depression had a higher hazard of death in both cause-specific and sub-distribution hazards models. Conclusion: Sub-distribution and cause-specific hazards models gave same results for death not for NHP. Simulation will be performed to investigate applicability of the models.

Objectif : Identifier les prédicteurs du PMS et de la mortalité. Méthodes : Modèles de risque de Cox spécifiques à la cause et à sous-distribution. Résultats : Sur 343 participants, 14,6 % ont été placés en maison de soins, 21 % sont décédés sans avoir été placés en maison de santé. L'interaction entre le SD et l'âge, une plus grande dépression physique produisent un risque plus élevé de PMS dans les modèles de risques spécifiques à la cause, seule l'interaction entre le SD et l'âge est significatif dans le modèle de risques de PMS à sous-distribution. L'âge élevé, une plus grande dépression physique produisent un risque plus élevé de décès dans les deux modèles de risques. Conclusion : Les modèles de risque sous-distribution et spécifiques à la cause produisent les mêmes résultats concernant le décès, mais non le PMS. Nous effectuerons des simulations pour étudier l'applicabilité des modèles.

YAWEN XU & SICHENG WU (York University), **ZENY FENG & XIAOJIAN YANG** (University of Guelph), **XIAOGANG WANG & XIN GAO** (York University)

[Monday May 27 / lundi 27 mai, 16:15]

Multiple-Platform Data Integration Method with Application to Combined Analysis of Microarray and Proteomic Data
Méthode d'intégration de données de différentes plateformes et application à l'analyse combinée de données provenant de biopuces et protéomiques

It's desirable in genomic studies to select biomarkers that differentiate between normal and diseased populations based on related data sets from different platforms. Most recently developed integration methods focus on correlation analyses between gene and protein expression profiles. These methods select biomarkers with concordant behavior but do not directly select differentially expressed biomarkers. Other methods combine statistical evidence in terms of ranks and p-values,

Il est souhaitable, dans les études génomiques, de sélectionner des biomarqueurs qui distinguent entre les populations normales et malades en fonction d'ensembles de données apparentés tirés de plateformes distinctes. La plupart des méthodes d'intégration récentes mettent l'accent sur les analyses de corrélation entre les profils d'expression des gènes et des protéines. Elles sélectionnent des biomarqueurs dont le comportement concorde, mais ne sélectionnent pas directement des biomarqueurs exprimés différemment. D'autres méthodes combinent les preuves sta-

1D-C2: Biostatistics - Applications and Case Studies Biostatistique - applications et études de cas

but they don't account for the dependency relationships among the data across platforms. We propose an integration method to perform hypothesis testing and biomarkers selection based on multi-platform data sets observed from normal and diseased populations.

tistiques concernant le rang et les valeurs p , mais elles ne tiennent pas compte des relations de dépendance entre les données sur différentes plateformes. Nous proposons une méthode d'intégration pour tester des hypothèses et sélectionner des biomarqueurs en fonction d'ensembles de données multi-plateformes observées dans des populations normales et malades.

XUECHEN ZHANG (University of Alberta)

[Monday May 27 / lundi 27 mai, 16:30]

Power Comparison of Discrete Data Methods with Repeated Measurement Data in Small Sample

Comparaison de la puissance de méthodes basées sur des mesures répétées discrètes avec des échantillons de petite taille

Various analytical methods are available to analyze repeated measurement data for both continuous and discrete data. In the case of discrete data, most methods are based on asymptotic normality, requiring large sample size. Naturally, their small sample performance may not match the expectation satisfactorily. Two main methods, non-linear mixed effect model and generalized estimating equation method, are investigated for their small sample performance on repeated binary responses. Repeated binary data sets are generated with sample size of 20, 40, 60, 80 and 100. Two methods are applied to each data set and power is compared, along with their coverage probabilities.

Diverses méthodes d'analyse sont disponibles pour analyser les données de mesure répétées à la fois continues et discrètes. Dans le cas de données discrètes, la plupart des méthodes sont basées sur la normalité asymptotique, nécessitant des échantillons de grande taille. Naturellement, avec des échantillons de petite taille, la performance risque de ne pas répondre aux attentes de façon satisfaisante. Nous étudions la performance de deux méthodes principales pour des réponses binaires répétées provenant d'échantillons de petite taille, à savoir le modèle non linéaire à effets mixtes et la méthode des équations d'estimation généralisées. Des ensembles de données binaires répétées sont générées avec des tailles d'échantillon de 20, 40, 60, 80 et 100. Les deux méthodes sont appliquées à chaque jeu de données, leur puissance et leurs probabilités de couverture sont comparées.

NICHOLAS MITSAKAKIS (Toronto Health Economics and Technology Assessment Collaborative), **KAREN BREMNER** (University Health Network, Toronto), **MURRAY KRAHN** (Toronto Health Economics and Technology Assessment Collaborative)

[Monday May 27 / lundi 27 mai, 16:45]

Methodological Challenges in Mapping a Disease Specific Psychometric Instrument to a Disease Specific Utility Instrument

Défis méthodologiques de la cartographie d'un instrument psychométrique spécifique à une maladie en un instrument utilitaire spécifique à une maladie

Regression techniques used to map disease-specific psychometric instruments to disease specific utility instruments have been proposed for the estimation of utility scores when data on the latter are not available. Here, using data from prostate cancer patients, we investigate potential improvement of the mapping when the utility scores are

Des techniques de régression utilisées pour cartographier les instruments psychométriques spécifiques à une maladie en instruments utilitaires spécifiques à une maladie ont été proposées pour l'estimation des scores d'utilité lorsque les données sur les instruments utilitaires ne sont pas disponibles. À l'aide de données de patients atteints du cancer de la prostate, nous étudions l'amélioration potentielle de

1D-C2: Biostatistics - Applications and Case Studies

Biostatistique - applications et études de cas

transformed prior to fitting the regression model. We also discuss the hypothesis that the performance of the mapping models is affected by the low correlation between the disease specific and non-specific items of the preference-based instrument and we present a couple of approaches that could be used for investigating this hypothesis.

la cartographie quand les scores d'utilité sont transformés avant d'ajuster le modèle de régression. Nous discutons également de l'hypothèse que la performance des modèles de cartographie est affectée par la faible corrélation entre les éléments spécifiques et non spécifiques à la maladie de l'instrument fondé sur les préférences, et nous présentons quelques approches qui pourraient être utilisées pour étudier cette hypothèse.

1D-C3: Dependence I: Correlation and Copulas Dépendance I : corrélation et copules

Chair/Président: Bruno Rémillard (HEC Montréal)

Room/Salle: EdSouth 221

Abstracts/Résumés

AYI AJAVON & FRANÇOIS PERRON (Université de Montréal)

[Monday May 27 / lundi 27 mai, 15:30]

Copulas Extensions for Non Continuous Random Variables

Extensions continues pour les sous copules des variables aléatoires discrètes

The copula of a non continuous random vector (Y_1, \dots, Y_d) is uniquely determined only on the range $R_1 \times \dots \times R_d$, where R_i is the range of $F_i(Y_i)$, F_i is the distribution function of Y_i , $i = 1, \dots, d$. It admits extensions on $[0, 1]^d$. The most used are the “starred copula” and the “standard extension copula” of Schweizer. It is shown that they are equivalent in dimension $d = 2$ but not yet in dimension $d > 2$. We prove that these copulas are the same in dimension $d \geq 2$. In addition, we provide the minimal and maximal extensions copulas in dimension $d > 2$.

La copule d'un vecteur aléatoire discret (Y_1, \dots, Y_d) est définie de façon unique sur un sous-ensemble discret de $[0, 1]^d$. Elle admet des extensions sur $[0, 1]^d$. Les plus utilisées sont la copule “étoilée” et l’ “extension standard” de Schweizer. Il est démontré qu’elles sont équivalentes en dimension $d = 2$ mais pas encore en dimension $d > 2$. Nous montrons que ces copules sont les mêmes en dimension $d > 2$. En outre, nous fournissons les copules minimales et maximales des extensions en dimension $d > 2$.

NIROSHAN WITHANAGE & ALEXANDER R. DE LEON (University of Calgary)

[Monday May 27 / lundi 27 mai, 15:45]

Joint Modeling of Hierarchically Clustered Non-Gaussian Continuous Outcomes via the Gaussian Copula

Modélisation conjointe de résultats continus non gaussiens en grappes hiérarchisées au moyen d'une copule gaussienne

In many biomedical studies, researchers simultaneously record multiple outcomes. With non-Gaussian continuous responses with disparate distributions, joint analysis is rendered complicated by the lack of multivariate analogue to the multivariate normal distribution. We propose a copula-based approach to joint modeling of non-Gaussian clustered continuous outcomes. We use Gaussian copula to glue multiple outcomes and adopt GLMMs for the marginal models of the outcomes. The approach does not assume conditional independence of outcomes unlike in conventional GLMMs. It is thus able to capture all relevant associations in the data. The proposed methodology is illustrated using data from a comet assay.

Dans plusieurs études biomédicales, les chercheurs enregistrent simultanément de multiples résultats. Pour des réponses continues non gaussiennes aux distributions disparates, l'analyse conjointe est compliquée par le manque d'alternatives à la loi normale multivariée. Nous proposons une approche basée sur les copules pour la modélisation conjointe de résultats continus non gaussiens en grappes. Nous utilisons une copule gaussienne pour regrouper des résultats multiples et nous adoptons des modèles linéaires mixtes généralisés (GLMM) comme modèles marginaux des résultats. Contrairement aux GLMM classiques, notre approche ne suppose pas l'indépendance conditionnelle des résultats. Elle permet ainsi de saisir tous les liens pertinents dans les données. La méthodologie proposée est illustrée au moyen d'un test des comètes.

1D-C3: Dependence I: Correlation and Copulas Dépendance I : corrélation et copules

YANLING CAI, JAMES V. ZIDEK & WILLIAM WELCH (University of British Columbia)

[Monday May 27 / lundi 27 mai, 16:00]

Breaking the Same Board Twice: the Magic of Statistics!

Casser la même planche deux fois: la magie de la statistique!

Stresses affect strength properties of a lumber specimen simultaneously. A strong stochastic dependence between these properties would diminish a wooden structure's reliability. Assessing strength relationships is a seemingly insurmountable challenge as every one of these strengths needs a destructive testing to measure. Here is where statistics comes to the rescue. This paper describes a large-scale experiment designed and conducted by statistics graduate students in a commercial testing lab. It proposes the new theory to infer the stochastic dependence, which applies proof load technique with a penalizing likelihood approach motivated by a Bayesian conjugate prior. The result: asymptotically consistent, more efficient.

Les tensions appliquées à une pièce de bois affectent l'ensemble de ses propriétés de résistance simultanément. Une forte dépendance stochastique entre ces propriétés diminuerait la fiabilité de la structure en bois. L'évaluation de la relation entre ces propriétés de résistance semble être un défi insurmontable, puisque la mesure de chacune d'elle nécessite un essai destructif. C'est ici que la statistique vient à la rescousse. Cet article décrit une expérience à grande échelle conçue et menée par des étudiants en statistique des cycles supérieurs dans un laboratoire d'essais commercial. Une nouvelle théorie est proposée pour inférer la dépendance stochastique qui fait appel à la technique de charge d'essai avec une approche de vraisemblance pénalisée inspirée de la loi a priori conjuguée dans un contexte bayésien. Résultat : convergence asymptotique et plus grande efficacité.

ZIHANG LU (University of Toronto), **JENNA SYKES & MELANIA PINTILIE** (Ontario Cancer Institute, Princess Margaret Hospital)

[Monday May 27 / lundi 27 mai, 16:15]

The Bias for Sub-Hazard Ratio when the Event Times are Correlated

Biais pour le rapport de sous-risques lorsque les temps d'événements sont corrélés

In the analysis of competing risk data, the effect of a covariate is modeled using the Fine and Gray approach, yielding a sub-distribution hazard ratio. The sub-hazard ratio provided by the Cox model cannot be interpreted as the degree of correlation between the two types of events is unknown. Our study investigates to what degree the sub-hazard ratio approximates the marginal hazard ratio for varying degrees of correlation. Using simulation, it was found that the bias was large for high correlations (≥ 0.5), depends on the direction of correlation and can overestimate as well as underestimate the marginal hazard ratio.

Dans l'analyse de données à risques concurrents, l'effet d'une covariable est modélisé selon la méthode Fine et Gray, donnant lieu à un rapport de risques pour la distribution sous-jacente. Le rapport des risques sous-jacents produit par le modèle de Cox ne peut être établi, puisque le degré de corrélation entre deux types d'événements est inconnu. Nous examinons dans quelle mesure le rapport de risques sous-jacents approxime le rapport de risques marginaux pour différents degrés de corrélation. Une simulation a montré un grand biais pour les fortes corrélations (≥ 0.5) qui dépend de l'orientation de la corrélation et qui peut donner lieu à un rapport de risques marginaux surestimé ou sous-estimé.

1D-C3: Dependence I: Correlation and Copulas Dépendance I : corrélation et copules

CHAFIK BOUHADDIoui (UAE University)

[Monday May 27 / lundi 27 mai, 16:30]

Statistical Inference in Infinite-Order Cointegrated Vector Autoregressive Processes under Uncorrelated but Dependent Errors

Estimation des paramètres d'un processus autorégressif cointégré sous des erreurs non-corrélées mais dépendantes

The concept of co-integration processes is one of the most used concepts in economics and finance. Mainly, researchers are interested in the behavior of the estimators of the model parameters. In this paper, we will investigate the asymptotic behavior of the estimators of an infinite-order co-integrated vector autoregressive series under nonindependent errors by showing its asymptotic distribution. Using this result, we will construct a Likelihood Ratio (LR) test of the co-integration rank. One can also develop a method under unrestrictive assumptions to select the autoregressive order. Monte Carlo experiments illustrate the finite sample performance of the LR test.

Le concept des processus autorégressifs cointégrés est très utilisé en finance et en économie. Généralement, les chercheurs sont intéressés par l'estimation des paramètres du modèle sous des conditions plus faibles. Dans ce papier, on étudiera le comportement asymptotique des estimateurs des paramètres d'un modèle autorégressif cointégré d'ordre infini sous des erreurs non-corrélées mais dépendantes. On construira ensuite le test de rapport de vraisemblance pour tester le rang de cointégration. Une méthode de choix de l'ordre autorégressif sera aussi présentée. Finalement, une étude par simulation nous permettra de vérifier les performances du test de rapport de vraisemblance sous un échantillon fini.

1D-C4: Estimation Methods Méthodes d'estimation

Chair/Président: Jose Garrido (Concordia University)

Room/Salle: EdSouth 177

Abstracts/Résumés

GOLSHID CHATRCHI (Carleton University)

[Monday May 27 / lundi 27 mai, 15:30]

Robust Estimation of Variance Components in Small Area Estimation

Estimation robuste des composantes de la variance dans le contexte de l'estimation sur petits domaines

Modern techniques for Small Area Estimation (SAE) rely on explicit modeling assumptions. These methods can be highly affected by the occurrence of outliers. Robust SAE methods have been developed to remedy this problem. This study considers estimators based on Robustified Maximum Likelihood (RML) equations (Sinha and Rao, 2009) and mainly concentrates on the estimation of variance components. An alternative iterative algorithm, fixed-point method, is proposed and Henderson's method III estimator is also suggested for the initial values of the algorithm. Moreover, the RML method is extended to the two-fold nested error regression model with equal error variances.

Les techniques modernes pour l'estimation sur petits domaines (EPD) reposent sur des hypothèses de modélisation explicites. Ces méthodes peuvent être très affectées par des valeurs aberrantes. Des méthodes robustes pour EPD ont été développées pour remédier à ce problème. Cette étude considère les estimateurs basés sur des équations de maximum de vraisemblance robustifiées (MVR) et se concentre sur l'estimation des composantes de la variance. Un algorithme itératif, point fixe, est proposé et l'estimateur d'Henderson III est utilisé pour les valeurs initiales. En outre, la méthode MVR est étendue à un modèle de régression doublement emboîté avec des variances d'erreur égales.

CHRISTOPHER A. GRAVEL (Carleton University), **PALASH GHOSH & ANUP DEWANJI** (Indian Statistical Institute), **PATRICK FARRELL** (Carleton University), **DANIEL KREWSKI** (University of Ottawa)

[Monday May 27 / lundi 27 mai, 15:45]

Analysis of Electronic Health Records for Active Pharmacovigilance with Misclassification of the Study Variable

Analyse de dossiers électroniques en pharmacovigilance avec erreur de classification pour la variable à l'étude

In Pharmacovigilance, electronic patient health records (EHRs) are used to estimate the risk of adverse drug reactions (ADRs). These datasets can contain misclassification errors in the study variable, due to misdiagnosis for example, shifting the estimates. Using maximum likelihood estimation under a multi-sample framework, we obtain corrected estimates of the ADR risk by incorporating additional information from internal validation samples. For the simple case of the odds ratio we derive closed form expressions for the adjusted estimators under differing assumptions. More complicated modelling using numerical methods will also be discussed.

La pharmacovigilance utilise les dossiers médicaux électroniques des patients afin d'estimer leur risque de subir un effet indésirable lié à un médicament. Ces bases de données peuvent contenir des erreurs de classification, ce qui peut altérer la valeur des estimés. En utilisant l'estimation du maximum de vraisemblance, on obtient des estimés corrigés des effets indésirables en incorporant des informations supplémentaires provenant d'échantillons de validation interne. Pour les rapports de chance, on dérive des solutions en forme fermée pour les estimateurs ajustés sous différentes suppositions. Des modélisations plus complexes utilisant les méthodes numériques sont aussi discutées.

1D-C4: Estimation Methods Méthodes d'estimation

CAREN HASLER & YVES TILLÉ (Université de Neuchâtel)

[Monday May 27 / lundi 27 mai, 16:00]

Balanced k -nearest Neighbor Imputation
Imputation équilibrée par les k plus proches voisins

In order to overcome the problem of item nonresponse, random imputations are often used because they tend to preserve the distribution of the imputed variable. Among the methods of random imputation, the random hot-deck has the interesting property that the imputed values are observed values. We present a new random method of hot-deck imputation which enables us to select the imputed values such that some balancing equations are satisfied and such that the donors are selected in a neighborhood of the recipients.

Des méthodes d'imputation aléatoire sont souvent utilisées pour traiter la non-réponse car de telles méthodes ont tendance à préserver la distribution des variables que l'on impute. Parmi les méthodes d'imputation aléatoire, on trouve les méthodes dites hot-deck ; celles-ci jouissent de l'intéressante propriété d'imputer des valeurs observées. Nous proposons une nouvelle méthode d'imputation hot-deck aléatoire permettant de sélectionner les valeurs imputées de telle manière que des équations d'équilibrage soient satisfaites et que les donateurs potentiels pour un receveur appartiennent à un voisinage de celui-ci.

FENG HE & GRACE YI (University of Waterloo)

[Monday May 27 / lundi 27 mai, 16:15]

Analysis of Progressive Multi-State Models with Misclassified States
Analyse de modèles multi-états progressifs avec états mal classés

Multi-state models are commonly used in studies of disease progression. Methods developed under this framework, however, are often challenged by misclassification in states. In this talk, we investigate issues concerning continuous time progressive multi-state models with misclassification in states. We develop inference methods that are based on joint modeling the progressive and misclassification processes. Moreover, we incorporate the mover-stayer feature in the progressive model to facilitate a simple type of heterogeneity where the underlying population consists of two subpopulations of movers and stayers. Estimation procedures are developed using both the likelihood and pairwise likelihood methods. Numerical studies will be presented.

Les modèles multi-états sont utilisés couramment dans les études de progression de maladie. Cependant, les méthodes développées dans ce cadre sont souvent mises à l'épreuve par une mauvaise classification des états. Dans cet exposé, nous étudions des problèmes liés aux modèles multi-états progressifs en temps continu comportant des erreurs de classification des états. Nous élaborons des méthodes d'inférence basées sur la modélisation conjointe des processus de progression et de mauvaise classification. Par ailleurs, nous intégrons la caractéristique mobile-stable dans le modèle progressif afin de faciliter un type d'hétérogénéité où la population sous-jacente est constituée de deux sous-populations d'individus mobiles et stables. Les procédures d'estimation sont élaborées au moyen des méthodes de vraisemblance et de vraisemblance par paires. Nous présentons aussi des études numériques.

HUIJING WANG & X. JOAN HU (Simon Fraser University), **MARY MCBRIDE & MARIA LORENZI** (British Columbia Cancer Agency)

[Monday May 27 / lundi 27 mai, 16:30]

Analysis of Counts with Two Latent Classes
Analyse de dénombrements avec deux classes latentes

Motivated partly by the CAYACS program at BC Cancer Agency, we consider a latent class model

Inspirés notamment par le programme CAYACS de la BC Cancer Agency, nous étudions un modèle de classes la-

1D-C4: Estimation Methods Méthodes d'estimation

for event counts and associated likelihood-based inference procedures. We start with the maximum likelihood estimation procedure and the associated EM algorithm. Following it, a pseudo maximum likelihood procedure is presented using additional information about one of the latent classes. We establish the consistency and asymptotical normality of the pseudo-MLE. Simulation studies are conducted to examine finite sample performance of the pseudo-MLE in efficiency and robustness. The CAYACS physician visit data are used to illustrate the methodology throughout the talk.

tentes pour des dénombrements d'événements et les procédures d'inférence basées sur la vraisemblance qui y sont associées. Nous présentons d'abord la procédure d'estimation du maximum de vraisemblance et son algorithme EM. Par la suite, nous présentons une procédure du maximum de pseudo-vraisemblance à l'aide de renseignements supplémentaires sur l'une des deux classes latentes. Nous établissons la convergence et la normalité asymptotique de l'estimateur du maximum de pseudo-vraisemblance. Nous examinons pour un échantillon fini la performance de cet estimateur en matière d'efficacité et de robustesse. Les données sur les visites des médecins du programme CAYACS sont utilisées pour illustrer la méthodologie tout au long de l'exposé.

1D-C5: Experimental and Sampling Design Plan d'expérience et d'échantillonnage

Chair/Président: Derek Bingham (Simon Fraser University)

Room/Salle: EdSouth 206

Abstracts/Résumés

YIN LI (University of Alberta), **YUANYUAN LIANG** (University of Texas at San Antonio), **JING WANG & KEUMHEE C. CARRIERE** (University of Alberta)

[Monday May 27 / lundi 27 mai, 15:30]

Multiple-Objective Response-Adaptive Repeated Measurement Designs in Clinical Trials for Binary Responses
Plan d'expérience multi-critères s'adaptant aux données pour les études cliniques comportant des mesures binaires répétées

A multiple-objective allocation strategy was recently proposed for constructing response-adaptive repeated measurement designs for continuous responses. We extend the allocation strategy to constructing response-adaptive repeated measurement designs for binary responses. Through computer simulations, we find that the allocation strategy developed for continuous responses also works well for binary responses and it can successfully allocate more patients to better treatment sequences without sacrificing much of estimation precision. However, design efficiency in terms of mean squared error drops sharply as more emphasis is placed on increasing treatment benefit than estimation precision.

Une stratégie d'allocation multi-critères a récemment été développée pour construire un plan s'adaptant aux données pour des mesures continues répétées. Nous généralisons cette stratégie à la construction de plans d'expérience s'adaptant aux données pour des mesures binaires répétées. À l'aide de simulations, nous constatons que la stratégie d'allocation développée pour les variables continues fonctionne bien pour les données binaires et qu'elle peut allouer plus de patients à un meilleur traitement sans pour autant trop sacrifier la précision des estimés. Par contre, l'efficacité du plan en termes d'erreur quadratique moyenne diminue brusquement puisque l'augmentation des bénéfices se trouve plus valorisée que la précision des estimations.

MICHAEL A ROTONDI (York University), **ALLAN DONNER** (University of Western Ontario)

[Monday May 27 / lundi 27 mai, 15:45]

A Confidence Interval Approach to Sample Size Estimation for Interobserver Agreement Studies with Multiple Raters and Outcomes

Approche par intervalle de confiance pour l'estimation de la taille d'échantillon dans des études sur la concordance inter-observateurs à évaluateurs et résultats multiples

Studies measuring interobserver agreement are common in clinical practice. The authors propose a sample size estimation technique to achieve a pre-specified lower and upper limit for a confidence interval for the kappa coefficient. This technique is illustrated using two examples. The first considers a pilot study in oral radiology, whose authors studied the reliability of the mandibular cortical index. The second example examines the level of interob-

Les études mesurant la concordance inter-observateurs sont courantes dans la pratique clinique. Nous proposons une méthode d'estimation de la taille d'échantillon permettant d'obtenir des limites inférieure et supérieure préétablies pour l'intervalle de confiance du coefficient kappa. Nous illustrons cette méthode par deux exemples. Le premier considère une étude pilote en radiologie buccale dont les auteurs ont examiné la fiabilité de l'indice cortico-mandibulaire. Le second examine le niveau de concordance

1D-C5: Experimental and Sampling Design Plan d'expérience et d'échantillonnage

server agreement among four nurses with respect to five triage levels used in the Canadian Triage and Acuity Scale. This method should be useful in the planning stages of an interobserver agreement study.

inter-observateurs parmi quatre infirmières basée sur les cinq niveaux de triage de l'Échelle canadienne de triage et de gravité. Cette méthode pourrait s'avérer utile dans la planification des étapes d'une étude sur la concordance inter-observateurs.

CHRIS WILD (University of Auckland), **JOHN M. NEUHAUS** (University of California, San Francisco), **ALASTAIR J. SCOTT** (University of Auckland)

[Monday May 27 / lundi 27 mai, 16:00]

Likelihood-Based Analysis of Longitudinal Data from Outcome-Dependent Sampling Designs

Analyse basée sur la vraisemblance de données longitudinales provenant de plans d'échantillonnage dépendant des résultats

Investigators commonly gather longitudinal data to assess changes in responses over time and to relate these changes to within-subject changes in predictors. With rare or expensive outcomes such as uncommon diseases and costly radiologic measurements, outcome-dependent sampling plans can improve estimation efficiency and reduce cost. Longitudinal follow up of subjects gathered in an initial outcome-dependent sample can then be used to study the trajectories of responses over time and to assess the association of changes in predictors within subjects with change in response. We develop two likelihood-based approaches for fitting generalized linear mixed models (GLMMs) to such data.

Les chercheurs recueillent souvent des données longitudinales pour évaluer les changements des réponses dans le temps et les associer à des changements des prédicteurs intra-sujets. Pour des résultats peu communs ou dispendieux, comme des maladies rares ou des mesures radiologiques coûteuses, les plans d'échantillonnage qui dépendent des résultats peuvent améliorer l'efficacité de l'estimation et réduire les coûts. Le suivi longitudinal de sujets recueilli dans un échantillon initial dépendant des résultats permet d'étudier les trajectoires des réponses dans le temps et d'évaluer l'association des changements dans les réponses à ceux des prédicteurs intra-sujets. Nous développons deux approches basées sur la vraisemblance pour ajuster des modèles linéaires mixtes généralisés à de telles données.

YANQING YI (Memorial University of Newfoundland)

[Monday May 27 / lundi 27 mai, 16:15]

The Most Powerful Test and the Optimal Design of Response Adaptive Clinical Trials

Test le plus puissant et plan optimal pour essais cliniques adaptatifs à la réponse

Response adaptive clinical trials have the advantages of allocating more patients to the potential better treatment but lead to dependent samples. This talk will introduce the most powerful test for response adaptive clinical trials and explore its properties. Based on the most powerful test statistic, an optimal design will be discussed in the considering statistical power and type I error rate.

Les essais cliniques adaptatifs à la réponse présentent l'avantage d'affecter plus de patients au meilleur traitement potentiel, mais ils donnent lieu à des échantillons dépendants. Dans cet exposé, nous présentons le test le plus puissant pour les essais cliniques adaptatifs à la réponse et nous explorons ses propriétés. Nous discutons ensuite du plan optimal fondé sur le test proposé en évaluant sa puissance statistique et son taux d'erreur de type 1.

2A-A: Gold Medal Address
Allocution du récipiendaire de la médaille d'or

Organizer and Chair / Responsable et président: Bovas Abraham (University of Waterloo)

Room/Salle: CCIS 1-430

Abstract/Résumé

ROB TIBSHIRANI (Stanford University)

[Tuesday May 28 / mardi 28 mai, 8:45]

The Lasso: A Review and Some Recent Developments

Le Lasso: une revue et certains développements récents

The lasso, or L1-penalization, is an active area of research in statistics, and is especially useful for modelling high-dimensional data. I will describe some new developments in this area, including new efficient, computational algorithms, modelling of interactions, and matrix completion. Finally, I will present some new work, describing an exact test for testing significance in an adaptive regression model fit via the lasso. This is joint work with many co-authors, the most recent work co-authored with Richard Lockhart, Jon Taylor, and Ryan Tibshirani

Le lasso, ou L1-pénalisation, est un domaine de recherche actif en statistique, et est particulièrement utile pour la modélisation des données de grande dimension. Je vais décrire quelques nouveaux développements dans ce domaine, y compris de nouveaux algorithmes efficace pour la modélisation des interactions et l'achèvement de matrices. Enfin, je vais vous présenter un nouveau travail, décrivant un test exact pour tester la signification dans un modèle de régression adaptative via le lasso. Ce travail est en commun avec de nombreux co-auteurs, le travail le plus récent avec Richard Lockhart, Jon Taylor et Ryan Tibshirani.

2B-V: Methods to Correct Measures of Effect for Bias Due to Exposure Measurement Error
Méthodes pour corriger la mesure des effets en tenant compte du biais dû aux erreurs de mesure d'exposition

Organizer and Chair / Responsable et présidente: Debbie J. Dupuis (HEC Montréal)

Room/Salle: MedSci 231

Abstract/Résumé

DONNA SPIEGELMAN (Harvard School of Public Health)

[Tuesday May 28 / mardi 28 mai, 10:30]

Methods to Correct Measures of Effect for Bias Due to Exposure Measurement Error

Méthodes pour corriger la mesure des effets en tenant compte du biais dû aux erreurs de mesure d'exposition

Error in the assessment of individual exposure levels usually leads to bias towards the null. In addition, the variability of effect estimates is underestimated, leading to anti-conservative inference. With exposure validation data, methods are available to empirically adjust point and interval estimates of health effects for measurement error. The regression calibration method for single and multiple surrogates will be reviewed and applied to studies of indoor and outdoor air pollution in relation to respiratory health. Extensions to survival data will be presented and applied to a study of the health effects of air pollution in relation to all-cause mortality.

Une erreur dans l'évaluation des niveaux d'exposition individuels entraîne habituellement un biais vers l'hypothèse nulle. De plus, la variabilité des estimations de l'effet est sous-estimée, ce qui entraîne une inférence non-conservatrice. Avec les données de validation de l'exposition, des méthodes sont disponibles pour ajuster de manière empirique l'erreur de mesure dans les estimations ponctuelles et les intervalles de confiance des effets de santé. Nous passerons en revue la méthode de calage par la régression pour des substituts simples et multiples et l'appliquerons aux études sur la pollution de l'air à l'intérieur et à l'extérieur par rapport à la santé respiratoire. Nous étendrons ceci aux données de survie et l'appliquerons à l'étude des effets de la pollution de l'air sur la santé en relation avec toutes les causes de mortalité.

2B-11: Effective Use of Laboratory Sessions in Undergraduate Statistics Teaching **Du bon usage des séances de laboratoire dans les cours de statistique de premier cycle**

Organizer and Chair / Responsable et président: Kevin E. Thorpe (University of Toronto)

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

Room/Salle: ECHA 1-190

Abstracts/Résumés

W. JOHN BRAUN (University of Western Ontario)

[Tuesday May 28 / mardi 28 mai, 10:30]

Naive Analysis of Variance

Analyse usuelle de la variance

Analysis of Variance is often covered in introductory statistics courses, but students often fail to understand it. The derivation of the test statistic and p-value require a mathematical background which may not be understood, thus obscuring the essential ANOVA concept. This could give students an inadequate view of the power of statistics. It is possible to provide students with a graphical technique which clearly conveys the essential concept. The technique can be understood by students with little or no background in probability or statistics. In fact, only the ability to add, subtract and compute averages is required.

L'analyse de la variance est souvent couverte dans les cours d'introduction à la statistique, mais bien souvent les étudiants ne la comprennent pas bien. Dériver la statistique du test et la valeur p nécessite une formation mathématique qui peut ne pas être bien comprise, cachant ainsi le concept essentiel d'ANOVA. Ceci pourrait fournir aux étudiants une vue incomplète de la puissance des statistiques. Il est possible de fournir aux étudiants une technique graphique qui communique clairement le concept essentiel. La technique peut être comprise par des étudiants possédant peu ou pas de bagage en probabilité ou statistique. En fait, l'habileté à additionner, soustraire et à calculer des moyennes est la seule requise.

BRUCE DUNHAM & GAITRI YAPA (University of British Columbia)

[Tuesday May 28 / mardi 28 mai, 11:00]

Using Group-Based Lab Activities in Undergraduate Statistics Teaching

Utilisation des activités de laboratoire en groupe dans l'enseignement de la statistique au premier cycle

Regular laboratory sessions can provide students with valuable learning experiences in their Statistics courses. Yet as with other modes of instruction, lab classes must be carefully planned to be most effective. Described here are suggestions for best practice in enhancing student learning during lab sessions, as implemented in several different courses at UBC. The approach involves group work, pre-reading, and concept-oriented activities using software and on-line applets. Examples are discussed, and contrasts are made with activities considered to be less beneficial to learning.

Des sessions régulières en laboratoire peuvent fournir aux étudiants des expériences d'apprentissage enrichissantes dans leurs cours de statistique. Comme plusieurs modes d'instruction, les classes de laboratoire doivent être planifiées soigneusement pour être des plus efficaces. Des suggestions de meilleures pratiques pour améliorer l'apprentissage des étudiants durant les séances de laboratoire, telles que mises en oeuvre dans différents cours à l'UBC sont ici décrites. L'approche nécessite le travail d'équipe, la pré-lecture et des activités orientées sur les concepts utilisant des logiciels et des applications en ligne. Des exemples sont traités et des contrastes sont faits avec des activités considérées moins bénéfiques à l'apprentissage.

2B-I1: Effective Use of Laboratory Sessions in Undergraduate Statistics Teaching Du bon usage des séances de laboratoire dans les cours de statistique de premier cycle

JIM STALLARD (University of Calgary)

[Tuesday May 28 / mardi 28 mai, 11:30]

The Use of Small-Group Activities in a Post-Calculus, Mathematical Statistics Course

Les activités en petits groupes dans le cadre d'un cours de statistique mathématique dont les étudiants ont déjà suivi un cours de calcul différentiel et intégral

Cobb's 1992 call for changes to statistics education included the use of active-learning activities that enable students to create knowledge. Cooperative learning involving student-work in small groups is a form of active-learning. A pedagogical model involving cooperative learning has been identified as one of the best-practices in the delivery of undergraduate statistics. The personal employment of such has been limited to service-courses. My presentation will focus on the recent implementation of a cooperative pedagogy in a lab setting of a post-calculus introduction to Mathematical Statistics course, and how this affected learning outcomes.

Dans son appel aux changements à l'éducation Statistique de 1992, Cobb inclue activités d'apprentissage actif permettant aux étudiants de créer des connaissances. L'apprentissage coopératif impliquant le travail des étudiants en petits groupes est une telle forme d'activités d'apprentissage actif. Un modèle pédagogique impliquant l'apprentissage coopératif a été identifié comme l'une des meilleures pratiques pour les cours de statistiques de premier cycle. L'emploi personnel de telles méthodes a été restreint cours généraux. Ma présentation se concentrera sur les implémentations récentes d'une pédagogie coopérative dans environnement de laboratoire d'un cours d'introduction aux statistiques mathématiques post-calcul différentiel, et sur les implications aux résultats d'apprentissage.

2B-I2: Functional Data Analysis Analyse de données fonctionnelles

Chair/Président: Duncan Murdoch (University of Western Ontario)
Organizer/Responsable: Jiguo (Jack) Cao (University of Western Ontario)

Room/Salle: ECHA 1-498

Abstracts/Résumés

DAVID CAMPBELL (Simon Fraser University), **FREDERIC FERRATY & PHILIPPE VIEU** (Université Paul Sabatier)
[Tuesday May 28 / mardi 28 mai, 10:30]

Inferring Nonlinear Differential Equation Models from Non-Parameteric Functional Estimators

Comment inférer des modèles d'équations différentielles non linéaires à partir d'estimateurs fonctionnels non paramétriques

Differential equation models are a type of functional regression model where the response is a (functional) derivative, and the covariates are the state functions. Parameters quantifying the relationship between derivative and states are typically easily interpretable, making this class of model appealing for understanding mechanisms and relationships. We outline an exploratory method for reverse engineering relationships based on non-parametric methods with the goal of proposing possible parametric nonlinear differential equation models. The method is based on linking kernel estimates of states and derivatives from data through generalized additive models to infer nonlinear relationships.

Les modèles d'équations différentielles sont un type de modèle de régression fonctionnelle où la réponse est une dérivée (fonctionnelle) et les covariables sont les fonctions d'état. Il est généralement facile d'interpréter les paramètres qui quantifient la relation entre la dérivée et les états, si bien que ce genre de modèle est très utile pour comprendre les mécanismes et les relations. Nous présentons une méthode exploratoire permettant la rétroconception de relations sur la base de méthodes non paramétriques, en vue de développer des modèles d'équations différentielles non linéaires paramétriques. La méthode cherche à relier les estimations par méthode du noyau des états et dérivées de données via des modèles additifs généralisés pour inférer des relations non linéaires.

GILES HOOKER (Cornell University)

[Tuesday May 28 / mardi 28 mai, 11:00]

Selecting the Domain of Integration in a Functional Linear Model

Sélection du domaine d'intégration dans un modèle fonctionnel linéaire

This talk examines the problem of determining which parts of a functional covariate may contribute to a functional linear model. Specifically, we consider the situation in which the coefficient of a functional linear model is known to be zero beyond some unknown point. This arises naturally in the functional convolution model in which a functional response is represented as a functional linear model of the short-term past of the covariate. We examine a change-point approach to detecting the range of integration to be used. The method

Nous examinons le problème de la détermination des quelles parties d'une covariable fonctionnelle peut contribuer à un modèle linéaire fonctionnel. Précisément, nous considérons la situation dans laquelle le coefficient d'un modèle fonctionnel linéaire est nulle au-delà un point inconnu. Cela se produit naturellement dans le modèle de convolution fonctionnel dans lequel une réponse fonctionnelle est représentée comme un modèle fonctionnel linéaire du passé à court terme de la covariable. On examine une méthode de point de changement pour déterminer l'intervalle d'intégration à utiliser. La méthode est illustrée sur les données de

2B-I2: Functional Data Analysis Analyse de données fonctionnelles

is demonstrated on data from a vehicle emissions study. des émissions des véhicules.

JAMES O. RAMSAY (McGill University)

[Tuesday May 28 / mardi 28 mai, 11:30]

Representing and Quantifying Complexity in Functional Observations and Parameters

La représentation et quantification de complexité des observations et paramètres fonctionnels

Roughness-penalized functional data analyses, including smoothing, use a roughness penalty defined by a linear differential operator L to impose smoothness on functional parameter estimates. Finite element methods use the stiffness matrix for the same purpose. The matrix R , consisting of the possible inner products of basis functions transformed by L , defines the complexity of a functional parameter or functional data. The talk will show how the eigenanalysis of R can be used to quantify complexity in existing functions, stabilize functional estimates and show the consequences of choosing L , especially when it is determined by data-defined parameter values.

L'analyse des données fonctionnelles utilise souvent des pénalités de complexité, y compris des pénalités de lissage. La pénalité de complexité est basée sur un choix de l'opérateur linéaire différentielle L , qui donne une matrice R qui contient des produits scalaires de $L \phi_k$, où ϕ_k est une fonction de base. L'analyse de valeurs propres de R quantifie la complexité, et aussi le niveau de chaque type de complexité dans les observations fonctionnelles. Les autres applications de cette technique seront discutées.

2B-I3: Machine Learning and Data Analytics

Apprentissage automatique et analytique de données

Organizer and Chair / Responsable et président: Ali Ghodsi (University of Waterloo)

Room/Salle: EdSouth 254

Abstracts/Résumés

ABBAS KHALILI (McGill University)

[Tuesday May 28 / mardi 28 mai, 10:30]

Sparse Mixture-of-Experts Models in High-Dimensional Classification Problems

Modèles limités de mélange d'experts pour les problèmes de classification de grande dimension

Recent advancements in medical and other fields of scientific research have allowed scientists to collect data of unprecedented size and complexity. A common scientific problem of interest in such applications is to classify subjects into one of the two pre-defined classes 0/1 when there is information available on a large number of features. In this talk we present sparse mixture-of-experts models for 0/1 classification problems in high dimensional data. We will discuss estimation and feature selection methods in these models. Statistical properties of the proposed methods will be explained theoretically and via simulations, and real data analysis.

Les récents progrès en recherche médicale et scientifique permettent aujourd'hui de collecter des données d'une ampleur et complexité sans précédent. L'un des problèmes courants dans ces applications consiste à classer des sujets en l'une des deux classes prédéfinies, 0/1, lorsque des informations sont disponibles pour un grand nombre de caractéristiques. Nous présentons des modèles limités de mélange d'experts pour les problèmes de classification 0/1 des données de dimension élevée. Nous discutons des méthodes d'estimation et de sélection des caractéristiques. Nous expliquons les propriétés statistiques de ces méthodes en théorie et par le biais des simulations et proposons une analyse de données réelles.

CSABA SZEPESVARI, ARASH AFKANPOUR, ANDRAS GYORGY & MICHAEL BOWLING (University of Alberta)

[Tuesday May 28 / mardi 28 mai, 11:00]

A Randomized Mirror Descent Algorithm for Large Scale Multiple Kernel Learning

Algorithme de descente en miroir randomisé pour l'apprentissage de noyaux multiples de grande échelle

We consider the problem of simultaneously learning to combine a very large number of kernels and learn a good predictor based on the learnt kernel. We show the surprising result that in the case of learning the coefficients of a polynomial kernel, a randomized mirror descent algorithm that uses a special (history dependent) distribution can be run in $O(\log(d))$ time per iteration while keeping the variance of gradient estimates low. We prove that the total computational cost of the method to achieve an ϵ -optimal solution to be $O(\log(d)/\epsilon^2)$, thereby allowing our method to operate for very large values of d .

Nous explorons comment apprendre à combiner un très grand nombre de noyaux tout en dégageant un bon prédicteur en fonction du noyau étudié. Nous découvrons, étonnamment, que lorsqu'il s'agit d'identifier les coefficients d'un noyau polynomial, un algorithme de descente en miroir randomisé qui utilise une distribution spéciale (dépendant du passé) peut être exécuté en $O(\log(d))$ de temps par itération tout en minimisant la variance de l'estimation du gradient. Nous prouvons que le coût de calcul total de la méthode permettant de réaliser une solution ϵ -optimal est de $O(\log(d)/\epsilon^2)$, si bien que notre méthode peut fonctionner pour de très grandes valeurs de d .

2B-I3: Machine Learning and Data Analytics

Apprentissage automatique et analytique de données

ANDRAS GYORGY (University of Alberta)

[Tuesday May 28 / mardi 28 mai, 11:30]

Universal Prediction in Changing Environments

Prévision universelle dans des environnements changeants

We consider the problem of online prediction in changing environments. In this framework the performance of a predictor is evaluated as the loss relative to an arbitrarily changing predictor, whose individual components come from a base class of predictors. Typical results in the literature consider different meaningful base classes separately. We present two unifying approaches: one for the case when a black box predictor is able to compete with the base class of policies, and another when competitive prediction performance is achievable via the mirror descent algorithm. Based on joint work with Tamas Linder, Gabor Lugosi, Csaba Szepesvari, and Yaoliang Yu

Nous considérons le problème de la prévision en ligne dans des environnements changeants. Dans ce cadre, la performance d'une variable explicative est évaluée comme la perte par rapport à une variable explicative arbitraire changeante dont les composantes individuelles proviennent d'une classe de base de variables explicatives. Les résultats typiques dans la littérature considèrent séparément différentes classes de base significatives. Nous présentons deux approches unifiées : l'une dans le cas où la variable explicative boîte noire peut entrer en compétition avec la classe de base des politiques et l'autre où la performance de la variable explicative compétitive peut être obtenue à l'aide de l'algorithme de descente du miroir. Basé sur un travail réalisé conjointement avec Tamas Linder, Gabor Lugosi, Csaba Szepesvari et Yaoliang Yu.

2B-I4: Stochastic Models in Insurance Modèles stochastiques en assurance

Organizer and Chair / Responsable et président: Sheldon Lin (University of Toronto)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe d'actuariat

Room/Salle: ECHA L1-430

Abstracts/Résumés

XIAOWEN ZHOU (Concordia University)

[Tuesday May 28 / mardi 28 mai, 10:30]

Lévy Risk Processes

Processus de risque de Lévy

A Lévy process is a stochastic process with stationary independent increments. Its fluctuation behavior is relatively well understood when such a process allows only one-sided jumps. In this talk we are going to review several applications of the fluctuation theory for spectrally one-sided Lévy processes to risk models. In particular, we are going to discuss three classes of risk processes, the risk processes with dividend barriers, the risk processes with taxes and the risk processes with random observations.

Un processus de Lévy est un processus stochastique à accroissement indépendant stationnaire. Ses fluctuations sont relativement bien comprises lorsque le processus ne permet que des sauts unilatéraux. Dans cette présentation, nous explorons plusieurs applications aux modèles de risque de la théorie de fluctuation pour les processus de Lévy spectralement unilatéraux. En particulier, nous discutons de trois classes de processus de risque : les processus de risque avec barrières de dividende, les processus de risque avec taxes et les processus de risque avec observations aléatoires.

CHRISTIANE LEMIEUX, DAVID LANDRIAULT & GORDON WILLMOT (University of Waterloo)

[Tuesday May 28 / mardi 28 mai, 11:00]

Adaptive Premium Policies in Risk Theory

Taux de prime adaptatif en théorie du risque

We present an extension of the classical risk model in which the premium rate policy is adaptive to claims experience. We assume the premium rate is reviewed each time the surplus reaches a new descending ladder height. A choice between a number m of rates is then made depending on the time elapsed between successive ladder heights. We derive explicit expressions for the probability of ruin, assuming claim sizes are mixed Erlang. We then motivate further our model by using a mixed Poisson process for the claims arrival. Finally, we discuss other applications of this method and possible related extensions.

Nous présentons une extension du modèle classique de risque dans lequel le taux de prime est adaptatif à l'expérience de réclamations. Nous supposons que le taux de prime est revu chaque fois que le surplus atteint une nouvelle hauteur d'échelle descendante. On choisit ensuite entre m taux de prime, dépendamment du temps écoulé entre les hauteurs d'échelle successives. Nous trouvons une formule exacte pour la probabilité de ruine, en supposant que la taille des réclamations est Erlang mixte. Nous motivons ensuite notre modèle en utilisant un processus de Poisson mixte pour l'arrivée des réclamations. Finalement, nous discutons de possibles extensions.

2B-I4: Stochastic Models in Insurance Modèles stochastiques en assurance

TAEHAN BAE (University of Regina)

[Tuesday May 28 / mardi 28 mai, 11:30]

Pricing a Motor Insurance Loss Rate Swaption

Établir le prix du taux de perte d'une option d'échange d'une assurance automobile

A few types of motor insurance-linked securities have recently been introduced as an effective means of managing insurance companies' liability risk. In this talk, we introduce a motor insurance loss rate swaption which gives a right for the option holder to enter into a cash flow swap contract. A risk-neutral valuation of the security, assuming that the aggregate loss follows a discounted compound Poisson process will be discussed. As an effective alternative to the computationally intensive Monte Carlo or Fourier inversion methods, we suggest a saddlepoint approximation to the risk neutral price of the hybrid derivative. Some numerical examples will be given based on a set of assumptions.

Quelques types d'assurance automobile liée aux garanties ont récemment introduit une manière efficace de gérer le risque de responsabilité des compagnies d'assurance. Dans cet exposé, nous introduisons une option d'échange d'assurance automobile qui donne le droit au titulaire de l'option d'opter pour un contrat d'échange de liquidités. Nous traiterons d'une évaluation du risque neutre de la garantie, en supposant que la perte agrégée suit un processus de Poisson composé réduit. Comme alternative efficace aux méthodes d'inversion de Monte-Carlo ou de Fourier qui nécessitent d'importants calculs, nous suggérons une approximation du point de selle du prix du risque neutre de la dérivée hybride. Nous donnerons quelques exemples numériques utilisant ces hypothèses.

2B-C1: Modelling the Spread of Infectious Diseases Modélisation de la propagation des maladies infectieuses

Chair/Président: Patrick Brown (University of Toronto)

Room/Salle: EdSouth 228

Abstracts/Résumés

HSING-MING CHANG & RHONDA J. ROSYCHUK (University of Alberta)

[Tuesday May 28 / mardi 28 mai, 10:30]

Simulation Study for a Spatial Scan Test Based on Compound Poisson Model

Étude de simulation pour un test de balayage spatial au moyen d'un modèle de Poisson composé

Spatial cluster analysis has been useful in identifying spatial patterns for: organisms in ecology; precipitation in weather modeling; disease mapping in epidemiology; spatial relation in crime analysis. Very often, 'hot spots' are sought after to identify locations of unusual high incidence rate. We introduce a spatial scan test that identifies clusters of events in a study region based on a compound Poisson model. A simulation study will be presented to illustrate our method for cluster detection of emergency department visits, where individuals may make multiple visits.

L'analyse spatiale en grappes est utile pour déterminer des structures spatiales pour des organismes en écologie, des précipitations en modélisation météorologique, de la cartographie de maladies en épidémiologie ou encore des relations spatiales en analyse criminelle. Les « points chauds » sont souvent hautement recherchés pour repérer des emplacements où l'incidence est anormalement élevée. Nous proposons un test de balayage spatial qui identifie des grappes d'événements dans une région étudiée au moyen d'un modèle de Poisson composé. Nous illustrons notre méthode à l'aide d'une étude de simulation pour la détection de grappes dans les visites à l'urgence, où les personnes peuvent se présenter plus d'une fois.

ROB DEARDON, RAJAT MALIK & PUI SZE KWONG (University of Guelph)

[Tuesday May 28 / mardi 28 mai, 10:45]

Parameterizing Individual-Level Models of Infectious Disease Spread Using Sampling-Based Likelihood Approximations

Établissement des paramètres de modèles au niveau de l'individu pour la propagation de maladies infectieuses par des approximations de la vraisemblance basées sur l'échantillonnage

Individual-level models (ILMs), 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 modelling the infectious pressure exerted by infected individuals of 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 in order to estimate the infectious pressure on each susceptible and, thus, speed-up the likelihood calculation.

Ajustés dans un cadre bayésien de MCMC, les modèles au niveau de l'individu (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 et occasionner une perte de temps pour le calcul de la vraisemblance. Nous présentons ici des méthodes d'échantillonnage permettant d'estimer la pression infectieuse sur chaque individu sensible et, par conséquent, accélérer le calcul de la vraisemblance.

2B-C1: Modelling the Spread of Infectious Diseases Modélisation de la propagation des maladies infectieuses

ASHOK KRISHNAMURTHY (Mount Royal University), **LOREN COBB & JAN MANDEL** (University of Colorado), **JONATHAN BEEZLEY** (CERFACS)

[Tuesday May 28 / mardi 28 mai, 11:00]

Tracking Infectious Disease Epidemics Using Bayesian Data Assimilation

Suivi d'épidémies de maladies infectieuses à l'aide de l'assimilation de données bayésiennes

We explore the use of Bayesian data assimilation methods for the statistical tracking of infectious diseases and to study the spatial dynamics of a disease spreading rapidly from one city to another. The epidemic models that we used for this study are spatial variants of the common susceptible-infectious-removed (S-I-R) compartmental model of epidemiology. To this purpose, we are adapting a family of data assimilation methods known as "Ensemble Kalman Filters". Bayesian statistical tracking, in real-time, of an infectious disease as it unfolds is illustrated for a simulated epidemic wave originating in Calgary, AB and spreading towards Edmonton, AB.

Nous explorons l'utilisation de méthodes d'assimilation de données bayésiennes pour le suivi statistique de maladies infectieuses et l'étude de la dynamique spatiale d'une maladie qui se propage rapidement d'une ville à une autre. Les modèles épidémiques utilisés pour cette étude sont des variantes spatiales du modèle compartimenté sensible-infectieux-retiré (SIR) courant en épidémiologie. À cette fin, nous adaptons une famille de méthodes d'assimilation de données connues sous le nom de "filtres d'ensemble de Kalman". Le suivi statistique bayésien en temps réel d'une maladie infectieuse qui se propage est illustré pour une vague épidémique simulée émanant de Calgary et se propageant vers Edmonton en Alberta.

MAHMOUD TORABI (University of Manitoba)

[Tuesday May 28 / mardi 28 mai, 11:15]

Spatio-Temporal Modeling for Disease Mapping

Modélisation spatio-temporelle pour la cartographie de maladies

In this talk, spatio-temporal models that use conditionally autoregressive smoothing across the spatial dimension and B-spline smoothing over the temporal dimension are considered. These models fall in the class of generalized additive mixed models. The frequentist analysis of these complex models is computationally difficult. Recently developed data cloning method provides a frequentist approach to mixed models. We propose to use data cloning, which yields to maximum likelihood estimation, to conduct frequentist analysis of spatio-temporal modeling of disease ratios. The performance of the proposed approach is evaluated through a real dataset and also by a simulation study.

Dans cet exposé, nous étudions des modèles spatio-temporels qui utilisent le lissage autorégressif conditionnel sur la dimension spatiale et le lissage par B-splines sur la dimension temporelle. Ces modèles font partie de la catégorie des modèles additifs mixtes généralisés. L'analyse fréquentiste de ces modèles complexes est mathématiquement difficile. Une méthode de clonage de données récemment mise au point fournit une approche fréquentiste pour les modèles mixtes. Nous utilisons une approche de clonage de données, qui mène à l'estimation du maximum de vraisemblance, pour réaliser une analyse fréquentiste de la modélisation spatio-temporelle des ratios de la maladie. La performance de l'approche proposée est évaluée à l'aide de données réelles et simulées.

2B-C2: Robust Estimation Estimation robuste

Chair/Présidente: Julie Zhou (University of Victoria)

Room/Salle: ECHA L1-220

Abstracts/Résumés

SEPIDEH FARSINEZHAD, RUSSELL STEELE & MASOUD ASGHARIAN (McGill University)

[Tuesday May 28 / mardi 28 mai, 10:30]

Doubly Robust G-Estimation for Time Varying Outcome via the Kalman Filter

G-Estimation doublement robuste au moyen du filtre de Kalman pour des résultats variant dans le temps

Doubly Robust G-Estimation is a semiparametric method of estimation developed by Robins in 2004 that adjusts for heterogeneity within individuals and over time in a sequential randomized trial. In this paper, for repeated measures on the outcome, we show the utility of assuming a Markov model for the counterfactual when the subject does not receive treatment during study. Our new estimating equation is robust to propensity score mis-specification if the Kalman model holds and still performs reasonably well under assumption violation. Furthermore, analysts can avoid complicated specification of causal models at each time point.

La G-estimation doublement robuste est une méthode d'estimation semi-paramétrique développée par Robins en 2004 qui ajuste l'hétérogénéité dans le temps et au sein des individus dans un essai séquentiel randomisé. Dans cet article, en présence de mesures répétées, nous montrons la pertinence d'un modèle de Markov pour le contrefactuel lorsque le sujet ne reçoit pas de traitement pendant l'étude. Notre nouvelle équation d'estimation est robuste à la mauvaise spécification de scores de propension lorsque le modèle de Kalman est applicable et donne même des résultats raisonnables en cas de non respect des hypothèses. Par ailleurs, les analystes peuvent éviter la spécification complexe de modèles causaux à chaque temps.

GUN HO JANG & LAKSHMI MUTHUSWAMY (Ontario Institute for Cancer Research)

[Tuesday May 28 / mardi 28 mai, 10:45]

Copy Number Variation Detection with Estimated Parameters Using Robust Distribution

Détection de la variation du nombre de copies avec paramètres estimés d'une distribution robuste

High-density single nucleotide polymorphism arrays (SNP) are effective in identifying copy number variations (CNVs) in the whole genome. Existing methods to detect CNVs primarily use log R ratio and B allele frequency while sample-specific parameters are predominantly excluded. In this paper, we propose a robust hidden Markov model in which parameter values are estimated using a computationally efficient EM algorithm. Our model also takes into account inter-marker correlation. Finally, we demonstrated accuracy of our method in detecting various CNV regions using simulated data.

Les tableaux de polymorphismes mononucléotidiques (SNP) de haute densité sont efficaces pour identifier les variations du nombre de copies (VNC) dans l'ensemble du génome. Les méthodes existantes pour détecter les VNC utilisent principalement le rapport log R et la fréquence de l'allèle B, tandis que les paramètres spécifiques à l'échantillon sont majoritairement exclus. Dans cet article, nous proposons un modèle de Markov robuste caché dans lequel les paramètres sont estimés en utilisant un algorithme de calcul EM efficace. Notre modèle prend également en compte la corrélation inter-marqueurs. Enfin, nous démontrons la précision de notre méthode pour la détection de diverses régions VNC à l'aide de données simulées.

2B-C2: Robust Estimation Estimation robuste

ALEXANDRE LEBLANC, STEPHANE DUROCHER & MATTHEW SKALA (University of Manitoba)

[Tuesday May 28 / mardi 28 mai, 11:00]

The Projection Median: A Multivariate Robust Adaptive Weighted Average

La médiane projective: une moyenne multivariée pondérée adaptative robuste

In this talk, we will briefly discuss multivariate robust adaptive weighted averages. In particular, we will look at depth-weighted averages, such as the Stahel-Donoho estimator based on the notion of outlyingness. As an alternative to these computationally intensive methods, we will introduce the projection median, obtained by averaging the univariate medians of all univariate projections of the sample.

Lors de cette présentation, nous examinerons brièvement les moyennes multivariées pondérées adaptatives robustes. En particulier, nous présenterons les moyennes pondérées à l'aide du concept de profondeur d'un point, telles que l'estimateur de Stahel-Donoho. Nous présenterons une alternative à ces méthodes très exigeantes au niveau du temps de calcul, la médiane projective, qui est obtenue en calculant une moyenne des médianes univariées associées à toutes les projections univariées de l'échantillon.

OLLI SAARELA, LÉO RAYMOND-BELZILE & DAVID A. STEPHENS (McGill University)

[Tuesday May 28 / mardi 28 mai, 11:15]

Can There be Bayesian Double Robustness?

La double robustesse bayésienne existe-t-elle?

Doubly robust methods for controlling for confounding in estimation of marginal causal contrasts are difficult to interpret from the Bayesian perspective, since Bayesian inferences are naturally based on a model for the outcome instead of the treatment assignment. Although Bayesian versions of propensity score adjustment have been proposed, the theoretical and practical motivations for these have not been fully explored. We study the properties of predictive inferences drawn from outcome models adjusted for both confounders and some function of the propensity score, emphasizing the role of a well-defined joint probability model.

Les méthodes doublement robustes pour le contrôle des facteurs confondants dans l'estimation de contrastes marginaux causaux sont difficiles à interpréter du point de vue bayésien, car l'inférence bayésienne est naturellement basée sur un modèle pour le résultat au lieu de l'affectation aux traitements. Même si des versions bayésiennes d'ajustement par les scores de propension ont été proposés, les motivations théoriques et pratiques pour ces dernières n'ont pas été pleinement explorées. Nous étudions les propriétés des inférences prédictives issues des modèles ajustés pour deux facteurs confondants et une certaine fonction du score de propension, en insistant sur le rôle d'un modèle bien défini de probabilité conjointe.

2B-C3: Statistical Modeling Modélisation statistique

Chair/Présidente: Laura Cowen (University of Victoria)

Room/Salle: EdSouth 206

Abstracts/Résumés

RASHID AHMED & STEVE BROWN (University of Waterloo)

[Tuesday May 28 / mardi 28 mai, 10:30]

Methods for Missing Data in Discrete Hierarchical Models

Méthodes pour les données manquantes dans les modèles hiérarchiques discrets

The focus of this paper is to investigate inference for discrete hierarchical models in the presence of missing data. Pattern mixture models have been developed to estimate the parameters for the fixed and random effects in the context of generalized linear models. The imputation methods used in this paper are: available case missing value (ACMVPM) restriction, complete case missing value (CCMVPM) restriction, and neighboring case missing value (NCMVPM) restriction. Among the three restriction methods the ACMVPM method performs best. The proposed method provides an alternative to standard selection and pattern-mixture modeling frameworks when data are not missing at random.

Cet article est axé sur l'étude de l'inférence des modèles hiérarchiques discrets en présence de données manquantes. Un modèle de mélange des schémas a été développé pour estimer les paramètres des effets fixes et aléatoires dans le cadre de modèles linéaires généralisés. Les méthodes d'imputation utilisées dans cet article sont les suivantes : restriction sur les valeurs existantes de données manquantes (ACMVPM), restriction sur les valeurs complètes de données manquantes (CCMVPM), restriction sur les valeurs les plus proches de données manquantes (NCMVPM). Entre les trois méthodes de restriction, la première fonctionne le mieux. La méthode proposée offre une alternative aux cadres de sélection et de modélisation de mélange de schémas standards lorsque les données ne sont pas manquantes au hasard.

ESHETU G. ATENAFU (University Health Network, University of Toronto), **WENDY LOU** (University of Toronto)

[Tuesday May 28 / mardi 28 mai, 10:45]

Multistate Markov Models for Joint Modeling

Modèles de Markov multi-état pour modélisation conjointe

Motivated by a study examining the benefit of enteral tube feeding for cystic fibrosis patients, who often have limited follow-up information due to illness, a multistate Markov model approach is proposed. This approach requires minimal assumptions, and offers intuitive examination of changes over time. Following a brief introduction of the motivating example, the proposed method, along with a selection of existing approaches, will be presented. The pros and cons of the methods, based on both the motivating example and the simulation study, will be discussed, including issues such as computational feasibility and availability of statistical programs.

Cet article est motivé par une étude des avantages de l'alimentation entérale par tube pour des patients atteints d'une fibrose kystique, qui ont souvent de l'information de suivi limitée en raison de leur maladie. Nous proposons ainsi une approche fondée sur les modèles de Markov multi-état. Celle-ci nécessite un minimum d'hypothèses et offre un examen intuitif des changements sur la durée. Après une courte introduction de l'exemple motivant, nous présenterons notre méthode ainsi qu'une sélection d'approches existantes. Nous discuterons des avantages et des inconvénients de ces méthodes, basés sur l'exemple motivant et sur l'étude de simulation, y compris les problèmes de calculabilité et de disponibilité des programmes statistiques.

2B-C3: Statistical Modeling Modélisation statistique

NARGES NAZERI RAD & JERRY LAWLESS (University of Waterloo)

[Tuesday May 28 / mardi 28 mai, 11:00]

Multi-State Models of Intermittently Observed Processes with Internal Covariates

Modèles multi-états de processus observés de façon intermittente avec covariables internes

Researchers have studied biomarker processes that are only observed intermittently within a multi-state modelling framework. There are, however, considerable challenges in fitting and checking models incorporating internal covariates. In this talk, we consider multistate models to study viral failure (viral rebound) in HIV-positive persons whose virus has been suppressed by combination antiretroviral treatment. We use multistate models for the viral load values that are taken at intermittent observation times, with internal covariates that are functions of previous viral load history. We discuss some methods of fitting and checking these models, and use them to estimate risks of viral failure.

Des chercheurs ont étudié des processus impliquant des biomarqueurs observés de façon intermittente dans le cadre de modèles multi-états. Cependant, il existe des défis importants pour ajuster et vérifier des modèles qui comprennent des covariables internes. Dans cet exposé, nous considérons des modèles multi-états pour étudier le rebond de la charge virale chez des personnes séropositives dont le virus a été supprimé par un traitement antirétroviral d'association. Pour traiter les valeurs de charge virale prises à des temps d'observation intermittents, nous utilisons des modèles multi-états qui comprennent des covariables internes qui sont fonction de l'historique antérieur de la charge virale. Nous discutons de quelques méthodes d'ajustement et de vérification de ces modèles et nous les utilisons pour estimer les risques d'un rebond de la charge virale.

ARINDAM ROYCHOUDHURY (Columbia University)

[Tuesday May 28 / mardi 28 mai, 11:15]

A Composite Likelihood Method for Estimating Phylogenetic Tree from Dependent Loci

Méthode d'estimation d'arbres phylogénétiques par vraisemblance composite à partir de loci dépendants

We present a coalescent-based composite likelihood for estimating phylogenetic trees from dependent loci. In a set of individuals, the allele counts in nearby genetic loci are usually associated with each other. In some analyses the parameters of interest could be estimated from the marginal distributions of these loci where each locus is a separate data-point. As they are dependent, it is often difficult to derive their joint likelihood. To simplify the derivation, we treat the loci as if they are independent, computing a composite likelihood. We establish that the inferences based on the composite likelihood are consistent.

Nous présentons une vraisemblance composite basée sur la coalescence pour estimer des arbres phylogénétiques à partir de loci dépendants. Pour un groupe de personnes, les décomptes d'allèles dans les loci génétiques à proximité sont généralement liés les uns aux autres. Dans certaines analyses, les paramètres d'intérêt pourraient être estimés à partir des distributions marginales de ces loci, où chaque locus est un point individuel. Comme ils sont dépendants, il est souvent difficile de dériver leur vraisemblance conjointe. Afin de simplifier les calculs, nous supposons que les loci sont indépendants et nous calculons une vraisemblance composite. Nous montrons que les estimateurs basés sur la vraisemblance composite sont convergents.

TYLER WILLIAMSON (Queen's University), **MISHA ELIASZIW** (Tufts University), **GORDON FICK** (University of Calgary)

[Tuesday May 28 / mardi 28 mai, 11:30]

Log-Binomial Models: Concavity and Estimation

Modèles log-binomiaux: concavité et estimation

2B-C3: Statistical Modeling Modélisation statistique

Increasingly, biostatisticians are employing log-binomial models to describe the relationship between a set of predictor variables and a binary outcome. However, standard statistical software occasionally reports failed convergence when fitting log-binomial models. This research directly interrogates the log-likelihood function for a general log-binomial model and demonstrates that it is concave down. However, standard Newton based maximization techniques are shown to be inadequate through the careful consideration of an example from real data from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). The model is estimated by a brute force search of the parameter space to maximize the log-likelihood function.

Les biostatisticiens utilisent de plus en plus les modèles log-binomiaux pour décrire la relation entre un ensemble de variables prédictives et un résultat binaire. Cependant, les logiciels statistiques courants signalent parfois une erreur de convergence lors de l'ajustement de modèles log-binomiaux. Cet article interroge directement la fonction de log-vraisemblance pour un modèle log-binomial général et démontre qu'elle est concave vers le bas. Toutefois, un examen attentif d'un exemple provenant de données réelles du Réseau canadien de surveillance sentinelle en soins primaires (RCSSSP) montre que les techniques de maximisation standards de Newton sont inappropriées. Le modèle est estimé par une recherche assidue dans l'espace paramétrique pour maximiser la fonction de log-vraisemblance.

XIMING XU & JOANNA MILLS FLEMMING (Dalhousie University), **EVA CANTONI** (University of Geneva), **CHRIS FIELD** (Dalhousie University)

[Tuesday May 28 / mardi 28 mai, 11:45]

State-Space Modelling of Fish Stock Maturities

Modélisation à espace d'états de la maturité des stocks de poissons

We introduce a state-space approach to the 3Ps cod maturity data in a DFO's annual sampling program, where the density functions of the observations are Binomial, while the latent state processes are modeled as AR(1). The maximum likelihood estimators of the unknown parameters are obtained based on Laplace approximation to the marginal likelihood of the observations. The sequential particle filtering and smoothing techniques are then applied to recover the underlying dynamics of the maturities through time. Our approach is compared with the sequential Monte Carlo method with state augmentation. Issues on the robustness of our procedure are also discussed.

Nous présentons une méthode à espace d'états pour les données sur la maturité de la morue dans la zone 3P provenant d'un programme d'échantillonnage annuel du ministère des Pêches et des Océans (MPO), dans laquelle les fonctions de densité des observations sont binomiales et les processus de l'état latent sont modélisés en AR(1). Les estimateurs de maximum de vraisemblance des paramètres inconnus sont obtenus au moyen d'une approximation de Laplace de la vraisemblance marginale des observations. Nous appliquons les méthodes séquentielles de filtration et de lissage des particules pour retrouver la dynamique sous-jacente de la maturité dans le temps. Nous comparons notre approche avec la méthode de Monte Carlo séquentielle avec augmentation d'état. Nous discutons aussi des questions liées à la robustesse de notre méthode.

2B-C4: Statistical Theory and Applications I Statistique : théorie et applications I

Chair/Président: Richard Lockhart (Simon Fraser University)

Room/Salle: EdSouth 276

Abstracts/Résumés

MARTIN LYSY (University of Waterloo), **ALEKSANDER LABUDA** (Asylum Research)

[Tuesday May 28 / mardi 28 mai, 10:30]

Stochastic Simulation of a Nanoscale Experiment

Simulation stochastique à l'échelle nanoscopique

With recent developments in nanoscopic measurement technology, researchers can now observe the properties of atoms and molecules at an individual level. However, such observations are often corrupted by complex and poorly-understood sources of noise. A natural framework for modeling this noise is the Generalized Langevin Equation (GLE). In theory, a GLE is derived from Hamiltonian dynamics. In practice, it can be obtained from an intuitive decomposition of the forces acting on the measurement device.

Here, a simple method of simulation for GLEs is presented, and used to calibrate an Atomic Force Microscope (AFM) for a nanoscopic profiling experiment.

Les technologies de mesure nanoscopique permettent maintenant aux chercheurs d'observer les propriétés d'atomes et de molécules au niveau individuel. Cependant, ces observations sont souvent contaminées par de multiples et complexes sources de bruit. L'équation générale de Langevin (GLE) est un modèle naturel pour ces signaux bruités. En théorie, une GLE est dérivée à partir de dynamiques hamiltoniennes. En pratique, elle résulte d'une décomposition intuitive des forces agissant sur l'instrument de mesure. Une méthode de simulation stochastique des GLE sera présentée, ainsi qu'une application au calibrage d'un microscope à force atomique (AFM).

MARIA ELENA RIVERA MANCIA (McGill University)

[Tuesday May 28 / mardi 28 mai, 10:45]

Bayesian Methods in the Study of Extremes

Méthodes bayésiennes pour l'étude des extrêmes

Extreme Value Theory has been widely used for modeling extreme data; however, one of its weaknesses is the threshold selection. Most methods for the threshold choice suffer from concerns over subjectivity and not accounting for the uncertainty. In this study, an analysis of financial institutions internal loss data is performed by using a Bayesian approach. The proposed model considers the form of the distribution below and above the threshold, combining a parametric estimation with a Bayesian approximation to perform inference about the unknown parameters. The threshold is considered as another model parameter and estimation is carried out using MCMC methods.

La théorie des valeurs extrêmes a été largement utilisée pour modéliser les données extrêmes. Cependant, l'une de ses faiblesses est la sélection des seuils. La plupart des méthodes de sélection du seuil donnent lieu à des préoccupations en termes de subjectivité et ne tiennent pas compte de l'incertitude. Dans cette étude, une analyse des données de perte interne d'institutions financières est réalisée avec une approche bayésienne. Le modèle proposé considère la forme de distribution au-dessous et au-dessus du seuil en combinant une estimation paramétrique avec une approximation bayésienne pour faire des inférences sur les paramètres inconnus. Le seuil est considéré comme un autre paramètre de modèle et une estimation est effectuée avec les méthodes de Monte Carlo.

2B-C4: Statistical Theory and Applications I Statistique : théorie et applications I

JEFFREY PICKA (University of New Brunswick)

[Tuesday May 28 / mardi 28 mai, 11:00]

Doubt, Science, and Statistics

Doute, science et statistique

The distinction between statistical and non-statistical approaches to uncertainty in science is that statistics makes constructive use of doubt. These uses are strongly related to the concept of a scientific fact within each field of application. Common issues arising in statistical consulting and in the philosophy of science will be shown to distinguish the scientifically useful application of statistics from other approaches to coping with uncertainty. Implications for statistical education will also be considered.

Contrairement aux méthodes non statistiques, l'approche statistique à l'égard de l'incertitude en science fait un usage constructif du doute. Cet usage est étroitement lié au concept de fait scientifique au sein de chaque domaine d'application. Nous présentons des problèmes courants qui surgissent en consultation statistique et dans la philosophie des sciences afin de mettre en évidence l'utilité scientifique de l'approche statistique par rapport à d'autres méthodes pour traiter l'incertitude. Nous examinons aussi ses répercussions sur l'éducation statistique.

SUSAN ZHE SHENG (University of Western Ontario)

[Tuesday May 28 / mardi 28 mai, 11:15]

Univariate and Bivariate Polynomially Adjusted Saddlepoint Density

Densité en point de selle univariée et bivariée ajustée par polynôme

Improved saddlepoint density approximations are obtained by means of moment-based polynomial adjustments. Some hybrid density approximants are then introduced and applied to various distributions, including mixtures of beta, gamma and Gaussian densities. Additionally, a methodology that involves a standardizing transformation is developed for the bivariate case, and connections to copulas are pointed out. Extensions to univariate and bivariate density estimates, which rely on sample moments and empirical cumulant-generating functions, are also proposed. Several illustrative examples are presented.

Des approximations en point de selle de la densité sont améliorées grâce à un ajustement polynomial basé sur les moments. Des approximations hybrides de la densité sont ensuite présentées et appliquées à différentes distributions, y compris des mélanges de densités beta, gamma et gaussiennes. De plus, une méthodologie comportant une transformation de standardisation est développée pour le cas bivarié, et les liens avec les copules sont décrits. Des généralisations à l'estimation de densités univariées et bivariées, qui sont basées sur les moments observés et la fonction génératrice des cumulants empirique, sont aussi proposées. Plusieurs exemples illustratifs sont présentés.

NANA XING (University of Calgary)

[Tuesday May 28 / mardi 28 mai, 11:30]

Application of Maximum Lq-Likelihood Method

Application de la méthode du maximum de vraisemblance-Lq

This talk is about a new parameter, the maximum Lq-likelihood estimator (MLqE), which was introduced by Ferrari and Yang in 2007. The properties of the MLqE are studied via asymptotic analysis and computer simulations. Then we show that

Dans cet exposé, il est question de l'estimateur du maximum de vraisemblance-Lq (MLqE), un nouveau paramètre introduit par Ferrari et Yang en 2007. Nous étudions les propriétés du MLqE par une analyse asymptotique et des simulations. Nous montrons que le MLqE surpasse l'estimateur

2B-C4: Statistical Theory and Applications I

Statistique : théorie et applications I

the MLqE outperforms the standard MLE, when estimating tail probabilities and quantiles of the Gamma distribution. Here, we mainly assess the relative efficiency between the MLqE and the MLE (Maximum Likelihood Estimator) for various sample sizes, using Monte Carlo simulations.

classique du maximum de vraisemblance lors de l'estimation des probabilités codales et des quantiles de la distribution Gamma. Nous évaluons principalement l'efficacité relative des deux estimateurs pour diverses tailles d'échantillon au moyen de simulations de Monte Carlo.

ALWELL OYET (Memorial University)

[Tuesday May 28 / mardi 28 mai, 11:45]

Estimation of a Lag 2 Dynamic Model with Measurement Error in Covariates

Estimation d'un modèle dynamique à décalage de deux périodes avec erreurs de mesure dans les covariables

We will discuss the estimation of the parameters of a lag 2 dynamic model in the presence of covariates that are measured with error. The lag 2 model is an extension of a lag 1 model which has been used by several authors to model various data including Poisson count data, branching processes and infectious diseases to mention but a few. The performance of the techniques developed will be examined through a simulation study.

Nous discutons de l'estimation des paramètres d'un modèle dynamique à décalage de deux périodes en présence d'erreurs de mesure dans les covariables. Ce modèle est une extension d'un modèle à décalage d'une période utilisé par plusieurs auteurs pour modéliser un vaste éventail de données, y compris des données de dénombrement de Poisson, des processus de ramification et des maladies infectieuses. Nous évaluons la performance des méthodes proposées au moyen d'études de simulation.

2C-11: An Introduction to Relative Survival Analysis Introduction à l'analyse de survie relative

Organizer and Chair / Responsable et président: Russell Steele (McGill University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: ECHA 1-190

Abstract/Résumé

NOORI AKHTAR-DANESH (McMaster University)

[Tuesday May 28 / mardi 28 mai, 13:30]

An Introduction to Relative Survival Analysis

Une introduction à l'analyse de survie relative

Estimating survival measures in population-based cancer studies is challenging because it is difficult to distinguish between patients who die from cancer and those who die from other causes. Relative survival analysis is the recommended technique for estimating survival patterns in population-based cancer registry datasets. It is defined as the observed survival among cancer patients divided by the expected survival in the general population. In this presentation use of a flexible parametric model in estimating relative survival will be discussed. This model is also used in estimating cure fraction rate, excess mortality rate, and net probability and crude probability of death.

L'estimation des mesures de survie dans les études sur le cancer dans la population pose un défi car il est difficile de distinguer les patients qui meurent du cancer et ceux qui meurent d'autres causes. L'analyse de survie relative est la technique recommandée pour estimer les modèles de survie dans les registres sur le cancer dans la population. Elle est définie comme la survie observée parmi les patients cancéreux divisée par la survie prévue dans la population en général. Cette présentation portera sur l'utilisation d'un modèle paramétrique flexible pour estimer la survie relative. Ce modèle est également employé pour estimer le taux de guérison, le taux de mortalité excessive et les probabilités nette et brute de décès.

2C-I2: Design and Methodological Issues in Environmental Diseases Problématiques liées à la méthodologie et aux plans d'expérience en matière de maladies liées à l'environnement

Organizer and Chair / Responsable et président: Lehana Thabane (McMaster University)

Room/Salle: EdSouth 254

Abstracts/Résumés

JUDY-ANNE CHAPMAN (Queen's University)

[Tuesday May 28 / mardi 28 mai, 13:30]

Personal Environment: Design and Analysis of Breast Cancer Trials

Environnement personnel : conception et analyse d'essais cliniques sur le cancer du sein

With a move to personalized or precision medicine, it becomes important to assess the effects of host prognostic factors as part of design, analysis and interpretation. Breast cancer trials of the NCIC Clinical Trials Group will provide concrete examples of assessing personal exposures to vitamin D, celecoxib, aspirin, and osteoporosis therapy and personal risks from genetics, fragility fractures, bone mineral density, weight, and body mass index. Design considerations will include integral/integrated biomarkers, and handling representation and guarantee time bias in prospective and retrospective studies. A survival analysis framework will consider parametric, Cox, and K-M estimation of effects.

Au vu de la tendance actuelle vers une médecine personnalisée de précision, il devient important d'évaluer les effets des facteurs de pronostics individuels dans le cadre de la conception, de l'analyse et de l'interprétation. Les essais cliniques sur le cancer du sein du Groupe de l'INCC offrent des exemples concrets d'une évaluation de l'exposition individuelle à la vitamine D, au célécoxib, à l'aspirine et au traitement contre l'ostéoporose, ainsi que des risques personnels résultant de la génétique, des fractures de fragilité, de la densité minérale osseuse, du poids et de l'indice de masse corporelle. La conception des essais tiendra compte des biomarqueurs intégraux/intégrés et d'une représentation fiable et garantira un biais temporel dans les études prospectives et rétrospectives. L'analyse de survie prendra en considération l'estimation des effets paramétriques de Cox et K-M.

AFISI ISMAILA (GlaxoSmithKline & McMaster University)

[Tuesday May 28 / mardi 28 mai, 14:00]

Design and Analyses Issues in Asthma and COPD Clinical Trials

Problèmes de conception et d'analyse dans les essais cliniques sur l'asthme et la MPOC

This presentation will discuss some of the key methodological challenges and possible solutions for Asthma and COPD clinical trials. We will cover research designs for evaluating efficacy, effectiveness and safety of new interventions in asthma or COPD patients. We will also discuss methods for handling non-compliant patients and patients with preferences in active-controlled trials. The presentation will also include the challenges of endpoint selection when the objectives of the trial include both regulatory and economic evaluation.

Cette présentation discutera de certains problèmes méthodologiques et de solutions envisageables dans les essais cliniques sur l'asthme et la MPOC. Nous traiterons de plans de recherche permettant d'évaluer l'efficacité, l'utilité et la sécurité de nouvelles interventions chez les patients souffrant d'asthme ou de MPOC. Nous discuterons également de méthodes de traitement pour les patients réticents et les patients indiquant des préférences dans les essais avec traitement actif. Enfin, nous aborderons les difficultés que présente la sélection des critères d'effet lorsque l'essai vise une évaluation à la fois réglementaire et économique.

2C-I2: Design and Methodological Issues in Environmental Diseases Problématiques liées à la méthodologie et aux plans d'expérience en matière de maladies liées à l'environnement

TIM RAMSAY & LORELLE WEISS (University of Ottawa), **TYE ARBUCKLE & MANDY FISHER** (Health Canada)

[Tuesday May 28 / mardi 28 mai, 14:30]

Temporal Variability in Urinary Concentrations of Triclosan among Pregnant Women in Canada

Variabilité temporelle de la concentration urinaire de triclosan chez les femmes enceintes au Canada

Triclosan is found in anti-bacterial products and can be absorbed into the body. To explore its potential endocrine disrupting properties, surveys largely rely on a single urine void to quantify exposure. For chemicals with a short half-life, such as triclosan, information on the reliability of these indicators of exposure is lacking. We will evaluate the temporal variability in urinary triclosan levels and calculate the validity of a spot urine sample to predict average exposure to triclosan in pregnant women. Mixed effect models will be fit to determine the association of urinary triclosan concentrations with predictors.

Le triclosan est un des ingrédients des produits antibactériens qui peut être absorbé dans l'organisme. Pour en explorer les éventuels effets de perturbation endocrinienne, les enquêtes se fient généralement à une miction unique pour quantifier l'exposition. Pour des produits chimiques comme le triclosan dont la durée d'activité est courte, nous manquons d'informations concernant la fiabilité de ces indicateurs d'exposition. Nous évaluerons la variabilité temporelle des niveaux de triclosan urinaires et calculerons la validité d'un échantillon ponctuel d'urine pour prévoir l'exposition moyenne au triclosan des femmes enceintes. Nous ajusterons des modèles à effets mixtes pour déterminer l'association de la concentration urinaire de triclosan avec les indicateurs.

2C-I3: Environmental Sampling and Monitoring Échantillonnage environnemental et surveillance

Chair/Président: Martin Lysy (University of Waterloo)
Organizer/Responsable: Yulia R. Gel (University of Waterloo)

Session sponsored by the Survey Methods Section
Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: ECHA L1-430

Abstracts/Résumés

SYLVIA ESTERBY (University of British Columbia - Okanagan)

[Tuesday May 28 / mardi 28 mai, 13:30]

Issues in the Extraction of Information from Water Quality Monitoring Data

Problèmes dans l'extraction de l'information provenant de données sur la surveillance de la qualité de l'eau

Advocacy for monitoring programs and identification of program shortcomings are topics current to environmental monitoring. Several issues relevant to the analysis of existing data and the design of new monitoring programs will be discussed. What appears to be abundant data can be information poor due to spatial and temporal heterogeneity and the nature and number of variables. Large-p-small-n data sets provide poor spatial-temporal characterization. Toxic contaminant data include large numbers of non-detects. Real time monitoring generates questions of appropriate signal capture and online quality control. Examples of statistical approaches used for water quality data will be given.

La défense des programmes de surveillance et l'identification des lacunes du programme sont des sujets d'actualité en surveillance environnementale. Nous présenterons plusieurs problèmes reliés à l'analyse des données existantes et à la conception de nouveaux programmes de surveillance. Des données en apparence riches peuvent contenir peu d'information à cause de l'hétérogénéité spatiale et temporelle ainsi que de la nature et du nombre de variables. Des jeux de données avec grand p, petit n fournissent une mauvaise caractérisation spatio-temporelle. Les données sur les contaminants toxiques contiennent plusieurs valeurs non détectées. La surveillance en temps réel amène des questions sur la saisie adéquate de signaux et le contrôle de la qualité en ligne. Nous donnerons des exemples d'approches statistiques employées sur des données sur la qualité de l'eau.

BRIAN GRAY AND VYACHESLAV LYUBCHICH / BRIAN GRAY ET VYACHESLAV LYUBCHICH (US Geological Survey / University of Waterloo)

[Tuesday May 28 / mardi 28 mai, 14:00]

Regression Estimation of Trends in Temperature when Time and Date of Sampling are Haphazard

Estimation par la régression des tendances de température lorsque l'heure et la date de l'échantillonnage sont fortuites

Trends in air and water temperatures may be confounded with trends in time and/or date of sampling when the latter represent convenience samples. Regression methods for addressing such confounding face multiple challenges: temperature-time of day associations may vary by day; temperature-

Les tendances des températures de l'air et de l'eau peuvent être confondues avec les tendances de l'heure ou de la date d'échantillonnage lorsqu'il s'agit d'échantillons de commodité. Les méthodes de régression pour traiter ce problème posent de nombreux défis : les associations entre la température et le moment de la journée peuvent varier par

2C-I3: Environmental Sampling and Monitoring Échantillonnage environnemental et surveillance

date of sampling associations may vary by year; trends in time and/or date of sampling may be confounded with spatial trends (if spatial locations are sampled irregularly in time or date); and variance estimators of time \times day, space \times day and date \times year associations may be biased when sample sizes are small. I address these concerns using river water temperature and simulated data.

jour ; celles entre la température et la date d'échantillonnage peuvent varier par année ; les tendances entre l'heure ou la date de l'échantillonnage peuvent être confondues avec les tendances spatiales (si les emplacements spatiaux sont échantillonnés de manière irrégulière par heure ou date) ; les associations des estimateurs de variance d'heure \times jour, d'espace \times jour et de date \times année peuvent être biaisés lorsque les tailles d'échantillon sont petites. Je traite ces préoccupations à l'aide de la température de l'eau d'une rivière et de données simulées.

CAROLYN HUSTON (CSIRO), **STEVE THOMPSON** (Simon Fraser University)

[Tuesday May 28 / mardi 28 mai, 14:30]

Where? When? How? Why? Choosing a Sampling Strategy to Support a Bayesian/Complex Model

Où? Quand? Comment? Pourquoi? Choisir une stratégie d'échantillonnage pour supporter un modèle bayésien/complexé

Upon completion of a Bayesian hierarchical model that allowed imputation of fish stock estimates on missing days, the question was posed "Does this mean that we can take weekends off from sampling?" Bayesian models are often used to model complex systems, particularly environmental ones in a monitoring context. Here we consider the "optimality" of different sampling designs for a model when the response being optimized is a surface, and when there are multiple scientific questions of interest being evaluated. A simulation based approach tied to ideas from exploratory data analysis and Bayesian goodness-of-fit tests is proposed and demonstrated.

Après avoir complété un modèle hiérarchique bayésien qui permettait d'imputer les estimations des stocks de poissons pour les jours manquants, la question suivante s'est posée : « Est-ce que cela signifie que l'on peut ne pas sélectionner un échantillon les fins de semaine ? » Les modèles bayésiens sont souvent employés pour modéliser des systèmes complexes, en particulier ceux environnementaux dans un contexte de surveillance. Nous considérons ici l'« optimalité » des différents plans d'expérience d'un modèle lorsque la réponse qui est optimisée est une surface et lorsque nous évaluons plusieurs questions scientifiques complexes d'intérêt. Une approche par simulation liée à des idées provenant de l'analyse exploratoire de données et de tests d'adéquation bayésien est proposée et démontrée.

2C-14: Skew Elliptical Distributions: The State of the Art Les distributions asymétriques elliptiques : technologie de pointe

Chair/Président: Christian Genest (McGill University)
Organizer/Responsable: C. J. Adcock (Sheffield University)

Room/Salle: ECHA 1-498

Abstracts/Résumés

C. J. ADCOCK (Sheffield University)

[Tuesday May 28 / mardi 28 mai, 13:30]

The Multivariate Extended Normal-Gamma Distribution
Distribution normale - gamma étendue multivariée

This paper describes an extended version of the multivariate normal-gamma distribution and its properties. The distribution is closed under conditioning and exhibits the general property that it is possible for standardised values of skewness and kurtosis to become arbitrarily large. These properties imply that the model has potential for applications in financial economics, particularly for asset classes whose returns are severely asymmetric. All moments of the distribution exist; a property which is important for general portfolio selection. The multivariate extended skew-normal distribution arises as a special case.

Cet article décrit une version étendue de la distribution normale - gamma étendue multivariée et ses propriétés. La distribution est fermée sous conditionnement et les valeurs d'asymétrie et d'aplatissement standardisées peuvent devenir arbitrairement grandes. Ces propriétés permettent au modèle une application à l'économie financière, notamment aux catégories d'actifs dont le rendement est sévèrement asymétrique. Tous les moments de la distribution existent, propriété importante pour la sélection générale de portefeuille. La distribution asymétrique - normale étendue multivariée est un cas spécial.

ADELCHI AZZALINI (Università di Padova), **GIULIANA REGOLI** (Università di Perugia)

[Tuesday May 28 / mardi 28 mai, 14:00]

Perturbation of Symmetry in Non-Standard Settings
Perturbation de la symétrie dans les situations non standard

In the last 12-15 years, a great deal of work has been dedicated to the study of so-called 'skew-symmetric distributions'. These are obtained starting from a symmetric 'base' density via a simple mechanism of perturbation which conceptually involves another continuous distribution. The talk describes briefly various extensions of the common setting of this construction. Firstly the support does not need to be the d -dimensional Euclidean space. A more radical extension replaces the requirement of symmetry by a suitable form of generalized symmetry. Moreover, under certain conditions, the constructive mechanism can also be applied to discrete distributions.

Ces 12-15 dernières années, de nombreuses études ont été consacrées aux « distributions asymétriques - symétriques ». Celles-ci s'obtiennent à partir d'une densité « de base » symétrique via un simple mécanisme de perturbation qui implique dans son concept une autre distribution continue. Je décris brièvement plusieurs prolongements de la situation habituelle de cette construction. Tout d'abord, le support n'est pas forcément un espace euclidien à d dimensions. Ensuite, de manière plus radicale, le critère de symétrie peut être remplacé par une forme de symétrie généralisée adaptée. Enfin, sous certaines conditions, le mécanisme constructif peut aussi s'appliquer à des distributions discrètes.

2C-14: Skew Elliptical Distributions: The State of the Art Les distributions asymétriques elliptiques : technologie de pointe

MARC G. GENTON (King Abdullah University of Science and Technology), **YANYUAN MA & MIJEONG KIM** (Texas A&M University)

[Tuesday May 28 / mardi 28 mai, 14:30]

Semiparametric Efficient and Robust Estimation of the Center of an Unknown Symmetric Population under Arbitrary Sample Selection Bias

Estimation efficace et robuste semiparamétrique du centre d'une population symétrique inconnue sous biais de sélection d'échantillon arbitraire

We propose semiparametric estimation methods of the center of a symmetric population when a representative sample of the population is unavailable due to an arbitrary selection bias. We do not impose any parametric form on the population distribution. Under this general framework, we construct a family of consistent estimators that is robust to population model misspecification, and we identify the efficient member that reaches the minimum possible estimation variance. The asymptotic properties and finite sample performance of the estimation and inference procedures are illustrated through theoretical analysis and simulations. A data example illustrates the usefulness of the methods in practice.

Nous proposons des méthodes d'estimation semiparamétrique du centre d'une population symétrique lorsqu'aucun échantillon représentatif de la population n'est pas disponible en raison d'un biais de sélection arbitraire. Nous n'imposons aucune forme paramétrique à la distribution de la population. Dans ce cadre général, nous construisons une famille d'estimateurs convergents qui sont robustes concernant la spécification erronée du modèle de la population, puis nous identifions le membre efficace qui présente la variance d'estimation la moins élevée. Nous illustrons les propriétés asymptotiques et la performance de l'estimation et des procédures d'inférence dans un cas d'échantillon fini par le biais d'une analyse théorique et des simulations. Nos utilisons un exemple de données pour illustrer l'utilité pratique des méthodes.

2C-15: Statistical Learning Apprentissage statistique

Chair/Président: Rob Tibshirani (Stanford University)

Organizer/Responsable: Ryan Tibshirani (Carnegie Mellon University)

Room/Salle: EdSouth 276

Abstracts/Résumés

JACOB BIEN (Cornell University), **JONATHAN TAYLOR & ROB TIBSHIRANI** (Stanford University)

[Tuesday May 28 / mardi 28 mai, 13:30]

A Lasso for Hierarchical Interactions

Un lasso pour les interactions hiérarchiques

The Lasso has become a common tool for building predictive models in high-dimensional settings. Scientists are often interested in models that include interactions between predictor variables. Even when the number of predictors is moderate, the number of interactions can be very large. We develop a Lasso geared specifically to this situation. In particular, we propose a convex optimization problem that produces sparse models that honor the hierarchy restriction that an interaction only be included if the corresponding predictors are in the model as main effects. We show that imposing hierarchy can be thought of as a form of regularization.

Le lasso est devenu un outil courant pour construire des modèles prédictifs dans les situations à dimension élevée. Les chercheurs s'intéressent souvent aux modèles dans lesquels les variables de prévision interagissent. Même si les variables prédictives ne sont pas nombreuses, le nombre d'interactions peut être très grand. Nous mettons au point un Lasso conçu spécialement pour cette situation. En particulier, nous proposons un problème d'optimisation convexe qui produit des modèles limités respectant la limitation hiérarchique selon laquelle une interaction n'est incluse que si les prédicteurs correspondants sont des effets principaux dans le modèle. Nous montrons qu'une hiérarchie peut être une forme de régularisation.

JULIAN WOLFSON (University of Minnesota)

[Tuesday May 28 / mardi 28 mai, 14:00]

Risk Prediction from Electronic Health Record Data: A Naive Bayes Approach

Prévision des risques à partir de dossiers de santé électroniques : une approche bayésienne naïve

Predicting an individual's risk of a clinical outcome is a statistical task with important consequences for both clinicians and public health experts. One data source for this task is electronic health record (EHR) data, which are increasingly available on entire populations under care within U.S. health insurance systems. EHR data are recorded longitudinally, making them ideal for predicting risk, but the dimensionality and complexity of EHR data challenge standard statistical methods. We present an extension of the traditional Naive Bayes (NB) machine learning approach to risk prediction using EHR data with time-to-event outcomes subject to censoring.

Prévoir le risque d'un résultat clinique pour un individu est une tâche statistique ayant d'importantes conséquences à la fois pour les cliniciens et les experts en santé publique. Or les dossiers de santé électroniques (DSE) constituent une source de données de plus en plus accessible pour les populations couvertes par les systèmes d'assurance maladie américains. Les données DSE sont enregistrées sur une base longitudinale, ce qui est idéal pour la prévision des risques, mais leur haute dimensionnalité et leur complexité en font un défi pour les méthodes statistiques standard. Nous présentons un prolongement de l'approche d'apprentissage machine traditionnel bayésien naïf permettant une prévision des risques à partir de données DSE avec des résultats de temps avant l'événement soumis à censure.

2C-15: Statistical Learning Apprentissage statistique

RYAN TIBSHIRANI (Carnegie Mellon University)

[Tuesday May 28 / mardi 28 mai, 14:30]

Fast Stagewise Algorithms for Approximate Regularization Paths

Algorithmes par étapes rapides pour chemins de régularisation approximatifs

Forward stagewise regression enjoys an interesting connection to the lasso: under some conditions, the path of estimates constructed by forward stagewise exactly coincides with the lasso path, as the step size goes to zero. Essentially the same equivalence holds outside of regression, for arbitrary convex, differentiable loss functions subject to an ℓ_1 norm constraint. Stagewise estimates provide a useful approximation even when they do not match their ℓ_1 -constrained analogues, and are computationally appealing. In general, regularization can take many forms, beyond the ℓ_1 norm and sparsity; in this talk, we extend the stagewise idea to general convex regularization problems.

La régression par étapes successives présente des liens intéressants avec le lasso : sous certaines conditions, le chemin des estimations construites par régression par étapes successives coïncide exactement avec le chemin du lasso, à mesure que la taille du pas se rapproche de zéro. Essentially, cette équivalence vaut également en dehors de la régression, pour les fonctions de perte différentiables convexes arbitraires sujettes à une contrainte de norme ℓ_1 . Les estimations par étape offrent une approximation utile même lorsqu'elles ne correspondent pas à leurs analogues limitées par ℓ_1 et elles ne nécessitent pas beaucoup de calculs. En général, la régularisation peut prendre diverses formes, au-delà de la norme ℓ_1 et du manque de densité ; dans cette présentation, nous élargissons le concept des étapes aux problèmes de régularisation convexes généraux.

2C-I6: Statistics in Business and Finance Statistique en gestion et finance

Organizer and Chair / Responsable et président: Mark Reesor (University of Western Ontario)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: EdSouth 206

Abstracts/Résumés

ADAM METZLER (Wilfrid Laurier University)

[Tuesday May 28 / mardi 28 mai, 13:30]

Modeling Stochastic, State-Dependent Correlations

Modélisation des corrélations stochastiques, dépendantes de l'état

Empirical evidence indicates that financial assets tend to be most sensitive to systematic risk factors when these factors take on adverse values. Unfortunately this relationship is not accounted for in many standard approaches to modeling stochastic correlation, most notably in simple mixture models. In this talk we introduce a novel family of correlation models that explicitly accounts for this type of state-dependence, and contains both the simple mixture and RFL models as extreme cases.

Une évidence empirique indique que les actifs financiers semblent plus sensibles aux facteurs de risque systématiques lorsque ces facteurs prennent des valeurs défavorables. Malheureusement, de nombreuses approches habituelles de modélisation de la corrélation stochastique, particulièrement les mélanges simples de modèles, n'en tiennent pas compte. Dans cet exposé, nous introduisons une nouvelle famille de modèles de corrélation qui tient compte de ce type de dépendance d'état et contient les modèles de mélange simple et RFL dans des cas extrêmes.

JESSE COULL (ATB Financial)

[Tuesday May 28 / mardi 28 mai, 14:00]

Credit Risk and Stress Testing

Risque de crédit et test de stress

This presentation provides an introduction to the statistical analyses used to monitor credit risks within a bank's loan portfolio, and to some of the challenges that exist. Focus will be on the estimation of loan losses under adverse macro-economic conditions ("stress testing"), and includes a discussion of the use of Monte-Carlo "importance sampling" to estimate upper percentiles of the loss distribution.

Cette présentation fournit une introduction aux analyses statistiques utilisées pour surveiller les risques de crédit du portefeuille de crédits d'une banque et à quelques-uns de ses défis. L'emphase sera mis sur l'estimation des pertes de crédit sous des conditions macro-économiques défavorables (« test de stress ») et inclura une discussion sur l'utilisation de « l'échantillonnage d'importance » de Monte-Carlo pour estimer les percentiles supérieurs de la distribution de perte.

YURI LAWRYSHYN (University of Toronto)

[Tuesday May 28 / mardi 28 mai, 14:30]

Applied Research Program with Financial Institutions

Programme de recherche appliquée dans les institutions financières

2C-16: Statistics in Business and Finance Statistique en gestion et finance

We provide three examples of research we undertake with our bank sponsors. At the undergraduate level, we developed a hedging calculator for the commodities desk. Simple one/two-factor mean reverting models were used to simulate profit and loss resulting from different trading strategies. At the master's level, we implemented a methodology to value early-stage pharmaceutical startups. The method utilizes a unique real options approach that is easy to implement yet is consistent with economic theory. At the PhD level, we developed a portfolio selection methodology based on second order stochastic dominance that is able to account for investor risk aversion.

Nous fournissons trois exemples de recherche entreprise avec nos commanditaires bancaires. Au niveau sous-gradué, nous avons développé un calculateur de couverture pour le bureau des marchandises. Des modèles de retour moyen simple à un ou deux facteurs ont été employés pour simuler le profit et la perte résultant des différentes stratégies d'échange. Au niveau de la maîtrise, nous avons mis en oeuvre une méthodologie pour évaluer les nouvelles entreprises pharmaceutiques à un stade précoce. La méthode utilise une approche unique d'options réelles facile à mettre en oeuvre et cohérente avec la théorie économique. Au niveau du doctorat, nous avons développé une méthodologie de sélection de portefeuille basée sur la dominance stochastique de deuxième ordre qui peut tenir compte du niveau de risque de l'investisseur.

2C-C1: Probability II Probabilité II

Chair/Président: Hao Yu (University of Western Ontario)

Room/Salle: EdSouth 228

Abstracts/Résumés

HASSAN OMIDI FIROUZI (Université de Montréal), **HIRBOD ASSA** (Concordia University), **MANUEL MORALES** (Université de Montréal)

[Tuesday May 28 / mardi 28 mai, 13:30]

Capital Allocation Problem for a New Ruin-Based Coherent Risk Measure

Problème d'allocation de capital pour une nouvelle mesure cohérente du risque basée sur la ruine

This paper deals with one application of a ruin-based coherent risk measures to a basic problem of finance so called capital allocation problem. In this paper, we introduce a ruin-based coherent risk measure on the space of stochastic processes. As an application, we apply this risk measure to find the allocated capital for an insurer's surplus. In fact, we find that the capital allocation problem for this risk measure has a unique solution determined by Euler allocation method.

Cet article traite une application d'une mesure cohérente du risque basée sur la ruine pour un problème de base en finance, nommément l'allocation de capital. Nous présentons cette mesure de risque définie sur l'espace des processus stochastiques. Nous la mettons en application pour trouver le capital alloué pour l'excédent d'un assureur. Nous constatons que le problème d'allocation du capital pour cette mesure de risque présente une solution unique déterminée par la méthode d'allocation d'Euler.

CLARENCE D. KALITSI, JAN VRBIK & WAI KONG YUEN (Brock University)

[Tuesday May 28 / mardi 28 mai, 13:45]

Approximate Sampling Distributions of the Parameter Estimators in the AR(1)-Model

Distributions échantillonales approximées pour les estimateurs des paramètres dans un modèle AR(1)

We consider the first order auto-regressive model defined by $X_{i+1} = \rho X_i + \varepsilon_{i+1}$. Based on a fully analytical approach, we demonstrate how to obtain the first four moments of some well-known estimators of ρ . This enables us to utilize the Edgeworth series to approximate the corresponding sampling distributions, which vastly improves that of the central limit theorem. More importantly, the resulting approximate sampling distributions perform very well even when the sample size is relatively small. In the case of the maximum likelihood estimator $\hat{\rho}_M$, we further show how this technique can be extended to higher order auto-regressive models.

Nous considérons le modèle auto-régressif de premier ordre défini par $X_{i+1} = \rho X_i + \varepsilon_{i+1}$. Suivant une approche entièrement analytique, nous montrons comment obtenir les quatre premiers moments de certains estimateurs bien connus de ρ . Cela nous permet d'utiliser la série Edgeworth pour approximer les distributions échantillonales correspondantes, ce qui améliore considérablement celle du théorème central limite. Plus important encore, les distributions échantillonales approximées qui en découlent donnent de très bons résultats même pour les tailles d'échantillon relativement faible. Dans le cas de l'estimateur du maximum de vraisemblance $\hat{\rho}_M$, nous montrons en outre comment cette technique peut être étendue à des modèles auto-régressifs d'ordre supérieur.

2C-C1: Probability II Probabilité II

JINGYA LI (University of Western Ontario), **ADAM METZLER** (Wilfrid Laurier University), **MARK REESOR** (University of Western Ontario)

[Tuesday May 28 / mardi 28 mai, 14:00]

Valuation of Contingent Capital Bonds in First-Passage Structural Models

Évaluation d'obligations de capital conditionnel dans des modèles structurels de premier passage

Contingent capital bonds (CCB) are bonds that convert to common shares when a certain pre-determined trigger is breached. We use a first-passage structural model to price CCBs based on a capital structure including deposit, equity, and senior and subordinated debt. Under infinite maturity, we derive a closed-form formula for the CCB's fair price and discuss how various factors affect issuing institution's contingent capital cost. Additionally, simulations confirm that broad conclusions drawn in the perpetual case also hold in the finite-maturity case. All the numerical experiments are based on real data and parameters that are calibrated to Canadian banks.

Les obligations de capital conditionnel (OCC) sont des obligations convertibles en actions ordinaires à l'atteinte d'un point de déclenchement préétabli. Pour établir les prix des OCC, nous utilisons un modèle structurel de premier passage basé sur une structure de capital qui comprend le dépôt, l'équité, ainsi que les créances de premier et de second rang. Dans le cadre d'une échéance indéterminée, nous dérivons une formule analytique pour le juste prix de l'OCC et nous discutons de la façon dont divers facteurs affectent le coût de délivrance du capital conditionnel pour l'institution. En outre, les simulations confirment que les conclusions générales tirées dans ce cadre tiennent également dans le cas d'une échéance déterminée. Toutes les expériences numériques sont basées sur des données réelles et des paramètres calibrés pour les banques canadiennes.

VAHED MAROUFY & PAUL MARRIOTT (University of Waterloo)

[Tuesday May 28 / mardi 28 mai, 14:15]

Computational Aspects of Inference in Local Mixture Models

Aspects computationnels de l'inférence pour des modèles de mélanges locaux

Local Mixture Models (LMM) give a inferentially tractable but still flexible alternative to full mixture models. The parameter space naturally includes boundaries of different sorts. These boundaries mean that computing the maximum likelihood estimate (MLE) is not standard. This talk shows how the geometry of ruled and developable surfaces enable fast and efficient algorithms for finding the MLE to be developed.

Les modèles de mélanges locaux (MML) offrent une alternative à la fois souple et calculable aux modèles de mélanges complets. L'espace paramétrique comprend naturellement des frontières de toutes sortes. Ces limites signifient que le calcul de l'estimateur au maximum de vraisemblance n'est pas standard. Cet exposé montre comment la géométrie de surfaces réglées et développables fournit des algorithmes rapides et efficaces pour trouver cet estimateur à développer.

HARSHA PERERA & TIM SWARTZ (Simon Fraser University), **PARAMJIT GILL** (University of British Columbia - Okanagan)

[Tuesday May 28 / mardi 28 mai, 14:30]

Declaration Guidelines in Test Cricket

Directives pour la déclaration au Test cricket

This project considers the decision problem of when to declare during the third innings of a test cricket match. There are various factors that affect

Ce projet aborde le problème décisionnel du meilleur moment pour procéder à la déclaration durant la troisième manche d'un match de Test cricket. Divers facteurs influent

2C-C1: Probability II Probabilité II

the decision of the declaring team including the target score, the number of overs remaining, the relative desire to win versus draw, and the scoring characteristics of the particular match. Decision rules are developed and these are assessed against historical matches. It is found that teams have traditionally been cautious in declaring, and that optimal decision making would lead to more frequent and earlier declarations.

sur la décision de l'équipe déclarante, y compris le résultat souhaité, le nombre d'overs restants (séries de lancers consécutifs), le désir d'une victoire par rapport à un match nul et les caractéristiques de pointage du match en question. Des règles de décision sont développées et évaluées en fonction de matchs historiques. Nous constatons que les équipes ont traditionnellement été prudentes en déclarant et qu'une prise de décision optimale conduirait à des déclarations plus fréquentes et plus hâtives.

JOSE GARRIDO, IONICA GROPARU & XIAOWEN ZHOU (Concordia University)

[Tuesday May 28 / mardi 28 mai, 14:45]

The Finite-Time Gerber-Shiu Function as a Risk Measure

La fonction de Gerber-Shiu comme mesure de risque

We consider a finite-time version of the Gerber-Shiu (G-S) function defined as follows:

$m_{\delta}(u;t) = E[e^{-\delta(t \wedge T)} w(U_{(t \wedge T)-}, U_{t \wedge T}) | U_0 = u]$, $u \geq 0, t > 0$, for general surplus processes U_t and bi-variate penalty functions w (where T is the time to ruin).

For special choices of the penalty function w that discriminate between ruin and non-ruin events, we show that $m_{\delta}(u;t) = \rho_t(U_t)$ is a risk measure that can be used for hedging positions on the surplus processes U_t .

Numerical illustrations are given for different insurance surplus.

Nous considérons une version de la fonction de Gerber-Shiu (G-S) définie en temps fini comme suit :

$m_{\delta}(u;t) = E[e^{-\delta(t \wedge T)} w(U_{(t \wedge T)-}, U_{t \wedge T}) | U_0 = u]$, $u \geq 0, t > 0$, pour un processus de surplus U_t général et une fonction de pénalité bi-variée w (T est le temps de la ruine).

Pour certains choix de la fonction de pénalité w qui discriminent entre les événements de ruine et non-ruine, nous montrons que $m_{\delta}(u;t) = \rho_t(U_t)$ est une mesure de risque servant à la couverture de positions sur le processus de surplus U_t .

Des illustrations numériques sont présentées pour divers surplus d'assurances.

2C-C2: Statistical Theory and Applications II Statistique : théorie et applications II

Chair/Présidente: Mary Lesperance (University of Victoria)

Room/Salle: ECHA L1-220

Abstracts/Résumés

OKSANA CHKREBTII (Simon Fraser University), **ERIN CAMERON** (University of Alberta), **DAVID CAMPBELL** (Simon Fraser University), **ERIN BAYNE** (University of Alberta)

[Tuesday May 28 / mardi 28 mai, 13:30]

Transdimensional ABC for Inferring Introduction and Spread Rates of Invasive Species

Calculs bayésiens approximatifs transdimensionnels pour l'inférence des taux d'introduction et de propagation d'espèces envahissantes

We develop a new transdimensional Monte Carlo sampler for approximate Bayesian inference when the likelihood consists of intractable variable-dimensional integrals. For example, estimating individual parameters for complex, partially-observed germ grain models precludes the use of standard posterior sampling techniques because the likelihood typically cannot be evaluated. Approximate Bayesian Computation (ABC) methods provide an alternative, but can be slow when relying on inefficient conditional simulation. We show that a transdimensional approach overcomes this problem, and use it to estimate rates of introduction and spread for the invasive earthworms species *Dendrobaena octaedra* (Savigny) along roads in the boreal forest of northern Alberta.

Nous développons un nouvel échantillonneur transdimensionnel de Monte Carlo pour l'inférence bayésienne approximative lorsque la vraisemblance consiste en intégrales incalculables à dimensions variables. Par exemple, l'estimation de paramètres individuels pour des modèles de grains et germes complexes partiellement observés exclut l'utilisation de techniques standards d'échantillonnage a posteriori, parce que la vraisemblance ne peut généralement pas être évaluée. Les méthodes de calculs bayésiens approximatifs (CBA) fournissent une solution de rechange, mais peuvent être lentes lorsqu'elles se fondent sur une simulation conditionnelle inefficace. Nous montrons qu'une approche transdimensionnelle surmonte ce problème et nous utilisons pour estimer les taux d'introduction et de propagation de l'espèce envahissante de lombric *Dendrobaena octaedra* (Savigny) le long des routes dans la forêt boréale du nord de l'Alberta.

CAMILA DE SOUZA & NANCY HECKMAN (University of British Columbia)

[Tuesday May 28 / mardi 28 mai, 13:45]

Switching Nonparametric Regression Models

Modèles de régression non paramétriques à changements d'états

We analyze functional data arising from a curve that, over its domain, changes between J states. We consider a sequence of response variables, y_1, \dots, y_n , where y_i depends on a covariate x_i according to an unobserved state z_i . The z_i 's form a stochastic process with the possible values of z_i being $j = 1, \dots, J$. If $z_i = j$ the expected response of y_i is $f_j(x_i)$. We modify the EM algorithm to estimate the parameters from the state process and

Nous analysons des données fonctionnelles découlant d'une courbe qui, sur son domaine, change entre J états. Nous considérons une séquence de variables réponses, y_1, \dots, y_n , où y_i dépend d'une covariable x_i selon un état non observé z_i . Ces états forment un processus stochastique avec les valeurs possibles de z_i correspondant à $j = 1, \dots, J$. Si $z_i = j$, la réponse espérée de y_i est $f_j(x_i)$. Nous modifions l'algorithme EM pour estimer les paramètres du processus d'état et les fonctions f_1, \dots, f_J . Nous obtenons les erreurs-

2C-C2: Statistical Theory and Applications II

Statistique : théorie et applications II

the functions f_1, \dots, f_J . We obtain standard errors for the parameter estimators of the state process. We conduct simulation studies and an application to a data set.

types des estimateurs des paramètres du processus d'état. Nous réalisons des études de simulation et une application à un ensemble de données.

SHIRIN GOLCHI & RICHARD LOCKHART (Simon Fraser University)

[Tuesday May 28 / mardi 28 mai, 14:00]

A Decision Theoretic Approach for Hypothesis Testing in Particle Physics

Approche décisionnelle théorique pour les tests d'hypothèses en physique des particules

The statistical procedure used in a family of problems in particle physics, similar in nature to the discovery of Higgs Boson, is investigated in this presentation. The problem is considered from a decision theoretic point of view. We introduce a loss function that mimics the features of the current procedure and obtain the Bayes rule. Comparisons are made between the existing and proposed methods in terms of frequentist properties of statistical testing procedures. The comparison results are interpreted as the approximate equivalence of the currently used method with the Bayes rule with certain choice of loss values and/or priors.

Dans cet exposé, nous étudions la procédure statistique utilisée pour une famille de problèmes en physique des particules, de nature similaire à la découverte du boson de Higgs. Le problème est considéré d'un point de vue théorique décisionnel. Nous présentons une fonction de perte qui imite les caractéristiques de la procédure actuelle et nous obtenons la règle de Bayes. Des comparaisons sont faites entre les méthodes existantes et proposées en termes de propriétés fréquentistes des procédures de tests statistiques. Les résultats de la comparaison sont interprétés comme l'équivalence approximative de la méthode utilisée actuellement avec la règle de Bayes pour certains choix de valeurs de pertes et/ou de lois a priori.

JONATHAN LEE, REG KULPERGER & HAO YU (University of Western Ontario)

[Tuesday May 28 / mardi 28 mai, 14:15]

Parallelization of Single Chain MCMC with Applications in Spatial Statistics

Parallélisation de MCMC à chaîne unique pour des applications en statistiques spatiales

Markov Chain Monte Carlo methods are a popular way to sample from complex distributions. Parallel computing is often used to run multiple chains to get samples faster. However, in spatial statistics, we occasionally have chains with long burn-in and poor mixing, so parallelizing a single chain might be better. When there are a large number of dependent parameters, this is non-trivial. We will explore possible ways of using a message passing interface (MPI) framework to do this parallelization.

Les méthodes MCMC représentent un moyen populaire d'échantillonner des distributions complexes. Le calcul parallèle est souvent utilisé pour générer des chaînes multiples et ainsi obtenir des échantillons plus rapidement. Cependant, pour les statistiques spatiales, nous rencontrons parfois des chaînes à longue période de rodage et à mauvais mélange pour lesquelles la parallélisation d'une chaîne unique peut s'avérer une meilleure option. Lorsque les paramètres dépendants sont nombreux, il s'agit d'un défi non trivial. Nous explorons les différentes façons d'utiliser le cadre d'une interface de passage de messages pour procéder à la parallélisation.

2C-C2: Statistical Theory and Applications II Statistique : théorie et applications II

JOSEPH FRANCOIS TAGNE TATSINKOU, PIERRE DUCHESNE & PIERRE LAFAYE DE MICHEAUX (Université de Montréal)

[Tuesday May 28 / mardi 28 mai, 14:30]

Smooth Test of Goodness-of-Fit of Normality for the Errors of an ARMA Process with Unknown Mean
Test lisse d'ajustement pour les erreurs d'un modèle ARMA avec moyenne inconnue

Neyman (1937) derived a goodness-of-fit test for the uniform distribution. This test has been generalized to any distribution by Rayner and Best (1989). Ducharme and Lafaye de Micheaux combined this approach to the Ledwina (1994) data driven principle to find a test for the errors of an ARMA process with known mean. In this work, we generalize the results of Ducharme and Lafaye de Micheaux (2004) for the case of an ARMA process with unknown mean. Some simulations and an application to real data are provided.

Neyman (1937) a proposé un test d'adéquation à la loi uniforme. Le test a été généralisé à une densité quelconque par Rayner et Best (1989). Ducharme et Lafaye de Micheaux (2004) ont utilisé l'approche de Neyman et la procédure de sélection de modèle de Ledwina (1994) pour construire un test de normalité pour les erreurs d'un modèle ARMA avec moyenne connue. Dans ce travail, nous généralisons les résultats de Ducharme et Lafaye de Micheaux dans le cas d'un modèle ARMA avec moyenne inconnue. Des simulations et une application aux données réelles sont fournies.

KASRA YOUSEFI & TIM SWARTZ (Simon Fraser University)

[Tuesday May 28 / mardi 28 mai, 14:45]

Advanced Putting Metrics in Golf
Mesures avancées pour coups roulés au golf

Using ShotLink data that records information on every stroke taken on the PGA Tour, we introduce a new metric to assess putting. The methodology is based on ideas from spatial statistics where a spatial map of each green is constructed. The spatial map provides estimates of the expected number of putts from various green locations. The difficulty of a putt is a function of both its distance to the hole and its direction. A golfer's actual performance can then be assessed against the expected number of putts.

À l'aide de données Shotlink qui enregistrent des informations sur chaque coup joué sur le Circuit de la PGA, nous présentons une nouvelle mesure d'évaluation des coups roulés. La méthodologie est basée sur des principes de statistiques spatiales où une carte spatiale de chaque vert est construite. La carte spatiale fournit des estimés du nombre attendu de coups roulés pour divers emplacements de vert. La difficulté d'un coup roulé est fonction à la fois de la distance du trou et de sa direction. La performance réelle d'un golfeur peut alors être évaluée par rapport au nombre attendu de coups roulés.

2D-A: Isobel Loutit Invited Address Allocution invitée Isobel Loutit

Organizer and Chair / Responsable et président: Executive committee of BISS & Mu Zhu
(University of Waterloo)

Session sponsored by the Business and Industrial Statistics Section
Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: MedSci 231

Abstract/Résumé

YEHUDA KOREN (Google Research)

[Tuesday May 28 / mardi 28 mai, 15:30]

Challenges in Recommender Systems: Science, Art and Pragmatics

Défis des systèmes recommandeurs : science, art et pragmatique

Recommender systems are a main facilitator in personalizing the user experience. A successful design needs to model the elusive human taste, and deal with the fact that humans constantly re-define their preferences. Furthermore, balancing between a detailed modeling of heavy users and an adequate modeling of newcomers makes the task all the more challenging. In this talk I will survey some of the recent advances made in the field, including (as time permits): improving recommendation accuracy, scaling up learning methods, ranking-oriented metrics, fast retrieval of recommendation results, and handling of popularity effects.

Les systèmes recommandeurs sont un modérateur principal dans la personnalisation de l'expérience de l'utilisateur. Un plan fructueux doit modéliser le goût humain insaisissable et traite le fait que les humains redéfinissent constamment leurs préférences. De plus, la tâche devient des plus stimulante lorsqu'elle doit atteindre un juste équilibre entre une modélisation détaillée pour les gros utilisateurs et une modélisation adéquate pour les nouveaux. Dans cet exposé, je passerai en revue quelques propositions faites dans le domaine, incluant (si le temps le permet) : améliorer la précision de la recommandation, augmenter les méthodes d'apprentissage, les méthodes de classement orienté, l'extraction rapide des résultats de la recommandation et le traitement des effets de popularité.

2D-I1: Environmental and Ecological Quality Assessment Évaluation de qualité environnementale et écologique

Organizer and Chair / Responsable et président: Abdel H. El-Shaarawi (American University in Cairo and the National Water Research Institute)

Session sponsored by the Survey Methods Section
Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: ECHA L1-430

Abstracts/Résumés

CHRIS FIELD (Dalhousie University), **ALAN WELSH** (Australian National University)

[Tuesday May 28 / mardi 28 mai, 15:30]

Estimating Bird Abundance: Approaches and Issues
Estimer l'abondance des oiseaux : approches et enjeux

We consider two settings in which the objective is to estimate abundance of particular bird species. The first arises from the Maritime Breeding Bird Atlas. Here we have crude estimates of the numbers of breeding pairs of each species in a $10km^2$ and we want an overall abundance estimate for the species. The second arises where we want to establish a trend in the abundance of a species over a specific study area. Will discuss issues which arise in the use of occupancy models as surrogates for abundance and will show a number of potential problems with this approach.

Nous considérons deux cadres dans lesquels les objectifs sont d'estimer l'abondance d'une certaine espèce d'oiseau. Le premier cadre provient de l'Atlas des oiseaux nicheurs des Maritimes. Nous avons ici une estimation grossière du nombre de couples nicheurs pour chaque espèce à l'intérieur de $10km^2$ et nous voulons estimer l'abondance globale de cette espèce. Le second cadre vise à établir une tendance de l'abondance d'une espèce dans une zone d'étude précise. Nous présenterons les enjeux qui surviennent en employant les modèles d'occupation comme substitut d'abondance et montrerons certains problèmes potentiels de cette approche.

JONAH KEIM (Matrix Solutions), **SUBHASH LELE** (University of Alberta)

[Tuesday May 28 / mardi 28 mai, 16:00]

Statistical Analysis of Camera Trap Data to Study the Effect of Human Activity on Resource Use Patterns by Wildlife: Spatial Recurrent Event Process and Inference

Analyse statistique des données recueillies à l'aide de caméras-pièges pour étudier l'effet de l'activité humaine sur les modèles d'utilisation des ressources par la faune : processus d'événements récurrents spatiaux et inférence

Effect of human development and other activities on wildlife in Northern Alberta, especially on iconic species such as Woodland Caribou, Moose and Wolves, is of immense interest to various stakeholders. To study the effect of human activity on the seasonal and spatial pattern of habitat use by various species in a non-invasive fashion, we set up number of motion sensor cameras in our study area. In this paper, we use spatial recurrent event process models to analyze the patterns of use by

Différents intervenants s'intéressent grandement à l'effet du développement humain et d'autres activités sur la faune du nord de l'Alberta, en particulier sur des espèces emblématiques comme le caribou des bois, l'orignal et les loups. Pour étudier de manière non invasive l'effet de l'activité humaine sur le modèle saisonnier et spatial d'utilisation de l'habitat par différentes espèces, nous installons plusieurs caméras munies de détecteurs de mouvement dans la zone à l'étude. Dans cet article, nous utilisons les modèles de processus d'événements récurrents spatiaux pour analyser les

2D-I1: Environmental and Ecological Quality Assessment Évaluation de qualité environnementale et écologique

Caribou, Moose and Wolves as a function of habitat covariates, season and human activity.

modèles d'utilisation du caribou, de l'orignal et des loups comme une fonction des covariables d'habitat, de saison et d'activité humaine.

STEVE THOMPSON (Simon Fraser University)

[Tuesday May 28 / mardi 28 mai, 16:30]

Dynamic Spatial and Network Sampling

Échantillonnage spatial dynamique et par réseau

Adaptive sampling designs can be useful for sampling populations that are unevenly distributed in space and time. Close connections exist between spatial sampling and network sampling methodologies. In this talk I will describe some recent and new sampling procedures for spatial and network populations that change over time.

Les plans d'échantillonnage adaptatifs peuvent être utiles pour échantillonner les populations distribuées inégalement dans l'espace et le temps. Des liens étroits existent entre les méthodologies de l'échantillonnage spatial et de l'échantillonnage par réseau. Dans cet exposé, je décrirai quelques procédures d'échantillonnage récentes et nouvelles pour des populations spatiales et par réseau qui changent au fil du temps.

2D-I2: Promoting Statistical Literacy - CensusAtSchool and Beyond Promotion de la culture statistique - Recensement à l'école et au-delà

Organizer and Chair / Responsable et président: John Brewster (University of Manitoba)

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

Room/Salle: EdSouth 254

Description

CensusAtSchool (C@S) is an international project that promotes and develops statistical literacy in school-aged children. An international panel will demonstrate how, in different countries, C@S has benefited students and teachers, fostered relationships between statisticians and teachers, been used as a springboard for other statistical education activities and been used to assist in the study of other important topics (related to the environment, for example). In 2012, SSC assumed responsibility for the Canadian node of C@S and the discussion at this session should help to establish future directions for C@S Canada. You will also be shown how you can become involved!

Le Recensement à l'école est un projet international qui promeut et développe les connaissances de base en statistique chez les enfants d'âge scolaire. Un groupe international d'experts démontrera comment, dans différents pays, ce projet a bénéficié aux élèves et aux enseignants et favorisé les relations entre statisticiens et enseignants, en plus de servir de tremplin pour d'autres activités éducatives sur la statistique et de soutien dans l'étude d'autres sujets importants (liés à l'environnement, par exemple). En 2012, la SSC était responsable du noeud canadien de Recensement à l'école et la discussion engagée dans cette séance devrait aider à établir l'orientation future du projet au Canada. Vous découvrirez aussi comment vous pouvez vous impliquer !

2D-I3: Recurrent Event Analysis Analyse d'événements récurrents

Organizer and Chair / Responsable et président: Pierre-Jérôme Bergeron (University of Ottawa)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: ECHA 1-190

Abstracts/Résumés

CECILIA COTTON (University of Waterloo)

[Tuesday May 28 / mardi 28 mai, 15:30]

Causal Inference with Recurrent Events

Inférence causale des événements récurrents

We consider the setting in which a time-to-event outcome is observed in conjunction with a longitudinal process of intermediate, possibly recurrent exasperation events likely associated with the risk of outcome of interest. Advanced techniques are necessary to estimate the causal effect of a non-randomized treatment on the outcome of interest in the presence of the recurrent events and time-varying confounders. Several techniques including marginal structure models and propensity score analysis will be discussed and illustrated with special attention to the strong underlying assumptions being made.

Nous considérons le cadre dans lequel la variable d'intérêt est un temps d'événement observé en combinaison avec un processus longitudinal d'événements exaspérants intermédiaires, possiblement récurrents vraisemblablement associés au risque du phénomène étudié. Des techniques avancées sont nécessaires pour estimer l'effet causal d'un traitement non-aléatoire en présence d'événements récurrents et de facteurs de confusion variant au fil du temps. Nous traiterons plusieurs techniques incluant les modèles de structure marginale et l'analyse de la cote de propension et l'illustrerons en portant une attention spéciale aux fortes suppositions sous-jacentes qui sont faites.

JERRY LAWLESS (University of Waterloo)

[Tuesday May 28 / mardi 28 mai, 16:00]

Deconstructing Event History Data

Déconstruction des données de l'historique des événements

Processes where individuals experience repeated or multiple events arise in many contexts. Understanding the dynamics of such processes is important scientifically and has implications for risk assessment and process management. For example, an event's intensity might be related to previous events, external factors, or characteristics of the individual. Difficulties arise in assessing process dynamics when study groups consist of heterogeneous individuals for whom there is limited follow-up. This talk will discuss some common features of event processes. Ways to assess process

Les processus où les individus sont sujets à des événements répétés ou multiples surviennent dans plusieurs contextes. Comprendre la dynamique de tels processus revêt une grande importance scientifique et a des implications dans l'évaluation du risque et la gestion de processus. Par exemple, l'intensité d'un événement peut être relié aux événements précédents, à des facteurs externes ou aux caractéristiques de l'individu. Des difficultés surgissent dans l'évaluation des processus dynamiques lorsque les groupes à l'étude sont formés d'individus hétérogènes pour lesquels il y a un suivi limité. Dans cet exposé, nous discuterons quelques caractéristiques communes aux processus d'évé-

2D-I3: Recurrent Event Analysis Analyse d'événements récurrents

dynamics will be illustrated on data associated with chronic disease.

nements. Nous illustrerons les manières d'évaluer la dynamique d'un processus à l'aide de données sur une maladie chronique.

RUSSELL STEELE (McGill University), **PIERRE-JÉRÔME BERGERON** (University of Ottawa), **IAN SHRIER** (Jewish General Hospital)

[Tuesday May 28 / mardi 28 mai, 16:30]

Comparing Approaches for Causal Inference for Recurrent Injury Data

Comparaison des approches d'inférence causale sur des données de blessures récurrentes

Past injury history is generally believed to influence the risk of future injury. Standard approaches estimate a decrease in time to injury (increase in hazard) for each successive injury. However, for many approaches, these estimates are confounded if the subjects have heterogeneous frailty for injury. We will present a comparison of standard event time analysis approaches used in injury epidemiology and discuss the advantages and disadvantages of each approach. We present simulated results and a motivating example from the circus troupe Cirque du Soleil.

Les gens croient généralement que les antécédents de blessures passées influencent le risque futur de blessures. Les approches standards estiment une diminution du temps entre deux blessures (augmentation du risque) pour chaque blessure successive. Cependant, pour plusieurs approches, ces estimés sont confondus si les sujets ont une fragilité hétérogène pour les blessures. Nous comparerons les approches habituelles pour analyser les événements au fil du temps qui sont employées en épidémiologie des blessures et discuterons des avantages et des inconvénients de chaque approche. Nous présentons les résultats obtenus par simulation et un exemple du Cirque du Soleil.

2D-I4: Stochastic Models with Applications to Environmental Problems Modèles stochastiques avec applications aux problèmes environnementaux

Organizer and Chair / Responsable et président: Hao Yu (University of Western Ontario)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Room/Salle: ECHA 1-498

Abstracts/Résumés

PRISCILLA GREENWOOD (University of British Columbia), **MARK MCDONNELL** (University of South Australia), **LAWRENCE WARD** (University of British Columbia)

[Tuesday May 28 / mardi 28 mai, 15:30]

Dynamics of Gamma Bursts in Local Field Potentials

Dynamiques des éclatements gamma dans les potentiels locaux de champs

We study a linear stochastic model of gamma bursts in local field potential recordings of interacting populations of excitatory and inhibitory neurons. By analyzing this simple model, which represents a large class of more complicated models, we are able to proceed further with the analysis than has been done before. We show that our model can be approximated in terms of a rotation times a two-dimensional standard Ornstein Uhlenbeck process, and that gamma bursts arise as excursions of the modulus of this Ornstein Uhlenbeck process.

Nous étudions un modèle stochastique linéaire des éclatements gamma dans les enregistrements de potentiels locaux de champs de l'interaction des populations des neurones stimulateurs et inhibiteurs. En analysant ce modèle simple, qui représente une grande classe de modèles plus compliqués, nous sommes en mesure d'aller au-delà des analyses précédentes. Nous montrons que notre modèle peut être approximé en terme d'une rotation multipliée par un processus standard d'Ornstein Uhlenbeck en deux dimensions et des éclatements gamma surgissent comme des excursions du module de ce processus d'Ornstein Uhlenbeck.

YULIYA MARTSYNYUK (University of Manitoba)

[Tuesday May 28 / mardi 28 mai, 16:00]

Consistent Estimators without Identifiability Conditions in Linear Functional and Structural Errors-in-Variables Models

Estimateurs cohérents sans conditions d'identifiabilité dans des modèles linéaires fonctionnels et structurels d'erreurs dans les variables

Linear functional and structural errors-in-variables models will be considered under some new and most general conditions on explanatory and error variables. We will present estimators for unknown slope and intercept parameters of the models without any of the so-called identifiability conditions, and prove weak and strong consistency of these estimators. Typical well known estimators of the slope, intercept and variances of error variables under identifiability conditions, which assume preliminary knowledge of some hard-to-estimate pa-

Les modèles linéaires fonctionnels et structurels d'erreurs dans les variables seront considérés sous de nouvelles conditions, plus générales, sur les variables explicatives et les erreurs. Nous présenterons des estimateurs de paramètres inconnus de pente et d'ordonnée à l'origine pour des modèles sans aucune des conditions d'identifiabilité et prouverons la cohérence faible et forte de ces estimateurs. Nous passerons également en revue les estimateurs habituellement bien connus de la pente, de l'ordonnée à l'origine et des variances des variables d'erreur sous des conditions d'identifiabilité qui supposent une connaissance préalable

2D-I4: Stochastic Models with Applications to Environmental Problems Modèles stochastiques avec applications aux problèmes environnementaux

rameters in the models, will also be revisited. In addition, we will discuss applications of errors-in-variables models to some environmental problems.

de paramètres difficiles à estimer des modèles. De plus, nous traiterons d'applications des modèles d'erreurs dans les variables à quelques problèmes environnementaux.

NATALIA STEPANOVA (Carleton University), **YURI INGSTER** (Yuri Ingster passed away in Aug. 2012. He was a Professor at St. Petersburg State Electrotechnical University.)

[Tuesday May 28 / mardi 28 mai, 16:30]

Adaptive Variable Selection in Nonparametric Sparse Regression

Sélection adaptative des variables d'une régression parcimonieuse non-paramétrique

The problem of exact recovery of an unknown multivariate signal f observed in the Gaussian white noise is studied. It is assumed that, in addition to some smoothness constraints, f possesses an additive sparse structure determined by the sparsity index $\beta \in (0, 1)$. As a consequence of the additive sparsity assumption, the recovery problem transforms to a variable selection problem. Conditions for exact variable selection are provided, and a family of asymptotically minimax variable selection procedures is constructed. The procedures are adaptive in the sparsity index β . This is a joint work with the late Professor Yuri Ingster.

Nous étudions le problème de récupération exacte d'un signal multivarié inconnu f observé dans le bruit blanc gaussien. Nous supposons que f possède une structure additive parcimonieuse déterminée par l'index de sous-population β dans $(0, 1)$, en plus de certaines contraintes de douceur. Suite à l'hypothèse d'additivité parcimonieuse, le problème de récupération en devient un de sélection de variables. Nous fournissons les conditions de sélection exacte des variables et construisons une famille de procédures de sélection de variables asymptotiquement minimax. Les procédures sont adaptatives par rapport à l'index de parcimonie β . Ce travail a été réalisé conjointement avec le regretté professeur Yuri Ingster.

3A-A: CRM-SSC Prize in Statistics Address Allocution du récipiendaire du Prix CRM-SSC

Organizer and Chair / Responsable et président: David A. Stephens (McGill University)

Room/Salle: MedSci 231

Abstract/Résumé

DEREK BINGHAM (Simon Fraser University)

[Wednesday May 29 / mercredi 29 mai, 8:30]

Prediction and Extrapolation Using Calibrated Computer Models

Prédiction et extrapolation à l'aide de modèles informatiques calibrés

An increasingly common task for scientists is to use computational models to make inferences about a process. Oftentimes, this means combining outputs from simulators and field observations to make predictions of the system with associated uncertainty. The setting where one would frequently like to predict is not the one where the observations were taken. In this talk, an overview of the area of computer experiments will be presented. Issues relating to prediction and extrapolation using computational models will be discussed. Special attention will be paid to extrapolation of radiative shocks that were studied at the Center for Radiative Shock Hydrodynamics.

Une tâche de plus en plus commune pour les scientifiques est d'utiliser des modèles informatiques pour tirer des conclusions au sujet d'un processus. Souvent cela implique de combiner des résultats de simulateurs avec des observations sur le terrain afin de faire des prédictions du système avec l'incertitude associée. Fréquemment, le réglage pour lequel on souhaiterait faire des prédictions n'est pas celui pour lequel les observations ont été obtenues. Dans ce séminaire, un aperçu du domaine de l'expérimentation informatique est présenté. Des problèmes liés à la prédiction et l'extrapolation à l'aide de modèles informatiques seront discutés. Une attention particulière sera portée à l'extrapolation de chocs radioactifs étudiés au Centre sur l'hydrodynamique des chocs radioactifs.

3B-V: Ranking and Selection in High-Dimensional Inference Ordonnement et sélection dans l'inférence pour les données de grande dimension

Organizer and Chair / Responsable et présidente: Debbie J. Dupuis (HEC Montréal)

Room/Salle: MedSci 231

Abstract/Résumé

MICHAEL NEWTON & NICHOLAS HENDERSON (University of Wisconsin - Madison)

[Wednesday May 29 / mercredi 29 mai, 10:00]

Ranking and Selection in High-Dimensional Inference

Ordonnement et sélection dans l'inférence pour les données de grande dimension

Consider samples drawn independently from a large number of separate populations, or “inference units”, with each sample providing information about a unit-specific real-valued parameter. Of interest is to identify the units corresponding to large parameters. This “list inference” problem occurs frequently in statistical genomics and elsewhere. I will discuss instances of the problem and a variety of approaches based upon p-values, q-values, local false-discovery rates, posterior expectations, posterior expected ranks, among others. Sampling theory leads us to consider “list content” and “list variance”, which measure useful properties of random lists, and for which some intriguing list-making procedures are computable.

Considérons des échantillons sélectionnés indépendamment d'un grand nombre de populations séparées, ou des « unités d'inférence », où chaque échantillon fournit de l'information sur la vraie valeur du paramètre d'une unité spécifique. Il est d'intérêt d'identifier les unités qui correspondent aux grands paramètres. Ce problème de « liste d'inférence » survient fréquemment en statistique du génome et ailleurs. Je discuterai certains exemples du problème et plusieurs approches basées notamment sur les valeurs p, les valeurs q, les taux locaux de fausse découverte, les espérances a posteriori et les classements espérés a posteriori. La théorie de l'échantillonnage nous amène à considérer le « contenu de liste » et la « variance de liste », lesquelles mesurent des propriétés importantes des listes aléatoires et pour lesquelles il est possible de calculer quelques procédures intrigantes de création des listes.

3B-11: Bootstrap and Change-Point Analysis with Applications Bootstrap et analyse du point de rupture avec applications

Organizer and Chair / Responsable et président: Murray Burke (University of Calgary)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Room/Salle: ECHA 2-150

Abstracts/Résumés

EDIT GOMBAY (University of Alberta), **ABDULKADIR HUSSEIN** (University of Windsor), **KONSTANTINOS FOKIANOS** (University of Cyprus)

[Wednesday May 29 / mercredi 29 mai, 10:00]

Change Detection for Binary Time Series

Détection du changement dans les séries temporelles binaires

Binary time series are of great practical importance as it is used in measuring health care performance, in financial markets, in industrial processes, in climate studies. We survey recent theoretical developments concerning logistic and other regression models that allow AR(p)-type dependence structure and the presence of covariates. Conditions are set for the Maximum Partial Likelihood Estimator's existence and its convergence to the true value at the optimal rate. The performance of the score vector of the partial likelihood function is analysed and used for change detection. Its usefulness will be demonstrated with data on 30-day mortality rates after cardiac surgery.

Les séries temporelles binaires revêtent une grande importance pratique puisqu'elles sont utilisées pour mesurer la performance des soins de santé, dans les marchés financiers, dans les processus industriels, dans les études sur le climat. Nous étudions des développements théoriques récents au sujet des modèles de régression logistique et d'autres modèles qui permettent la structure de dépendance de type AR(p) et la présence de covariances. Des conditions sont établies sur l'existence de l'estimateur du maximum de vraisemblance partiel et sa convergence vers la vraie valeur du taux optimal. Nous analysons la performance du vecteur de scores de la fonction de vraisemblance partielle et l'utilisons pour détecter le changement. Nous démontrerons son utilité à l'aide de données sur les taux de mortalité 30 jours après une chirurgie cardiaque.

MIKLOS CSORGO (Carleton University)

[Wednesday May 29 / mercredi 29 mai, 10:30]

Random Walking around the Bootstrap I

Marche aléatoire autour du bootstrap I

Let X, X_1, \dots be i.i.d. random variables with mean $\mu = EX$. Sampling with replacement from X_1, \dots, X_n , a random sample on X , for each integer $n \geq 1$, we will define $T_{m_n}^*$ and $G_{m_n}^*$ as respective Efron bootstrapped versions of the classical Student t-statistic $T_n(X)$ and pivot $T_n(X - \mu)$. We study $T_{m_n}^*$ and $G_{m_n}^*$ as natural asymptotic pivots for the sample mean \bar{X}_n and the population mean $\mu = EX$ respectively. Both will be seen to relate naturally to sampling

Soient X, X_1, \dots des variables aléatoire i.i.d. de moyenne $\mu = EX$. En échantillonnant avec remise à partir de X_1, \dots, X_n , un échantillon aléatoire sur X , pour chaque entier $n \geq 1$, nous définirons $T_{m_n}^*$ et $G_{m_n}^*$ comme étant respectivement des versions du bootstrap d'Efron de la statistique t de Student $T_n(X)$ et du pivot $T_n(X - \mu)$. Nous étudions $T_{m_n}^*$ et $G_{m_n}^*$ comme des pivots naturels asymptotiques de la moyenne échantillonnale \bar{X}_n et de la moyenne de la population $\mu = EX$. Nous verrons que les deux sont reliés naturellement à l'échan-

3B-I1: Bootstrap and Change-Point Analysis with Applications Bootstrap et analyse du point de rupture avec applications

from a finite population, viewed as an imaginary random sample from an infinite super-population, rather than to sampling from a finite population of real numbers.

tillonnage d'une population finie qui est vue comme un échantillon aléatoire imaginaire d'une super-population infinie, plutôt qu'à un échantillon d'une population finie de nombres réels.

MASOUD NASARI (Carleton University)

[Wednesday May 29 / mercredi 29 mai, 11:00]

Random Walking around the Bootstrap II

Marche aléatoire autour du bootstrap II

The respective Efron bootstrapped versions $T_{m_n}^*$ and $G_{m_n}^*$ of the Student t-statistic $T_n(X)$ and pivot $T_n(X - \mu)$ as in part I (same title) will be defined. Via conditioning on the weights, we establish Berry-Esseen inequalities for both $T_{m_n}^*$ and $G_{m_n}^*$ that will show that, for the same sample size, the confidence intervals for \bar{X}_n and μ , respectively based on the pivots $T_{m_n}^*$ and $G_{m_n}^*$, have a significantly smaller error as compared to the classical t and the classical bootstrapped t confidence intervals.

Nous définirons les versions du bootstrap d'Efron $T_{m_n}^*$ et $G_{m_n}^*$ de la statistique t de Student $T_n(X)$ et du pivot $T_n(X - \mu)$ comme dans la partie I (même titre). En conditionnant sur les poids, nous établissons les inégalités de Berry-Essen pour $T_{m_n}^*$ et $G_{m_n}^*$ qui montreront que pour une même taille d'échantillon, les intervalles de confiance de \bar{X}_n et μ , basés sur les pivots $T_{m_n}^*$ et $G_{m_n}^*$, ont une erreur significativement plus petite que les versions classique t et classiques du bootstrap t des intervalles de confiance.

3B-I2: Methods for Cross-Country and Cross-Cultural Surveys Méthodes pour les enquêtes transnationales et transculturelles

Organizer and Chair / Responsable et président: Christian Boudreau (University of Waterloo)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: ECHA L1-430

Abstracts/Résumés

JEAN DUMAIS (Statistics Canada), **MARC JONCAS** & **SYLVIE LAROCHE** (Statistique Canada)

[Wednesday May 29 / mercredi 29 mai, 10:00]

Standardisation of Sampling Designs and Quality Assurance in International Comparative Studies
Standardisation des plans de sondage et assurance de la qualité dans les enquêtes comparatives

International comparative studies in education have standardised their statistical methods, in particular, the sampling design they use. Benefits and challenges associated with such standardized designs will be examined. Lessons learned during more than fifteen years of practice will be given in closing remarks.

Les études comparatives internationales en éducation ont éprouvé le besoin de normaliser leurs méthodes statistiques, en particulier, les plans de sondage. Nous examinons les bénéfices et défis liés à la mise en place d'un plan de sondage normalisé pour ces études. Nous terminons en discutant des leçons apprises au cours de plus de quinze ans de participation à différentes enquêtes internationales.

CELIA HUANG, DAGMAR MARIACA HAJDUCEK & **MARY THOMPSON** (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 10:30]

Cross-Country Comparisons in International Surveys: Examples Using Frequentist and Bayesian Approaches
Comparaisons entre les pays dans les enquêtes internationales : exemples employant les approches fréquentistes et bayésiennes

Comparisons of survey responses across many countries can be difficult, when the country samples have differing demographic profiles or survey histories. Cross-country modeling can help by adjusting country means and proportions to a standardized population. Comparisons may be confounded by differences in data collection methods and in the measurement properties of questionnaire items. Small measurement differences may be appropriately captured by the incorporation of random country effects into the cross-country model. The model is then useful as a null hypothesis in testing for country differences. The methods are illustrated with examples from the International Tobacco Control Survey.

Les comparaisons des réponses d'enquête entre les pays peuvent être difficiles lorsque les échantillons des pays possèdent des profils démographiques ou des historiques d'enquête différents. La modélisation entre les pays peut aider à ajuster les moyennes et les proportions des pays à des populations normalisées. Les comparaisons peuvent être confondues par les différences entre les méthodes de collecte et les propriétés de mesure des items du questionnaire. De petites différences de mesure peuvent être convenablement saisies en incorporant des effets aléatoires de pays dans le modèle entre les pays. Le modèle sert ensuite d'hypothèse nulle pour tester les différences entre les pays. Des exemples de l'International Tobacco Control Survey sont utilisés pour illustrer les méthodes.

3B-I2: Methods for Cross-Country and Cross-Cultural Surveys Méthodes pour les enquêtes transnationales et transculturelles

BETH-ELLEN PENNELL (University of Michigan)

[Wednesday May 29 / mercredi 29 mai, 11:00]

Toward Best Practice in Cross-National Data Collection: Development of Guidelines

Vers des pratiques exemplaires dans la collecte de données transnationales : développement de lignes directrices

The number and scope of studies covering many cultures, languages, nations, or regions have increased significantly over the past decade. This has led to a growing need to provide information on best practices across the multiple aspects of cross-cultural survey design and implementation to ensure the collection of high-quality comparative data. However, there is very little published material on the details of implementing surveys that are specifically designed for comparative research, including what aspects are to be standardized and when local adaptation is appropriate. The Cross-Cultural Survey Design and Implementation guidelines initiative was created to address this gap (<http://ccsg.isr.umich.edu>).

Le nombre et la portée des études qui couvrent plusieurs cultures, langues, nations ou régions ont augmenté de manière significative au cours de la dernière décennie. Ceci a mené à un besoin grandissant d'information sur les pratiques exemplaires sur plusieurs aspects d'un plan d'échantillonnage transnational et sa mise en oeuvre pour assurer la collecte de données comparables de grande qualité. Cependant, peu de matériel a été publié sur les détails de la mise en oeuvre d'enquêtes développées spécifiquement pour la recherche comparative, incluant quels aspects doivent être standardisés et quand l'adaptation locale est requise. L'initiative du plan d'échantillonnage de l'enquête transnationale et de la mise en oeuvre des lignes directrices a été créée pour combler cette lacune (<http://ccsg.isr.umich.edu>).

3B-I3: Recommender Systems Systèmes de recommandation

Organizer and Chair / Responsable et président: Xiaotong Shen (University of Minnesota)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: ECHA 1-498

Abstracts/Résumés

MU ZHU (University of Waterloo), **PETER FORBES** (University of Oxford), **JENNIFER NGUYEN** (University College London)

[Wednesday May 29 / mercredi 29 mai, 10:00]

Content-Boosted Matrix Factorization Techniques for Recommender Systems

Techniques de factorisation de matrices à contenu amélioré dans les systèmes recommandeurs

Recommender systems are becoming ubiquitous marketing tools. Recommendation algorithms can be either based on content or driven by collaborative filtering. The Netflix contest (2006-2009) rejuvenated a widespread interest in the matrix factorization approach to collaborative filtering. We have been studying ways to incorporate content information directly into this approach. In this talk, I will present two methods. One uses an extra penalty that has a selective shrinkage effect, and other uses a regression constraint. These content-boosted matrix factorization algorithms not only improve recommendation accuracy, but also provide useful insights about the contents, as well as make recommendations more easily interpretable.

Les systèmes recommandeurs deviennent des outils de marketing omniprésents. Les algorithmes de recommandation peuvent être basés sur le contenu ou provenir de filtrage collaboratif. Le concours Netflix (2006-2009) a suscité un intérêt considérable envers l'approche de factorisation de matrices appliquée au filtrage collaboratif. Nous avons étudié des méthodes permettant d'inclure l'information sur le contenu directement à cette approche. Dans cet exposé, je présenterai deux méthodes. L'une utilise une pénalité supplémentaire qui a un effet de rétrécissement sélectif et l'autre utilise une contrainte de régression. Ces algorithmes de factorisation de matrices à contenu amélioré permettent non seulement d'accroître la précision de la recommandation, mais également de fournir des renseignements utiles sur les contenus en plus de faciliter l'interprétation des recommandations.

YUNZHANG ZHU, XIAOTONG SHEN & CHANGQING YE (University of Minnesota)

[Wednesday May 29 / mercredi 29 mai, 10:45]

Partial Latent Models and Sparse Over-Complete Factorizations with Missing Values

Modèles latents partiels et factorisations parcimonieuses sur-complètes avec données manquantes

Personalized information filtering extracts the information specifically relevant to a user, predicting his/her preference over a large number of items. In this talk, we introduce partial latent models to leverage additional user-specific and content-specific predictors. In particular, we factorize a user-over-item preference matrix into a product of two matrices, each representing a user's preference and

Le filtrage d'information personnalisé extrait l'information particulièrement pertinente pour un utilisateur, prédisant sa préférence par rapport à un grand nombre d'items. Dans cet exposé, nous introduisons les modèles latents partiels pour tirer profit de variables explicatives additionnelles spécifiées par l'utilisateur et dont le contenu est spécifique. Plus précisément, nous factorisons une matrice de préférence de l'utilisateur envers l'item en un produit de deux matrices,

3B-I3: Recommender Systems Systèmes de recommandation

item preference respectively. We seek a sparsest latent factorization from a class of over-complete factorizations, possibly with a high percentage of missing values. Computationally, we design methods based on blockwise coordinate descent, decomposing a large optimization problem into many small subproblems, which permits parallelization.

chacune représentant la préférence d'un utilisateur et celle d'un item. Nous cherchons une factorisation latente parcimonieuse d'une classe de factorisations sur-complètes, possiblement avec un pourcentage élevé de valeurs manquantes. Relativement au calcul, nous concevons des méthodes basées sur une descente coordonnée par bloc en décomposant un vaste problème d'optimisation en plusieurs sous-problèmes ce qui permet la parallélisation.

3B-C1: Dependence II Dépendance II

Chair/Présidente: Johanna Nešlehová (McGill University)

Room/Salle: EdSouth 276

Abstracts/Résumés

HELA ROMDHANI (McGill University), **LOUIS-PAUL RIVEST** & **M'HAMED LAJMI CHAIEB-LAKHAL** (Université Laval)

[Wednesday May 29 / mercredi 29 mai, 10:00]

Kendall's tau for hierarchical data

Tau de Kendall pour des données hiérarchiques

We introduce two exchangeable Kendall's tau as non parametric measures of associations for hierarchical data. We provide estimators for these measures and investigate their asymptotic properties under hierarchical Archimedean and meta-elliptical copula models. For data drawn from meta-elliptical hierarchical copulas we deduce estimators for the two intra-class correlation coefficients. We also develop test procedures for cluster and sub-cluster effects. Performances are assessed using Pitman efficiencies and Monte Carlo study.

Nous introduisons deux tau de Kendall échangeables comme mesures d'association pour les données hiérarchiques. Nous en proposons des estimateurs dont nous étudions les propriétés asymptotiques sous des modèles de copules Archimédiennes et meta-elliptiques hiérarchiques. Pour les données provenant de copules meta-elliptiques hiérarchiques, nous déduisons des estimateurs pour les coefficients de corrélation intra-classe. Nous développons, enfin, des procédures de tests pour les effets des groupes et des sous-groupes. Leurs performances sont évaluées à l'aide de calculs d'efficacité de Pitman et de simulations.

LIQUN DIAO, RICHARD COOK & **KER-AI LEE** (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 10:15]

A New Copula Model for Marked Point Processes

Nouveau modèle de copule pour les processus ponctuels marqués

Many chronic diseases feature recurring clinically important events with which there often exists an associated random variable realized upon the occurrence of each event. We describe a novel model for a marked point process which incorporates a dependence between continuous marks and the event process through the use of a copula function. The copula formulation ensures that event times can be modeled by any intensity function for point processes and any joint model can be specified for continuous marks. The relative efficiency of separate and joint analyses is examined through simulation. An application in transfusion medicine is given for illustration.

Plusieurs maladies chroniques présentent des événements cliniques importants récurrents pour lesquels il existe souvent une variable aléatoire associée se réalisant en même temps que chaque événement. Nous décrivons un nouveau modèle pour un processus ponctuel marqué qui intègre la dépendance entre les marques continues et le processus de l'événement à l'aide d'une copule. La formulation de la copule permet que les temps d'événements soient modélisés par n'importe quelle fonction d'intensité des processus ponctuels et que n'importe quel modèle conjoint puisse être utilisé pour les marques continues. Nous examinons l'efficacité relative des analyses séparées et conjointe par voie de simulation et nous illustrons notre approche à l'aide d'une application en médecine transfusionnelle.

3B-C1: Dependence II Dépendance II

MÉLINA MAILHOT (Concordia University), **ÉTIENNE MARCEAU** (Université Laval), **MHAMED MESFIOUI** (Université du Québec à Trois-Rivières)

[Wednesday May 29 / mercredi 29 mai, 10:30]

*Bivariate Value-at-Risk and Tail Value-at-Risk
VaR et TVaR bivariées*

We establish several new propositions from reformulations of the bivariate lower and upper orthant Value-at-Risk (VaR). We study their behavior, and we motivate their use for risk allocation, to represent ruin probabilities and for risk comparison. In particular, new results on the convexity conditions for the marginals and the dependence model are provided. These theoretical results are used to build lower and upper confidence regions for random vectors, to establish new concepts of risk allocation, and to build dependence bounds for functions of two random variables. Then, we present the bivariate Tail Value-at-Risk, based on previous results.

En reformulant la lower et upper orthant VaR bivariée, nous présenterons plusieurs nouvelles propositions par rapport à ces mesures de risque. Nous étudierons les particularités des courbes et motiverons l'usage de la VaR bivariée pour l'allocation de capital, les probabilités de ruine et pour comparer des risques. En particulier, de nouveaux résultats sur la convexité des courbes seront présentés. Ils permettront d'établir des régions de confiance bivariées, de développer l'allocation de capital basée sur ces mesures bivariées et de borner des fonctions de variables aléatoires. Nous présenterons aussi la TVaR bivariée, basée sur les résultats pré-établis.

WAYNE OLDFORD & DEREK WANG (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 10:45]

*Visual Hypothesis Tests of Independence
Tests d'indépendance visuels*

Integrating traditional inferential thinking into exploratory data analysis has been a long standing challenge. Recently the logic of significance tests has been formally applied to data visualization methods (e.g. Buja et al., 2009, Phil. Trans. Royal Soc. A.). The test statistic becomes a combination of the display and the viewer. The observed significance level is determined empirically. In this talk, I will briefly review the logic of visual hypothesis tests and then conduct a few with the audience. The focus will be on assessing the independence of two variables. A variety of tests will be applied. Come be a test statistic.

L'intégration du raisonnement inférentiel classique dans l'analyse exploratoire des données est un défi de longue date. La logique des tests d'hypothèse a été appliquée de façon formelle à des méthodes de visualisation de données (comme dans Buja et coll., 2009, Phil. Trans. Royal Soc. A.). La statistique de test devient une combinaison de la présentation et de l'auditoire. Le niveau de signification observé est déterminé de façon empirique. Dans cet exposé, je passerai brièvement en revue la logique des tests d'hypothèses visuels, puis j'en effectuerai quelques-uns avec l'auditoire. L'accent sera mis sur l'évaluation de l'indépendance entre deux variables. Un éventail de tests sera appliqué. Venez et transformez-vous en statistique de test.

ANTONIO PUNZO (Università di Catania), **LUCA BAGNATO** (Università Cattolica del Sacro Cuore), **LUCIO DE CAPITANI** (Università di Milano-Bicocca)

[Wednesday May 29 / mercredi 29 mai, 11:00]

*Improving the Autodependogram Using the Kullback-Leibler Divergence
Amélioration de l'autodépendogramme à l'aide de la divergence de Kullback-Leibler*

3B-C1: Dependence II Dépendance II

The autodependogram is a graphical device recently proposed in the literature to analyze autodependencies. It is defined computing the classical Pearson statistics of independence at various lags in order to point out the presence lag-dependencies. This paper proposes an improvement of this diagram obtained by substituting the Pearson statistics with an estimator of the Kullback-Leibler divergence between the bivariate density of two delayed variables and the product of their marginal distributions. A simulation study, on well-established time series models, shows that this new autodependogram is more powerful than the previous one. An application to financial data is also shown.

L'autodépendogramme est un dispositif graphique apparu récemment dans la littérature permettant d'analyser les autodépendances. Il est défini en calculant la statistique d'indépendance classique de Pearson à divers décalages en vue de révéler la présence de dépendances dans ces décalages. Cet article propose une amélioration de ce diagramme obtenu en substituant la statistique de Pearson par un estimateur de la divergence de Kullback-Leibler entre la densité de deux variables différées et le produit de leurs distributions marginales. Une étude de simulation basée sur des modèles de séries chronologiques bien établis montre que ce nouvel autodépendogramme est plus puissant que le précédent. Une application à des données financières est également présentée.

JULES JOSEPH S. DE TIBEIRO (Université de Moncton), **BERNARD COLIN** (Université de Sherbrooke)

[Wednesday May 29 / mercredi 29 mai, 11:15]

Optimal Quantization: An Information Theory Approach

Le codage optimal : un point de vue de la théorie de l'information

Based on the notion of mutual information between the components of a random vector, we construct, for data reduction reasons, an optimal quantization of the support of its probability measure. More precisely, we propose a simultaneous discretization of the whole set of the components of the random vector which takes into account, as much as possible, the stochastic dependence between them. Examples are presented.

Partant de la notion d'information mutuelle entre les composantes d'un vecteur aléatoire, nous construisons, dans l'optique de réduction des données, un codage optimal de partitionnement d'un support de mesure de probabilité d'un vecteur aléatoire donné. Cela consiste précisément, à minimiser la perte de l'information mutuelle en définissant l'ensemble de tous les intervalles tels que le codage du support qui en résulte puisse conserver la dépendance stochastique entre les composantes du vecteur aléatoire. Des exemples sont présentés.

3B-C2: Estimation and Hypothesis Testing Estimation et tests d'hypothèses

Chair/Présidente: Joan Hu (Simon Fraser University)

Room/Salle: EdSouth 228

Abstracts/Résumés

MOHAMED AMEZZIANE (DePaul University), **SYED EJAZ AHMED** (Brock University)

[Wednesday May 29 / mercredi 29 mai, 10:00]

The Limiting Distribution of a Class of Shrinkage Estimators and Related Inference
La distribution asymptotique des estimateurs obtenus par rétrécissement

We derive the limiting distribution for estimators that are obtained by shrinking unbiased estimators towards points in the parameter space. The family of these limiting distributions depends on the distance between the "guess" and the true parameter. The limiting distribution is used to derive the null distribution for a new class of test statistics that tend to outperform classical ones under certain common conditions. We compare the performance of the new testing methodology to that of classical ones for different types of parameters. Theoretical results are supported by a simulation study.

Nous étudions la distribution asymptotique des estimateurs obtenus par rétrécissement des estimateurs non-biaisés envers des points dans l'espace du paramètre. La famille de ces distributions asymptotiques dépend de la distance entre le point et le paramètre. La distribution sous l'hypothèse nulle est obtenue à partir de cette distribution asymptotique pour une nouvelle classe de statistiques de tests dont la performance est meilleure que celle des tests classiques, sous des conditions usuelles. Les résultats théoriques sont illustrés par des simulations Monte Carlo.

SONG CAI & JIAHUA CHEN (University of British Columbia)

[Wednesday May 29 / mercredi 29 mai, 10:15]

Dual Empirical Likelihood Ratio Test under Semi-Parametric Density Ratio Models for Multiple Samples
Test dual du rapport de vraisemblances empirique sous les modèles de rapport de densité semi-paramétriques pour échantillons multiples

Density ratio model (DRM) is a flexible semi-parametric model for the relationship among the underlying densities of multiple samples. Empirical likelihood (EL) is a powerful tool for the inference of DRM parameters. However, if any two underlying distributions are identical, the EL is not defined in the neighbourhood of the true parameter. For two sample problem, a dual-likelihood approach has been proposed by Keziou and Leoni-Aubin to get around this problem. We extended this dual approach to multiple sample situations, and showed that the corresponding dual-likelihood ratio has a simple Chi-squared limiting distribution under a class of general composite hypothesis.

Le modèle relatif au rapport de densité est un modèle semi-paramétrique permettant de déterminer la relation entre les densités sous-jacentes de plusieurs échantillons. La vraisemblance empirique (VE) est un outil puissant pour l'inférence des paramètres des modèles de ratio de densité. Cependant, si aucune des deux distributions sous-jacentes n'est identique, la vraisemblance empirique n'est pas définie dans l'environnement du vrai paramètre. Keziou et Leoni-Aubin ont proposé une approche duale de vraisemblance pour contourner ce problème. Nous avons étendu cette approche duale à plusieurs échantillons et avons montré que le rapport de vraisemblance duale correspondant a un simple chi carré limitant la distribution sous une classe d'hypothèses composite générale.

3B-C2: Estimation and Hypothesis Testing Estimation et tests d'hypothèses

SHAKHAWAT HOSSAIN & MELODY GHAHRAMANI (University of Winnipeg)

[Wednesday May 29 / mercredi 29 mai, 10:30]

Shrinkage Estimation in Linear Regression with AR(m)-GARCH(p,q) Errors

Estimation par rétrécissement dans des modèles de régression linéaire avec erreurs AR(m)-GARCH(p,q)

We consider the shrinkage estimation method for linear regression models with AR(m)-GARCH(p,q) errors when some of the regression parameters are restricted to a subspace. We derive the asymptotic distributional biases and risks of the shrinkage estimators using a large sample theory. A Monte Carlo simulation study is conducted to examine the relative performance of the proposed estimators with the maximum likelihood estimator (MLE). Our simulation study shows that the shrinkage estimators dominate the MLE in the entire parameter space. We illustrate the proposed method by analyzing a real life data set from econometrics.

Nous considérons la méthode d'estimation par rétrécissement pour des modèles de régression linéaire avec erreurs AR(m)-GARCH(p,q) lorsque certains paramètres de régression sont restreints à un sous-espace. Nous dérivons les biais et les risques distributionnels asymptotiques des estimateurs de rétrécissement au moyen de la théorie asymptotique. Nous réalisons une étude de la simulation de Monte Carlo pour examiner l'efficacité relative des estimateurs proposés avec l'estimateur de vraisemblance maximum (EVM). Notre étude de simulation montre que les estimateurs par rétrécissement dominent l'estimateur de maximum de vraisemblance dans tout l'espace du paramètre. Nous illustrons cette méthode en analysant un jeu de données réelles en économétrie.

YINGLI QIN (University of Waterloo), **WEIMING LI** (Beijing University of Posts and Telecommunications)

[Wednesday May 29 / mercredi 29 mai, 10:45]

Hypothesis Testing for Large Dimensional Covariance Matrices

Tests d'hypothèses pour des matrices de covariance de grande taille

In this paper, we propose statistics for testing whether a covariance matrix is an identified matrix and the equality of two covariance matrices when the data dimension is high compared to the sample size. The proposed test statistics are based on the Stieltjes transform of the spectral distribution of the sample covariance matrix. Through Random Matrix Theory, we prove that the proposed test statistics are asymptotically generalized chi-square distributed under null hypotheses. Simulation results are presented to show that the proposed tests outperform some existing methods in various cases.

Dans cet article, nous présentons des statistiques pour tester si une matrice de covariance est égale à une matrice donnée et si deux matrices de covariance sont égales alors que la dimension des données est élevée par rapport à la taille de l'échantillon. Les tests statistiques proposées sont basées sur une transformation de Stieltjes de la distribution spectrale de la matrice de covariance empirique. Selon la théorie des matrices aléatoires, nous prouvons que sous les hypothèses nulles, les statistiques proposées suivent une distribution de χ^2 généralisée asymptotique. Nous présentons des simulations qui montrent que les tests proposés surpassent certaines méthodes existantes dans plusieurs cas.

JOEL SANGO & PIERRE DUCHESNE (Université de Montréal)

[Wednesday May 29 / mercredi 29 mai, 11:00]

Modelization of Multivariate Cointegrated Time Series

Modélisation des séries chronologiques multivariées et cointégrées

This article considers the estimation and the validation of multivariate periodic cointegrated times series models. To fit the model, two approaches

L'article porte sur l'estimation et la validation des modèles périodiques multivariés cointégrés. Deux approches sont discutées afin d'estimer le modèle : l'estimation de plein

3B-C2: Estimation and Hypothesis Testing Estimation et tests d'hypothèses

are discussed: full rank estimation and reduced rank estimation. To validate the estimated model, we study the sample matricial autocovariances of the innovations, the residual autocovariances and their asymptotic distributions. With these results, portmanteau type tests are proposed and their finite-sample properties are examined through a simulation study.

Key Words: order of cointegration, reduced rank estimation, portmanteau-type tests.

rang et l'estimation de rang réduit. Pour la validation du modèle estimé, nous étudions les matrices d'autocovariance échantillonales reposant sur le processus innovation non-observable, les matrices d'autocovariance résiduelles ainsi que leurs distributions asymptotiques. Basés sur ces résultats, des tests de type portmanteau sont construits. Nous procéderons aussi à des simulations pour étudier le comportement de ces tests statistiques en échantillons finis.

Mot clés : rang de cointégration, estimation de rang réduit, tests portmanteau.

EKATERINA TUZOV & YULIYA MARTSYNYUK (University of Manitoba)

[Wednesday May 29 / mercredi 29 mai, 11:15]

Exploring Functional CLT Confidence Intervals for a Population Mean in the Domain of Attraction of the Normal Law
Exploration d'intervalles de confiance du TCL fonctionnel pour la moyenne d'une population dans le domaine d'attraction de la loi normale

We take a Studentized partial sums process based on independent copies of a random variable X , with trajectories in the function space $D[0, 1]$. As a consequence of a functional central limit theorem for this process with X in the domain of attraction of the normal law, we consider convergence in distribution of several functionals of this process and derive respective asymptotic confidence intervals for the mean of X . We explore the expected lengths of the obtained confidence intervals and, in particular, conclude that some of our confidence intervals are shorter than that obtained from a Studentized central limit theorem.

Nous supposons un processus de sommes partielles studentisées basé sur des copies indépendantes d'une variable aléatoire X dont les trajectoires sont dans l'espace de fonctions $D[0, 1]$. En conséquence d'un théorème central limite (TCL) fonctionnel pour ce processus avec X dans le domaine d'attraction de la loi normale, nous étudions la convergence en loi de plusieurs fonctionnelles de ce processus et nous dérivons les intervalles de confiance asymptotiques respectifs pour la moyenne de X . Nous examinons les longueurs prévues des intervalles de confiance obtenus et concluons notamment que certains de nos intervalles sont plus courts que ceux obtenus d'un théorème central limite studentisé.

3B-C3: Models Accounting for Measurement Error and Meta-Analysis Modèles avec erreurs de mesure et méta-analyse

Chair/Président: Duncan Murdoch (University of Western Ontario)

Room/Salle: EdSouth 206

Abstracts/Résumés

MIREYA DIAZ (Henry Ford Hospital)

[Wednesday May 29 / mercredi 29 mai, 10:00]

Performance of Multivariate Logistic Regression for Meta-Analysis of Diagnostic Accuracy

Performance de la régression logistique multivariée pour la méta-analyse de la précision diagnostique

There is a scarcity of information regarding the statistical performance of multivariate logistic regression models for meta-analysis of non-comparative categorical outcomes. The presence of considerable bias in the variance components for the univariate case indicates that this knowledge is necessary. The present simulation study fills in this gap, assessing the fixed and random effects estimators in a comprehensive set of scenarios for the case of meta-analysis of diagnostic accuracy using both frequentist and Bayesian approaches. We will discuss the virtues and limitations of these approaches and provide practical recommendations.

Il existe peu d'information sur la performance statistique des modèles de régression logistique multivariée pour la méta-analyse de résultats catégoriels non comparatifs. L'existence de biais importants dans la variance pour des réponses univariées indique qu'une connaissance plus approfondie est nécessaire. Notre étude de simulation comble cette lacune en évaluant les estimateurs des effets fixes et aléatoires dans une gamme complète de scénarios pour la méta-analyse de la précision diagnostique à l'aide des approches fréquentiste et bayésienne. Nous discutons des vertus et des limites de ces approches et nous présentons des recommandations pratiques.

MARIA ESTHER PEREZ TREJO & ROBERT PLATT (McGill University)

[Wednesday May 29 / mercredi 29 mai, 10:15]

Statistical Modeling of Outcomes Subject to Measurement Error Induced by Double Clustering via a Bayesian Hierarchical Model

La modélisation statistique des réponses exposées à une erreur de mesure induite pour double conglomération avec un modèle hiérarchique bayésien

A type of measurement error, known as double clustering, occurs in cluster randomized trials when the measurement of the outcome is naturally clustered within the same groups used as units for randomization. The variation due to inherent clustering cannot be separated from the variation between evaluators, unless additional audited observations (assumed free of variation due to double clustering) are available. A Bayesian hierarchical model is proposed, and estimation of the treatment effect conducted via MCMC methods. A simulation-based study is conducted, under different scenarios of double-clustering sever-

Un type d'erreur de mesure, la double grappe, advient dans les études randomisées par grappe quand la mesure de la réponse forme une grappe naturelle dans les mêmes groupes utilisés comme unités de randomisation. Cette variation ne peut être séparée de la variabilité entre les évaluateurs, à moins que d'autres observations vérifiées (ne souffrant pas de double grappe) sont disponibles. Un modèle bayésien hiérarchique est proposé, et l'estimation de l'effet du traitement est fait par MCMC. Une étude par simulation est effectuée, selon plusieurs divers degrés de double grappe et proportion de sujets vérifiés, pour des réponses normales et binaires.

3B-C3: Models Accounting for Measurement Error and Meta-Analysis Modèles avec erreurs de mesure et méta-analyse

ity/proportion of audited subjects, for normally distributed and binary outcomes.

YING YAN & GRACE YI (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 10:30]

Corrected Profile Likelihood for Cox Model with Covariate Measurement Error

Profil de vraisemblance corrigé pour le modèle de Cox avec erreur de mesure des covariables

Cox proportional hazards models with covariate measurement error have been studied extensively, and various approaches have been proposed to adjust for measurement error effects under the classical additive measurement error models. However, similarity and difference of the existing methods are rarely studied. In this talk, we propose the corrected profile likelihood approach, and show that many existing methods can be unified by our approach. We further generalize our corrected profile likelihood approach to accommodating general measurement error models. Various simulation studies and a real data example are presented.

Les modèles à risques proportionnels de Cox avec erreur de mesure des covariables ont été largement étudiés et plusieurs approches ont été proposées pour traiter les effets des erreurs de mesure sous les modèles additifs classiques d'erreur de mesure. Cependant, les similarités et différences des méthodes existantes sont rarement étudiées. Dans cet exposé, nous proposons l'approche par profil de vraisemblance corrigé et nous montrons qu'elle peut rassembler plusieurs méthodes existantes. Nous généralisons davantage notre approche pour traiter des modèles généraux d'erreur de mesure. Nous présentons diverses études de simulation et un exemple fondé sur des données réelles.

QIAOHAO ZHU (Alberta Health, Government of Alberta), **KIMMIE C. CHOUGH** (University of Alberta)

[Wednesday May 29 / mercredi 29 mai, 10:45]

Detecting and Correcting Publication Bias in Meta-Analysis: A Maximum Likelihood Approach

Détection et correction du biais de publication dans les méta-analyses: une méthode basée sur le maximum de vraisemblance

We formulate publication bias in meta-analysis as truncated normal distribution problem, and propose a new parametric model for detecting and correcting publication bias. We develop maximum likelihood method of estimating the underlying overall effect size and the severity of publication bias. We derive ML functions corresponding to the two situations that cause publication bias: small effect size and large p-value. We carried out extensive simulation study to evaluate the performance of our methodology, and compared our method to the Trim and Fill method. We find that our method performs consistently well in detecting and correcting publication bias in various situations.

Nous définissons le biais de publication dans les méta-analyses comme un problème de distribution normale tronquée et nous proposons un nouveau modèle paramétrique pour la détection et la correction de ce biais. Nous développons la méthode du maximum de vraisemblance pour estimer l'ampleur de l'effet sous-jacent et l'importance du biais de publication. Nous dérivons des fonctions du maximum de vraisemblance correspondant aux deux situations à l'origine de ce biais : une faible ampleur de l'effet et une grande value-p. Nous avons mené une vaste étude de simulation pour évaluer la performance de notre approche, que nous avons comparée à la méthode «trim and fill». Nous concluons que notre méthode donne de bons résultats de façon constante pour détecter et corriger le biais de publication dans divers contextes.

3C-A: SSC Impact Award Address

Allocution du récipiendaire du prix pour impact de la SSC

Organizer and Chair / Responsable et présidente: Georgia Roberts (Statistics Canada)

Session in honour of the 2012 SSC Impact Award winner, David Binder

Séance en l'honneur du récipiendaire du prix pour impact de la SSC 2012, David Binder

Room/Salle: MedSci 231

Abstracts/Résumés

JOHN ELTINGE (Bureau of Labour Statistics)

[Wednesday May 29 / mercredi 29 mai, 13:00]

Variance Function Models for Complex Sample Surveys

Modèles de fonction de la variance dans les enquêtes complexes

In work with complex sample surveys, variance function models can provide useful tools for improvement of design, for integration of multiple data sources, and for construction of confidence intervals and other inferential statistics. This paper explores the conceptual basis for these variance functions; develops related approximations based on large-sample and small-deviation conditions; reviews estimating-equation approaches to estimation and inference for variance-function models with complex survey data; and presents related diagnostics. The main ideas are illustrated with examples from the U.S. Consumer Expenditure Survey and the U.S. Current Employment Statistics program. Keywords: Composite estimation; design effect; generalized variance function; superpopulation model

En travaillant sur des enquêtes complexes, les modèles de fonction de la variance peuvent fournir des outils utiles pour améliorer la conception, pour fusionner plusieurs sources de données et pour construire des intervalles de confiance et d'autres statistiques inférentielles. Cet article explore le cadre conceptuel de ces fonctions de variance, développe des approximations basées sur des conditions de grande taille d'échantillon et de faible variation, passe en revue les approches d'estimation reliées aux équation dans l'estimation et l'inférence pour des modèles de fonction de la variance dans les données d'enquêtes complexes et présente des diagnostics qui y sont associés. Les principales idées sont illustrées à l'aide d'exemples du U.S. Consumer Expenditure Survey et du U.S. Current Employment Statistics program. Mots clés : Estimation composite ; effet de plan ; fonction de variance généralisée ; modèle de super-population

J.N.K. RAO (Carleton University)

[Wednesday May 29 / mercredi 29 mai, 13:30]

Weighted Estimating Equations Approach to Inference from Complex Survey Data: Impact of David Binder

Approche des équations estimantes pondérées à l'inférence de données d'enquêtes complexes : impact de David Binder

Standard methods for inference using estimating equations require modifications for analyzing complex survey data. David Binder's pioneering 1983 ISR paper introduced design-weighted estimating equations that lead to design-consistent estimators of parameters defined as solutions to population estimating equations. A linearization method

Il importe d'apporter des modifications aux méthodes usuelles d'inférence à l'aide d'équations estimantes pour analyser les données provenant d'enquêtes complexes. L'article de David Binder dans l'ISR en 1983 a introduit les équations estimantes pondérées selon le plan qui mènent à l'estimation de paramètres cohérente avec le plan comme solutions des équations estimantes de la population. Une méthode

3C-A: SSC Impact Award Address

Allocution du récipiendaire du prix pour impact de la SSC

was used to derive design-consistent variance estimators. I will present several applications of the weighted estimating equations approach as well as extensions based on weighted composite score equations to handle more complex problems of making inference on variance and covariance components associated with multi-level models.

par linéarisation est utilisée pour dériver des estimateurs de variance cohérents avec le plan. Je présenterai plusieurs applications de l'approche des équations estimantes pondérées ainsi que des extensions basées sur les équations pondérées du score composite pour traiter des problèmes plus complexes tels que l'inférence des composantes de variance et de covariance associées aux modèles à multiveaux.

MARY THOMPSON (University of Waterloo)
[Wednesday May 29 / mercredi 29 mai, 14:00]
Measuring Causal Effects With Surveys
Mesure des effets causaux dans les enquêtes

David Binder's work has had a profound impact on the analysis of survey data by researchers in the social and health sciences, where causation and the measurement of effects of interventions are of great importance. The talk will review David's writing on the subject of modeling and measuring causation. A discussion will follow of recent advances in mediational modeling in intervention research with population surveys.

Le travail de David Binder a eu un profond impact sur l'analyse des données d'enquêtes par les chercheurs en sciences sociales et de la santé, où la causalité et la mesure des effets des interventions revêtent une grande importance. L'exposé passera en revue les articles de David sur la modélisation et la mesure de la causalité. Une discussion suivra sur les progrès récents de la modélisation médiationnelle dans la recherche sur les interventions dans les enquêtes sur la population.

3C-I1: Issues and Challenges Faced by New Researchers Problèmes et défis auxquels font face les nouveaux chercheurs

Organizer and Chair / Responsable et présidente: Cecilia Cotton (University of Waterloo)

Session sponsored by the Committee on New Investigators

Séance parrainée par le Comité des nouveaux chercheurs

Room/Salle: EdSouth 254

Description

The transition from graduate school to postdoctoral fellow to assistant professor is both challenging and rewarding. The purpose of this panel discussion is to discuss various issues of primary importance to new researchers such as: the job interview process, writing grant proposals, teaching, publishing research, building a research program, service, collaborations, work-life balance and others. Panel participants will briefly discuss their personal experiences and then offer insights and advice in a moderated open discussion format.

La transition des études supérieures au postdoctorat, puis à un poste de professeur adjoint est une expérience à la fois exigeante et enrichissante. L'objectif de cette discussion en groupe est d'aborder diverses questions d'importance capitale pour les nouveaux chercheurs, y compris le processus d'embauche, la rédaction de demandes de subventions, l'enseignement, la publication d'articles, l'élaboration d'un programme de recherche, le service à la communauté, les collaborations et la conciliation entre la vie professionnelle et privée. Les participants pourront exposer brièvement leurs expériences personnelles et offrir leurs observations et conseils dans le cadre d'une discussion libre en présence d'un modérateur.

3C-I2: Resources of Energy and Environmental Issues Ressources en énergie et enjeux environnementaux

Organizer and Chair / Responsable et président: Pritam Ranjan (Acadia University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: ECHA 1-498

Abstracts/Résumés

DEREK BINGHAM & SHIRIN GOLCHI (Simon Fraser University), **CHIPMAN HUGH** (Acadia University)

[Wednesday May 29 / mercredi 29 mai, 13:00]

Monotone Computer Model Emulation

Émulation d'un modèle informatique monotone

Computer simulators are commonly used to study physical phenomena. Often the computer model is computationally expensive, and a Gaussian process emulator is used in its place. In many applications, the simulator output is known to be monotone with respect to some of the inputs. Usefully using this prior knowledge is challenging for emulation because the function space becomes restricted to monotone sample paths that are also required to interpolate the outputs. We develop a Bayesian approach to Gaussian process modeling capable of incorporating monotonicity information. The performance of the proposed approach is demonstrated via simulated examples and a real application.

Les simulateurs informatiques sont couramment utilisés pour étudier des phénomènes physiques. Le modèle informatique est souvent coûteux en ressources informatiques et est remplacé par un émulateur de processus gaussien. Dans plusieurs applications, la sortie du simulateur est monotone par rapport à certains intrants. L'utilisation efficace de cette information connue a priori pose un défi pour l'émulation puisque l'espace de la fonction devient restreint aux chemins de l'échantillon monotone qui servent à interpoler les sorties. Nous développons une approche bayésienne à la modélisation d'un processus gaussien capable d'inclure l'information sur la monotonie. Nous démontrons la performance de l'approche proposée à l'aide d'exemples simulés et d'une application véritable.

VERA BULAEVSKAYA, SONIA WHARTON, MATTHEW SIMPSON, DON LUCAS & WAYNE MILLER (Lawrence Livermore National Laboratory)

[Wednesday May 29 / mercredi 29 mai, 13:30]

Statistical Methods for Wind Power Forecasting

Méthodes statistiques pour prédire l'énergie éolienne

Accurate forecasting of power output at a wind park is essential for successfully integrating wind energy into a power grid, but it is also challenging. While sophisticated computer models exist for forecasting wind speed, a key predictor of power, there is uncertainty in their outputs, stemming from incomplete knowledge about the initial conditions, the proper physical representation of the system, and the parameter values in a given representa-

Prédire avec précision la production électrique d'un parc éolien est essentiel pour intégrer avec succès l'énergie éolienne à un réseau électrique, mais cela représente également un défi. Alors qu'il existe des modèles informatiques sophistiqués pour prédire l'énergie éolienne, un indicateur clé de la production électrique, il y a des incertitudes quant aux extrants, provenant d'une connaissance incomplète des conditions initiales, de la représentation physique adéquate du système, ainsi que des valeurs des paramètres d'une

3C-I2: Resources of Energy and Environmental Issues Ressources en énergie et enjeux environnementaux

tion. Moreover, wind power also depends on other factors, all of which interact with wind speed and each other in a complex fashion. The talk will describe statistical methods to address these issues.

représentation donnée. De plus, l'énergie éolienne dépend également d'autres facteurs qui interagissent tous avec la vitesse du vent et entre eux de manière complexe. Cet exposé décrira des méthodes statistiques pour traiter ces questions.

RICHARD KARSTEN (Acadia University), **TRACEY KUTNEY** (Natural Resources Canada), **BRAIN POLAGYE** (University of Washington)

[Wednesday May 29 / mercredi 29 mai, 14:00]

Quantifying the Uncertainty in Predicting Tidal Power Resources

Quantifier l'incertitude dans la prévision des ressources d'énergie marémotrice

It has been estimated that the tidal currents in the Bay of Fundy could generate 2500 MW of power, more than enough to meet Nova Scotia's electricity needs. As the development of commercial tidal-turbine farms begins, it is important to quantify the uncertainty in such power estimates. In this talk, we examine how to combine the uncertainties in measuring and modelling the tidal flow, in making long term predictions of the flow, and in predicting turbine performance into an uncertainty for an annual energy estimate. We also examine methods to reduce this uncertainty.

Selon des estimations, les courants des marées de la Baie de Fundy pourraient générer 2 500 MW d'électricité, plus que nécessaire pour répondre aux besoins en électricité de la Nouvelle-Écosse. Alors que le développement de usines commerciales marémotrices débute, il est important de quantifier l'incertitude de telles estimations de production électrique. Dans cet exposé, nous examinons comment combiner les incertitudes dans la mesure et la modélisation du débit de marée dans les prévisions à long terme du débit et dans la prévision de la performance d'une turbine par rapport à l'incertitude de l'estimation annuelle d'énergie. Nous examinons également des méthodes pour réduire cette incertitude.

3C-I3: Statistical Methods for Modeling Biology and Ecology Processes Méthodes statistiques pour la modélisation des processus en biologie et en écologie

Chair/Présidente: Llwellyn Armstrong (Ducks Unlimited Canada)
Organizer/Responsable: Ying Zhang (Acadia University)

Session sponsored by the Committee on Women in Statistics

Séance parrainée par le Comité de promotion de la femme en statistique

Room/Salle: ECHA 2-150

Abstracts/Résumés

LIHUI LIU & SUBHASH LELE (University of Alberta)

[Wednesday May 29 / mercredi 29 mai, 13:00]

Statistical Analysis of Serial Dilution Assays Using Estimating Functions and Data Cloning

Analyse statistique d'essais de dilution en série par fonctions d'estimation et clonage de données

Serial dilution assay data are commonly analyzed using non-linear regression analysis of the standard samples alone. Recently hierarchical models are proposed to combine data across patients. These hierarchical models are usually analyzed using non-informative Bayesian approach. The non-linear regression method suffers from the instability of the solutions and the Bayesian approach suffers from difficult MCMC convergence. In this paper, we use data cloning and estimating functions to obtain maximum likelihood estimators and best predictors of the individual concentration while removing the instability of the solutions for non-linear models and non-convergence issues associated with the MCMC.

Les données d'essais de dilution en série sont généralement analysées par analyse de régression non linéaire des seuls échantillons types. Récemment, des modèles hiérarchiques ont été proposés pour combiner les données de plusieurs patients. Ces modèles hiérarchiques sont habituellement analysés par le biais d'une approche bayésienne non informative. La méthode de régression non linéaire souffre d'une instabilité de ses solutions et l'approche bayésienne de problèmes de convergence MCMC. Dans cet article, nous utilisons le clonage de données et les fonctions d'estimation pour obtenir des estimateurs du maximum de vraisemblance et les meilleurs indicateurs de la concentration individuelle tout en écartant l'instabilité des solutions des modèles non linéaires et les problèmes de non convergence associés au MCMC.

YING ZHANG, JING DING & PAUL CABILIO (Acadia University)

[Wednesday May 29 / mercredi 29 mai, 13:30]

Modelling Age-Specific Harvest Data

Modélisation des données sur les prises selon l'âge

Hunting and trapping are long standing traditions in Nova Scotia. Harvest of furbearers is regulated by the Nova Scotia Department of Natural Resources (DNR), and harvest data have been collected by the department since the 1980s. This wealth of data can be used to gain information about the populations of harvested species in the province. We evaluated several catch effort meth-

La chasse et la trappe sont des traditions de longue date en Nouvelle-Écosse. La collecte des animaux à fourrure est réglementée par le ministère des Ressources naturelles de la Nouvelle-Écosse (RNNE), qui collecte des données sur la récolte depuis les années 1980. Toutes ces données peuvent servir à mieux comprendre les populations d'espèces capturées dans la province. Nous avons évalué plusieurs méthodes de prise selon l'effort pour modéliser

3C-I3: Statistical Methods for Modeling Biology and Ecology Processes

Méthodes statistiques pour la modélisation des processus en biologie et en écologie

ods for modeling age-at-harvest data on furbearers, so as to provide wildlife managers with applicable population reconstruction approaches. In this talk, we will present the results of our comparisons and then explore possibilities for improvement.

les données sur les prises selon l'âge des animaux à fourrure, afin d'offrir aux gestionnaires de la faune diverses approches de reconstruction de la population. Dans cette présentation, nous donnerons les résultats de nos comparaisons avant d'explorer des possibilités d'amélioration.

LIWEN ZOU (Memorial University), **ERIC STONE & JEFFREY THORNE** (North Carolina State University)

[Wednesday May 29 / mercredi 29 mai, 14:00]

Incorporating Human Disease Data into Models of Molecular Evolution

Inclusion de données sur la maladie humaine dans les modèles d'évolution moléculaire

The purpose of our project is to use intraspecific and interspecific samples to make improvement on population genetic inference about the selective impact of molecular phenotypes. In the first stage, we constructed a mixture model derived from Tuffley-Steel model. Our model combines intraspecific data, which provides an opportunity to study ongoing or currently completed selective events, and interspecific data, which interspecific phenotypic variation mostly results from fixed genetic differences, in a sensible way. Using our model, we are able to analyze P53 gene evolution using interspecific sequence data and human disease data.

L'objectif de notre projet est d'utiliser des échantillons intraspécifiques et interspécifiques pour améliorer l'inférence génétique au niveau de la population concernant l'impact sélectif des phénotypes moléculaires. Dans un premier temps, nous avons construit un modèle de mélange dérivé du modèle Tuffley-Steel. Celui-ci combine de manière judicieuse des données intraspécifiques, qui permettent d'étudier des événements en cours ou récemment complétés, et interspécifiques, dont la variation phénotypique interspécifique résulte principalement de différences génétiques fixes. À l'aide de notre modèle, nous analysons l'évolution du gène P53 à l'aide de données de séquence interspécifiques et de données sur la maladie humaine.

3C-I4: Targeted Maximum Likelihood Estimation (TMLE) and Related Methods Estimation ciblée par maximum de vraisemblance (TMLE) et méthodes s'y rattachant

Organizer and Chair / Responsable et présidente: Mireille Schnitzer (McGill University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: ECHA L1-430

Abstracts/Résumés

SUSAN GRUBER (Harvard School of Public Health), **MARK VAN DER LAAN** (University of California, Berkeley)
[Wednesday May 29 / mercredi 29 mai, 13:00]

Triple-Robust Targeted Minimum Loss-Based Estimation
Estimation triple robuste ciblée de la perte minimale

An efficient double robust estimator solves the efficient influence curve estimating equation for the parameter of interest. New results show that solving one particular component of the efficient influence curve estimating equation guarantees consistent parameter estimation, even under dual misspecification of the outcome regression and censoring/treatment mechanism models. This result motivates a new triple-robust targeted minimum loss-based estimation (TMLE) procedure that ensures the TMLE solves the required D_{CAR} component of the efficient influence curve estimating equation. Simulation studies demonstrate the consistency properties under correct specification of at least one of the three components of the estimation procedure.

Un estimateur efficace robuste double résout l'équation estimante de la courbe d'influence efficace du paramètre d'intérêt. De nouveaux résultats montrent que résoudre une composante particulière de l'équation estimante de la courbe d'influence efficace garantit l'estimation cohérente de paramètres, même sous une mauvaise spécification combinée des modèles du résultat de la régression et du mécanisme de censure/traitement. Ce résultat motive une nouvelle procédure d'estimation triple robuste ciblée de la perte minimale qui assure la résolution de la composante D_{CAR} de l'équation estimante de la courbe d'influence efficace. Des études par simulation démontrent les propriétés cohérentes sous une spécification correcte d'au moins un des trois paramètres de la procédure d'estimation.

ORI STITELMAN, CLAUDIA PERLICH & BRIAN DALESSANDRO (Media6Degrees), **FOSTER PROVOST** (NYU - Stern School Of Business)

[Wednesday May 29 / mercredi 29 mai, 13:30]

The Challenges of Estimating the Causal Effect of Online Display Advertising
Les défis de l'estimation de l'effet causal dans l'affichage publicitaire en ligne

In general the metrics used to evaluate success in online display advertising have nothing to do with measuring effectiveness. However, the effectiveness of the advertising is ultimately what any marketer should care about. There are many challenges associated with estimating the effect of the advertising. Targeted Maximum Likelihood Estimation (TMLE) can address many of these chal-

En général, les métriques employées pour évaluer le succès de l'affichage publicitaire en ligne n'ont rien à voir avec les mesures d'efficacité. Cependant, l'efficacité de la publicité correspond à ce à quoi un distributeur devrait s'intéresser. Il existe plusieurs défis associés avec l'estimation de l'effet de la publicité. L'estimateur ciblé du maximum de vraisemblance (ECMV) peut traiter plusieurs de ces défis. Dans cet exposé, nous présenterons une analyse des campagnes

3C-I4: Targeted Maximum Likelihood Estimation (TMLE) and Related Methods Estimation ciblée par maximum de vraisemblance (TMLE) et méthodes s’y rattachant

lenges. In this talk we will present an analysis of online display advertising campaigns using TMLE. We will compare and contrast TMLE versus other methods. In addition, we will discuss pragmatic issues regarding estimating causal effects in the online display ecosystem.

d’affichage publicitaire en ligne utilisant l’ECMV, nous comparerons l’ECMV à d’autres méthodes. De plus, nous discuterons des considérations pragmatiques de l’estimation des effets causaux dans l’univers de la publicité en ligne.

MARK VAN DER LAAN (University of California, Berkeley)

[Wednesday May 29 / mercredi 29 mai, 14:00]

Targeted Minimum Loss-Based Estimation

Estimation ciblée de la perte minimale

I will review Targeted Minimum Loss-Based Estimation (TMLE) that provides a template for the construction of semiparametric efficient substitution estimators of low dimensional estimands in large semiparametric models for which maximum likelihood estimation breaks down due to the curse of dimensionality. We will review various examples, including estimation problems in causal inference. We also review a formal theorem that establish asymptotic normality and efficiency of TMLE, whose condition on the remainder demonstrate that the incorporation of super learning (an ensemble algorithm using cross-validation) is fundamentally important.

Je passerai en revue l’estimation ciblée de la perte minimale qui fournit un cadre pour la construction d’estimateurs efficaces de substitution semi-paramétriques en faible dimension dans de grands modèles semi-paramétriques pour lesquels l’estimation du maximum de vraisemblance s’effondre à cause de la malédiction multidimensionnelle. Nous passerons en revue différents exemples, incluant les problèmes d’estimation en inférence causale. Nous passons également en revue un théorème formel qui établit la normalité asymptotique et l’efficacité de l’estimation ciblée de la perte minimale dont la condition sur le reste démontre que l’intégration de l’apprentissage super (un ensemble algorithmique employant une validation croisée) est fondamentalement importante.

3C-C1: Probability in Actuarial Science and Finance Probabilité en actuariat et en finance

Chair/Président: Don McLeish (University of Waterloo)

Room/Salle: EdSouth 228

Abstracts/Résumés

ALEXANDER MELNIKOV & HAO LI (University of Alberta)

[Wednesday May 29 / mercredi 29 mai, 13:00]

Orthogonal Polynomials and their Applications in Financial and Actuarial Modeling
Polynômes orthogonaux et leurs applications en modélisation financière et actuarielle

The primary goal of the talk is to present a new approach to financial and actuarial modeling which is based on polynomial extensions of probability distributions for financial assets and insurance claims. Besides improvements in modeling of asset returns and insurance claim sizes a better option pricing as well as risk measures calculations become possible using this approach. Theoretical findings are supported with numerical examples and illustrations exploiting financial and actuarial statistical data.

L'objectif principal de cet exposé est de présenter une nouvelle approche de modélisation financière et actuarielle basée sur les extensions polynomiales de distributions de probabilité pour les actifs financiers et les réclamations d'assurance. Outre les améliorations en matière de modélisation du rendement des actifs et de la taille des réclamations d'assurance, une meilleure tarification des options ainsi que des calculs de mesures du risque deviennent possibles grâce à cette approche. Nous appuyons nos résultats théoriques par des exemples numériques et des illustrations exploitant des données financières et actuarielles.

SHANOJA NAIK & BOVAS ABRAHAM (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 13:15]

The Fractional Diffusion Equation and a New Distribution with Heavy Tails
Équation de diffusion fractionnelle et nouvelle distribution à ailes lourdes

In the classical diffusion process, particles in a physical system spread according to the Gaussian process. Using fractional derivative on the partial differential equation, the solution turns out to be non Gaussian and we demonstrated that the invariance solution is a particular form of Wright function. We refer to the density function as the Wright distribution. It can also be thought of as a convolution of two random variables having inverse Gaussian and gamma distributions. The Wright distribution has heavy tails with other suitable properties to model financial data. We studied some properties and developed an estimation procedure.

Dans le processus de diffusion classique, les particules dans un système physique se propagent suivant un processus gaussien. En utilisant une dérivée fractionnelle de l'équation différentielle partielle, la solution s'avère non gaussienne et nous montrons que la solution invariante est une forme particulière de la fonction de Wright. Nous dirons la distribution de Wright pour référer à la densité. Elle peut également être considérée comme une convolution de deux variables aléatoires de distributions gaussienne inverse et gamma. La distribution de Wright présente des ailes lourdes et d'autres propriétés appropriées pour modéliser des données financières. Nous étudions quelques propriétés et présentons une méthode d'estimation.

RUODU WANG (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 13:30]

Completely Mixable Distributions
Distributions complètement mélangeables

3C-C1: Probability in Actuarial Science and Finance Probabilité en actuariat et en finance

A completely mixable distribution, defined as the marginal distribution of identically distributed dependent random variables having a constant sum, was recently introduced in 2011. In this talk, I will present the theory and applications of the family of complete mixable distributions. This family of distributions is relevant to Actuarial Science and Quantitative Risk Management; it plays an important role in solving the problems of risk aggregation with unknown dependence structure.

Une distribution complètement mélangeable, définie comme la distribution marginale de variables aléatoires dépendantes et identiquement distribuées ayant une somme constante, a été proposée en 2011. Dans cet exposé, je présente la théorie et les applications de la famille des distributions complètement mélangeables. Cette famille de distributions est pertinente en actuariat et en gestion quantitative du risque ; elle joue un rôle important dans la résolution de problèmes d'agrégation des risques avec une structure de dépendance inconnue.

TAEHAN BAE (University of Regina), **IAN ISCOE** (IBM), **CHANGKI KIM** (Korea University)

[Wednesday May 29 / mercredi 29 mai, 13:45]

Valuing Retail Credit Tranches under Conditional Dependence

Valorisation de tranches de crédit de détail sous l'hypothèse de dépendance conditionnelle

Under the single-factor structural credit risk modeling framework, we consider the double mixtures to model a general dependence structure beyond the typical conditional independence assumption between the entities in a homogeneous credit pool. For a large portfolio, the fair spread of a structured retail credit tranche is expressed in terms of the sums of single integrals which can be easily computed by a numerical method. We discuss the behaviors of tranche spreads by the level of dependence under four double mixture models, Gaussian-ICM, t -ICM, Gaussian-Beta and t -Beta, and calibrate these models to market data.

Selon le cadre structurel à facteur unique de modélisation du risque de crédit, nous considérons les mélanges doubles pour modéliser une structure de dépendance générale au-delà de l'hypothèse d'indépendance conditionnelle typique entre les entités d'un pool de crédit homogène. Pour un grand portefeuille, la répartition équitable d'une tranche structurée de crédit de détail est exprimée par la somme des intégrales simples qui peuvent être facilement calculées par une méthode numérique. Nous discutons le comportement des écarts de tranches selon le niveau de dépendance à l'aide de quatre modèles de mélanges doubles, à savoir ICM gaussien, t -ICM, Beta-gaussien et t -Beta, et nous calibrons ces modèles aux données du marché.

3C-C2: Variable Selection, Clustering and Classification Sélection de variables, partitionnement et classification

Chair/Président: Jean-François Plante (HEC Montréal)

Room/Salle: EdSouth 206

Abstracts/Résumés

FOLLY ADJOGOU & ALEJANDRO MURUA (Université de Montréal)

[Wednesday May 29 / mercredi 29 mai, 13:00]

Functional Model-Based Clustering for Longitudinal Data

Analyse de données longitudinales selon une approche fonctionnelle de la classification non supervisée

We develop a flexible model for the analysis and clustering of complete or sparse longitudinal data. The model combines functional principal component analysis (based on splines) and model-based clustering. The main data groups arise from clusters in the space of spline coefficients. The clusters are modeled by a mixture of Student's t-distributions with unknown degrees of freedom. We develop an approximation of the marginal log-likelihood (MLL) for model selection which compares favorably with other popular criteria (AIC and BIC). We also consider an extension of our model to curves in multiple dimensions. We will present applications to gene expression and clinical data.

Nous développons un modèle flexible pour l'analyse et la classification de données longitudinales (balancées ou non). Le modèle combine l'analyse fonctionnelle en composantes principales et le regroupement en classes qui repose sur l'espace des coefficients dans la base des splines et un modèle de mélange de distributions de student de degrés de liberté inconnus. Nous développons une approximation de la log-vraisemblance marginale (MLL) pour la sélection de modèles qui se compare favorablement aux critères usuels (AIC et BIC). Nous considérons également une extension du modèle aux courbes multidimensionnelles. Nos applications principales concernent les données d'expression génétique et les essais cliniques.

ANNE-SOPHIE CHAREST (Université Laval), **RUSSELL STEELE** (McGill University)

[Wednesday May 29 / mercredi 29 mai, 13:15]

Imputation for Nonresponse via Clustering of High-Dimensional Categorical Variables

Imputation pour la non-réponse via un classement basé sur plusieurs variables catégoriques

When using multiple imputation for inference from incomplete data, finding an appropriate imputation model may be challenging. For example, fitting a saturated general location model, as suggested in Schafer (1997), rapidly becomes infeasible as the number of categorical variables increases. We propose to first reduce the dimensionality of the categorical data by clustering the observations using the nonparametric HD-algorithm of Zhang et al. (2006). The cluster indicators can then be used within a general location model to impute missing values of the continuous random variables. We describe the proposed methodology, discuss its properties, and present an application to scleroderma data.

Choisir un modèle approprié pour l'imputation multiple de données manquantes peut être difficile. Par exemple, ajuster un modèle de position général saturé, tel que suggéré par Schafer (1997), devient rapidement impossible quand le nombre de variables catégoriques croît. Nous suggérons de réduire d'abord la dimensionnalité des données catégoriques en classifiant les observations à l'aide de l'algorithme non-paramétrique de Zhang et al. (2006). Les variables indicatrices de l'appartenance aux groupes sont ensuite utilisées dans le modèle de position général pour imputer les variables continues manquantes. Nous décrivons la méthode proposée et ses propriétés et présenterons un exemple concret concernant la sclérodémie.

3C-C2: Variable Selection, Clustering and Classification Sélection de variables, partitionnement et classification

KEVIN KEEN (University of British Columbia), **WILLIAM PETRCICH** (University of Northern British Columbia), **STEPHAN VAN EEDEN & JAMES DUNNE** (University of British Columbia)

[Wednesday May 29 / mercredi 29 mai, 13:30]

Bumping Bagged Classification Trees and Random Forests

Bumping d'arbres de décision et de forêts aléatoires agrégés par bootstrap

Two algorithms are proposed that incorporate bumping and bagging, one with classification trees and the other with random forests, to find the subset of all possible initial choices of explanatory variables with the smallest possible misclassification rate. These algorithms can be used as a check of the robustness of the original selection of potential explanatory variables. Two examples are given involving the classification of interstitial lung disease from twelve biomarkers in patients with the diffuse and limited forms of scleroderma, a very rare rheumatic disease of unknown etiology.

On propose deux algorithmes qui incorporent bumping et bagging, l'un avec des arbres de décision et l'autre avec les forêts d'arbres décisionnels (forêts aléatoires), pour trouver le sous-ensemble de tous les choix possibles initiales des variables explicatives avec le plus petit taux d'erreur possible. Ces algorithmes peuvent être utilisés pour le contrôle de robustesse de la sélection originale des variables explicatives. On donne deux exemples impliquant la classification de la maladie pulmonaire interstitielle avec douze biomarqueurs des patients atteints des formes diffuse et limitée de la sclérodémie : une maladie très rare, une maladie rhumatismale, et d'étiologie inconnue.

MACIAK MATUS & IVAN MIZERA (University of Alberta)

[Wednesday May 29 / mercredi 29 mai, 13:45]

Change in Behaviour with Lasso Methods

Changement de comportement avec les méthodes Lasso

In modern regression approaches, one can also deal with situations when the underlying dependence structure of interest suddenly changes its functional form. In such cases, we refer to regression models with change-points. But for common estimation techniques used in statistics to handle such situations, one has to know the location of changes in advance. This is however, not possible in many practical applications. We therefore, discuss an automatic and fully data-driven estimation method based on group lasso, that can easily handle both, flexibility in the unknown regression function and an absence of any prior information regarding the change-point occurrences.

Dans les approches de régression moderne, on peut traiter des situations seulement lorsque la structure de dépendance sous-jacente d'intérêt change soudainement sa forme fonctionnelle. Dans ces cas, nous nous référons aux modèles de régression avec changement. Cependant, pour les techniques d'estimation courantes utilisées dans les statistiques pour traiter ce genre de situations, on doit connaître à l'avance l'endroit des changements, ce qui n'est pas possible dans de nombreuses applications pratiques. Nous allons ainsi présenter une méthode d'estimations automatique et complètement basées sur des données à l'aide de la méthode Lasso en groupe qui peut facilement traiter autant la flexibilité dans la fonction de régression inconnue que l'absence d'information a priori concernant les occurrences du point de changement.

PAULA MURRAY, RICHARD BROWNE & PAUL MCNICHOLAS (University of Guelph)

[Wednesday May 29 / mercredi 29 mai, 14:00]

Model-Based Clustering Using Mixtures of Skew-t Factor Analyzers

Modèle de regroupement en classes au moyen de mélanges d'analyseurs de facteurs t asymétriques

A mixture of skew-t factor analyzers is introduced for the clustering of high-dimensional data such as

Nous présentons un mélange d'analyseurs de facteur t asymétriques permettant de regrouper en classes de données

3C-C2: Variable Selection, Clustering and Classification Sélection de variables, partitionnement et classification

bioinformatics data. Notably, we make use of a version of the skew-t distribution which has not previously appeared in the mixture-modelling literature. The models are applied to simulated data as well as two high-dimensional microarray gene expression data sets, obtaining clustering results which are equivalent or superior to those of established clustering methods.

de grande dimension, comme les données bioinformatiques. Il faut remarquer que nous utilisons une version de la distribution t asymétrique qui n'apparait pas auparavant dans la littérature de modèles de mélanges. Les modèles sont appliqués à des données simulées ainsi qu'à deux jeux de données de grande dimension d'expression génétique de micropuces, ce qui permet d'obtenir des résultats de regroupement qui sont équivalents ou supérieurs à ceux des méthodes de regroupement bien connues.

MONICA H.T. WONG & PAUL MCNICHOLAS (University of Guelph)

[Wednesday May 29 / mercredi 29 mai, 14:15]

Parsimonious Biclustering of Gene Expression Data with Applications in Nutrigenomics

Classification double parcimonieuse de données d'expression génique avec 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.

Nous proposons une famille de modèles de mélange gaussiens pour la classification double de données d'expression génique de grande dimension. Un algorithme de classification double a déjà été présenté dans le cadre duquel un modèle de mélange à covariance d'analyse factorielle était généralisé pour inclure une matrice de loadings binaire et stochastique. Cette forme particulière de la matrice des loadings conduit à une matrice de covariance diagonale par blocs. Notre famille est définie en imposant une structure d'analyse factorielle d'abord, puis en y ajoutant des contraintes. L'estimation des paramètres est effectuée à l'aide d'une variante de l'algorithme d'espérance-maximisation et la sélection de modèle est basée sur le critère d'information bayésien.

3D-A1: CJS Award Address
Allocution du récipiendaire du Prix de la RCS

Chair/Président: David A. Stephens (McGill University)
Organizer/Responsable: Denis Larocque (HEC Montréal)

Room/Salle: EdSouth 254

Abstract/Résumé

JING QIN (National Institute of Allergy and Infectious Diseases), **GUAN ZHONG** (Indiana University South Bend), **BIAO ZHANG** (University of Toledo)

[Wednesday May 29 / mercredi 29 mai, 15:00]

Information Borrowing Methods for Covariate-Adjusted ROC Curve

Méthodes d'extraction d'information pour courbes ROC ajustées à l'effet de covariables

In medical diagnostic testing problems, the covariate adjusted receiver operating characteristic (ROC) curves have been discussed recently for achieving the best separation between case and control. Due to various constraints, the sample sizes for some covariate values are not large enough to support reliable direct estimations of ROCs for all the underlying covariates of interest. We develop statistical methods to effectively utilize the information provided by the data using the semi-parametric exponential tilting models. In these models, the density functions from different covariate levels share a common baseline density and the parameters in the exponential tilting component reflect the difference among covariates. The new covariate adjusted ROC is much smoother and more efficient than the nonparametric counterpart. A simulation study and a real data application are reported.

Dans les problèmes de tests diagnostics médicaux, les courbes ROC ajustées à l'effet de covariables ont été discutées récemment à l'effet qu'elles permettent la meilleure discrimination entre les groupes traité et contrôle. À cause de contraintes diverses, les tailles d'échantillons pour certaines covariables ne sont toutefois pas assez grandes pour assurer une estimation directe fiable des courbes ROC pour toutes les covariables d'intérêt. Nous développons ici des méthodes statistiques afin d'utiliser de façon efficace l'information fournie par les données en utilisant des modèles de «tilting» exponentiel semi-paramétriques. Dans ces modèles, les densités de probabilité des différents niveaux des covariables partagent une densité de référence commune où les paramètres reflètent les différences parmi les covariables. La nouvelle courbe ROC ainsi proposée est beaucoup plus lisse et plus efficace que sa version non-paramétrique. Les résultats d'une étude de simulation ainsi qu'une application sur de vraies données seront présentés.

3D-A2: Pierre Robillard Award Address
Allocution du récipiendaire du Prix Pierre-Robillard

Organizer and Chair / Responsable et présidente: Johanna Nešlehová (McGill University)

Room/Salle: MedSci 231

Abstract/Résumé

LUKE BORNN (Harvard University)

[Wednesday May 29 / mercredi 29 mai, 15:00]

Nonstationary Modeling Through Dimension Expansion

Modélisation non stationnaire par l'agrandissement de la dimension

In this talk, we propose a novel approach to modeling nonstationary spatial fields. The proposed method works by expanding the geographic plane over which these processes evolve into higher dimensional spaces, transforming and clarifying complex patterns in the physical plane. By combining aspects of multi-dimensional scaling, group lasso, and latent variable models, a dimensionally sparse projection is found in which the originally nonstationary field exhibits stationarity. Following a comparison with existing methods in a simulated environment, dimension expansion is studied on a classic test-bed data set historically used to study nonstationary models. Following this, we explore the use of dimension expansion in modeling air pollution in the United Kingdom, a process known to be strongly influenced by rural/urban effects, amongst others, which gives rise to a nonstationary field.

Dans cet exposé, on propose une nouvelle approche pour la modélisation de champs spatiaux non stationnaires. La méthode proposée consiste à agrandir le plan géographique sur lequel ces processus évoluent vers des espaces dont la dimension est plus élevée, transformant et exhibant ainsi des comportements complexes dans le plan physique. En combinant des éléments de «scaling» multidimensionnel, de lasso groupé et de modèles à variables latentes, une projection est ainsi découverte pour laquelle le champ initialement non stationnaire se comporte de façon stationnaire. Faisant suite à une comparaison avec des méthodes existantes dans un environnement simulé, l'agrandissement de la dimension proposée est étudié sur un jeu de données classique utilisé historiquement pour investiguer des modèles non stationnaires. On explore ensuite l'usage de l'agrandissement de la dimension pour modéliser la pollution de l'air au Royaume-Uni, un processus reconnu comme étant fortement influencé par des effets ruraux et urbains, entre autres, ce qui contribue à induire un champ non stationnaire.

3D-C1: Analysis of Longitudinal Data Analyse de données longitudinales

Chair/Président: Joel Dubin (University of Waterloo)

Room/Salle: EdSouth 276

Abstracts/Résumés

CYNTHA STRUTHERS & DON MCLEISH (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 15:00]

A Multivariate Diffusion Model for AIDS Data

Modèle de diffusion multivarié pour des données sur le SIDA

Longitudinal studies are usually modelled with some form of regression model. When we study a process over time that one expects to be stationary in the absence of covariate changes, it seems more natural to use a potentially stationary stochastic process. We model continuous-time multivariate longitudinal data from the Multicenter AIDS Cohort Study using a multivariate diffusion model with diffusion parameters which are functions of time-varying covariates. If the covariates were fixed, the model would be a correlated multivariate Ornstein-Uhlenbeck process. The model is used to evaluate alternatives to CD4 counts as a surrogates for the health of AIDS patients.

Les études longitudinales sont généralement modélisées à l'aide d'une forme de modèle de régression. Lors de l'étude d'un processus dans le temps qu'on présume stationnaire en l'absence de changements dans les covariables, il semble plus naturel d'utiliser un processus stochastique potentiellement stationnaire. Nous modélisons des données longitudinales multivariées en temps continu de la Multicenter AIDS Cohort Study à l'aide d'un modèle de diffusion multivarié dont les paramètres de diffusion sont fonction de covariables variant dans le temps. Si les covariables étaient fixes, le modèle serait un processus corrélé multivarié d'Ornstein-Uhlenbeck. Le modèle sert à évaluer les alternatives aux dénombrements des CD4 comme substituts pour la santé des patients atteints du SIDA.

MIREILLE SCHNITZER, JUDITH J. LOK & RONALD J. BOSCH (Harvard School of Public Health)

[Wednesday May 29 / mercredi 29 mai, 15:15]

Determining the Predictors for Negative HIV Outcomes under a Suppressive ART Regime

Détermination des prédicteurs de résultats négatifs du VIH sous un régime de traitement antirétroviral suppressif

The clinical gold standard in HIV treatment is to continuously administer Anti-Retroviral Treatment (ART) after initiation. Clinical interest lies in improving patient outcomes under continuous treatment and successful viral suppression. We performed an analysis of the ALLRT longitudinal database to determine the predictors of long-term negative outcomes under the counterfactual setting of continuous treatment (and virologic suppression), and no censoring. A marginal structural logistic model was used to model AIDS and non-AIDS defining events, including mortality. We will present data analyses using two doubly-robust, semiparametrically efficient methods, namely the

La norme clinique par excellence dans le traitement du VIH consiste à administrer de façon continue des traitements antirétroviraux (ARV) après l'initiation. L'intérêt clinique réside dans l'amélioration des résultats des patients sous traitement continu et du taux de succès de la suppression virale. Nous avons effectué une analyse de la base de données longitudinales ALLRT afin de déterminer les prédicteurs de résultats négatifs à long terme dans le cadre contrefactuel d'un traitement continu (et d'une suppression virologique) sans censure. Un modèle structurel marginal logistique a été utilisé pour modéliser les événements déterminants liés au SIDA (ou à son absence), y compris la mortalité. Nous présentons des analyses de données à l'aide de deux méthodes semi-paramétriques efficaces et double-

3D-C1: Analysis of Longitudinal Data Analyse de données longitudinales

Bang and Robins estimator and Targeted Maximum Likelihood Estimation.

ment robustes, à savoir l'estimateur de Bang et Robins et l'estimation au maximum de vraisemblance ciblé.

ELEANOR M PULLENAYEGUM (McMaster University)

[Wednesday May 29 / mercredi 29 mai, 15:30]

Analysis of Longitudinal Data Subject to Irregular Follow-Up: Diagnosing and Correcting for Small-Sample Bias in Inverse-Intensity Weighted Estimating Equations

Analyse de données longitudinales sujettes à un suivi irrégulier: diagnostic et correction de biais dans des équations d'estimation pondérées par l'intensité inverse pour de petits échantillons

Observational longitudinal data are often subject to irregular follow-up, with no two individuals visiting at the same set of times. The most popular method for handling this, is inverse-intensity weighting, where the usual Generalized Estimating Equations are weighted by the inverse of the visit intensity. Whilst this yields consistent estimators for regression coefficients, simulation evidence shows that there can be important bias even with moderate sample sizes, due to the need to estimate visit intensities. This work explores higher-order expansions to diagnose and correct for this bias, and contrasts easily computed corrections with more computationally intensive bootstrap methods.

Les données longitudinales d'observation sont souvent sujettes à un suivi irrégulier, où les individus effectuent chacun leurs visites à des temps différentes. La méthode la plus populaire pour traiter ces données est une pondération par intensité inverse, selon laquelle les équations d'estimation généralisées sont pondérées par l'inverse de l'intensité des visites. Même si cette méthode produit des estimateurs convergents des coefficients de régression, des simulations montrent que d'importants biais peuvent survenir, y compris pour des échantillons de taille moyenne, découlant de la nécessité d'estimer l'intensité des visites. Cet article explore les expansions d'ordre supérieur visant à diagnostiquer et à corriger ces biais et compare des corrections faites à l'aide de calculs simples avec des méthodes bootstrap nécessitant des calculs intensifs.

JEMILA HAMID & WEILIANG HUANG (McMaster University)

[Wednesday May 29 / mercredi 29 mai, 15:45]

Residuals in the Growth Curve Model and their Application in the Analysis of Longitudinal Data

Les résidus dans le modèle de courbe de croissance et application de ceux-ci dans l'analyse de données longitudinales

In univariate models, residuals are relatively simple to examine and they have been studied extensively. However, in multivariate models and in the analysis of longitudinal data, residuals have complex structure, are correlated and do not often follow the normal distribution. In this study, we consider decomposed residuals defined using the growth curve model. We also transformed these residuals to uncorrelated residuals using Cholesky decomposition of the covariance matrix. Extensive simulations demonstrate that the decomposed and transformed residuals perform well for assessing model fit as well as checking the normality assumption. We illustrate performance of our method using a real data set.

Dans les modèles univariés, les résidus sont relativement simples à examiner, et ont été étudiés à grande échelle. Cependant, dans des modèles multivariés et dans l'analyse de données longitudinales, les résidus ont une structure complexe, sont corrélés et ne suivent pas souvent la distribution normale. Dans cette étude, nous considérons des résidus décomposés définis au moyen du modèle de courbe de croissance. Nous transformons également ces résidus en résidus non corrélés grâce à la décomposition de Cholesky de la matrice de covariance. Des simulations approfondies démontrent que les résidus décomposés et transformés sont efficaces pour analyser l'ajustement du modèle, ce qui permet de vérifier l'hypothèse de normalité. Nous illustrons l'efficacité de notre méthode avec un jeu de données réelles.

3D-C2: Business and Industrial Statistics II Statistique industrielle et de gestion II

Chair/Président: Fernando Camacho (Damos)

Room/Salle: EdSouth 221

Abstracts/Résumés

AMBER CORKUM, YING ZHANG & PAUL CABILIO (Acadia University)

[Wednesday May 29 / mercredi 29 mai, 15:00]

Confidence Interval Construction for Short Time Series with Applications in Administrative Health Databases

Construction de l'intervalle de confiance pour les séries chronologiques courtes et applications dans des bases de données administratives dans le domaine de la santé

In health policy research, we are interested in determining if there is a significant difference in the behaviour of prescribing physicians pre- and post-policy change. In this talk we use properties of the Wilcoxon Signed Rank Statistic (WSRS) to construct confidence intervals for this change. This involves computing the variance of the WSRS for time series data. We can derive this exactly when the error distribution is normal and dependence structure is known. For the more realistic case, where the error distribution or dependence structure is unknown, we demonstrate how to estimate the variance by block and sieve bootstrap methods.

En recherche sur la politique en matière de santé, nous sommes intéressés à déterminer si le comportement des médecins prescripteurs était très différents avant et après le changement de la politique. Dans cet article, nous utilisons les propriétés de la statistique de Wilcoxon pour observations appariées afin de construire des intervalles de confiance pour ce changement. Ceci implique de calculer la variance de la statistique de Wilcoxon pour observations appariées de données de séries chronologiques courtes. Nous pouvons obtenir la variance exacte lorsque la distribution de l'erreur est normale et que la structure de la dépendance est connue. Dans les cas les plus réalistes, lorsque la distribution de l'erreur ou lorsque la structure de la dépendance est inconnue, nous démontrons comment estimer la variance par les méthodes bootstrap par blocs et tamis.

VEERESH GADAG & ASOKAN MULAYATH VARIYATH (Memorial University), **UDITHA BALASOORIYA** (Nanyang Technical University)

[Wednesday May 29 / mercredi 29 mai, 15:15]

On Tests for Multiple Contaminants in Shocks Model and Labelled Slippage Models

Tests pour des modèles de chocs et de glissement marqué avec plusieurs contaminants

We consider life testing situation where the underlying lifelength distribution is Weibull with known shape parameter. We derive likelihood ratio test for testing multiple contaminants where contaminants may have occurred (i) due the shock(s) received by the system at known points in time and (ii) as per labeled slippage model. Through Monte-Carlo studies we obtain the respective cutoff points for given values of shape parameter. After power performance comparisons we show that of the respective likelihood ratio tests statistics derived earlier

Nous considérons un contexte de test de la durée de vie dans lequel la loi de durée de vie sous-jacente est celle de Weibull avec des paramètres de forme connus. Nous dérivons le test de rapport de vraisemblances pour examiner les nombreux contaminants dans lesquels les contaminants peuvent survenir (i) en raison de choc(s) reçu(s) par le système à des instants connus dans le temps, (ii) en fonction du modèle de glissement marqué. Les études de Monte Carlo ont permis d'obtenir des limites d'inclusion respectives pour les valeurs données des paramètres de forme. Après avoir effectué des comparaisons de puissance, nous montrons

3D-C2: Business and Industrial Statistics II Statistique industrielle et de gestion II

for exponential lifelength distribution are robust for different values of the shape parameter of Weibull distribution.

que les statistiques respectives des tests de rapport des vraisemblances dérivées précédemment pour la distribution de durée de vie exponentielle, sont robustes pour différentes valeurs du paramètre de forme de la loi de Weibull.

TOM LOUGHIN (Simon Fraser University Surrey), **ANDREW HENREY** (Simon Fraser University)

[Wednesday May 29 / mercredi 29 mai, 15:30]

Joint Model Selection for Location and Dispersion Effects in Un-Replicated Factorial Experiments

Sélection de modèle conjoint pour les effets de localisation et de dispersion dans des expériences factorielles non répétées

Most procedures that have been proposed to identify dispersion effects in un-replicated factorial designs assume that location effects have been identified correctly. Incorrect identification of location effects may impair subsequent identification of dispersion effects. We develop a model for joint identification of location and dispersion effects that can reliably identify active effects of both types. The joint model is estimated using maximum likelihood and hence effect selection is done using information-based criteria. An exhaustive search through a limited version of the space of possible models is conducted. Both a single-model output and model averaging are considered.

La plupart des procédures qui ont été proposées pour déterminer les effets de dispersion dans des plans factoriels non répétées suppose que les effets de localisation ont été identifiés correctement. Une identification incorrecte des effets de localisation peut altérer l'identification ultérieure des effets de dispersion. Nous avons développé un modèle conjoint d'identification des effets de localisation et de dispersion qui peut distinguer de façon fiable les effets actifs des deux types d'effet. Le modèle conjoint est estimé au moyen de la méthode du maximum de vraisemblance. Ainsi, la sélection de l'effet est effectuée au moyen de critères axés sur l'information. Une recherche exhaustive dans une version limitée de l'espace des modèles possibles est réalisée. Nous considérons une seule sortie du modèle simple et le moyennage de modèles.

XIANG YAN & MEI LING HUANG (Brock University)

[Wednesday May 29 / mercredi 29 mai, 15:45]

On Confidence Intervals of High Quantiles

Sur les intervalles de confiance pour quantiles supérieures

The high quantile estimation of heavy tailed distributions has many important applications. There are theoretical difficulties in studying the heavy tailed distributions since they often have infinite moments. There are also bias issues with the existing confidence intervals of high quantiles methods. This paper proposes a new estimation method on the confidence interval of high quantiles to avoid these difficulties and to improve the efficiencies and reduce bias. Comparisons of efficiencies and biases of new estimators relative to existing estimators are studied. The results of simulations confirm the theoretical results. Finally, we investigate a real-world example of applications.

L'estimation des quantiles supérieurs de distributions à ailes relevées a de nombreuses applications importantes. L'étude des distributions à ailes relevées présente des difficultés théoriques, car celles-ci ont souvent des moments infinis. Il y a également les problèmes de biais avec les intervalles de confiance de quantiles élevés. Cet article présente une nouvelle méthode d'estimation sur l'intervalle de confiance de quantiles élevés pour éviter ces difficultés, en améliorer l'efficacité et réduire le biais. Nous étudions des comparaisons d'efficacité et de biais des nouveaux estimateurs relatifs aux estimateurs existants. Les résultats des simulations confirment les résultats théoriques. Finalement, nous examinons un exemple d'applications.

3D-C3: Competing Risks and Disease Progression Risques concurrents et progression des maladies

Chair/Président: Thierry Duchesne (Université Laval)

Room/Salle: EdSouth 228

Abstracts/Résumés

YUN-HEE CHOI (Western University)

[Wednesday May 29 / mercredi 29 mai, 15:00]

On Combining Population-Based and Clinic-Based Time-to-Event Family Data in Disease Risk Estimation

Sur la combinaison de données de temps avant événement familiales pour les populations cliniques et temporelles dans l'estimation du risque de maladies

Disease family registries often use different sampling designs to ascertain families in a study. In the analysis of combined family data arising from different designs, frailty-based models are developed for estimating disease risks associated with mutated genes. Multiple frailty distributions are employed to account for different familial correlation structures induced by different study designs, along with appropriate correction for ascertainment. A simulation study is conducted to examine the performance of our proposed method. Lynch Syndrome families by population-based and clinic-based designs were obtained from the NCI Cancer Family Registries and their age-specific penetrance and relative risk associated with mutated genes were estimated.

Les registres de maladies familiales utilisent souvent différents plans d'échantillonnage pour faire une estimation des familles dans une étude. Dans l'analyse des données familiales combinées qui proviennent de différents plans, on développe des modèles basés sur la fragilité pour estimer les risques de maladie associés aux gènes modifiés. Plusieurs distributions de fragilité sont employées pour considérer les différentes structures de corrélations familiales induites par différents plans d'études, ainsi que la correction pour les estimations. Une étude de simulation est réalisée pour examiner l'efficacité de la méthode que nous proposons. Les familles de syndrome de Lynch par plans basés sur la population et cliniques ont été obtenues à partir des registres du cancer familial de l'Institut national du cancer. On a également estimé la pénétrance spécifique à l'âge de ces familles et le risque relatif associé aux gènes modifiés.

UTKARSH DANG & PAUL MCNICHOLAS (University of Guelph)

[Wednesday May 29 / mercredi 29 mai, 15:15]

Cluster-Weighted Modeling for Time-to-Event Data

La modélisation par pondération de grappes pour des données de temps avant événement

We present a mixture of accelerated failure time models for a competing risks situation in a cluster-weighted modeling (CWM) framework. CWM models the joint probability of data arising from a population of sub-populations using combinations of local models. Both reliability and survival models analyze data on time to some event of interest in the engineering and medical fields. Here, we present a novel approach to group mixture estimation and classification. Finally, we present our results on some simulated and real censored data

Nous présentons un mélange de modèles de temps de défaillance accélérée pour une situation de risques concurrents dans un cadre de modélisation par pondération de grappes (MPG). Cette dernière modélise la probabilité conjointe de données provenant d'une population de sous-populations au moyen de combinaisons de modèles locaux. Les deux modèles de fiabilité et de survie analysent les données sur le temps avant certains événements d'intérêt dans les domaines respectifs d'ingénierie et médical. Dans cet article, nous présentons une nouvelle approche d'estimation de mélange fondée sur les groupes ainsi qu'une classifi-

3D-C3: Competing Risks and Disease Progression Risques concurrents et progression des maladies

where the time and cause of failure was recorded only on some of the observations.

Finally, we present our results on simulated and real censored data for which the time and cause of failure were not recorded for some observations.

YAN YAN WU & LAURENT BRIOLLAIS (Samuel Lunenfeld Research Institute, Mount Sinai Hospital), **AUGUSTINE WONG** (York University)

[Wednesday May 29 / mercredi 29 mai, 15:30]

An Accurate Method for Testing Genetic Heterogeneity in Families from Studies of Rare Genetic Variants

Une méthode précise pour tester l'hétérogénéité génétique dans des familles à partir d'études de variantes génétiques rares

Rare genetic variants contribute to complex disease risks. The analysis of genetic heterogeneity in families remains challenging. Mixed-effects models provide a powerful tool for analyzing the effect of rare variants after accounting for the familial dependencies. However, the classical methods for testing heterogeneity in families are often biased. Here, we propose a higher-order likelihood based method that dramatically reduces the bias. The proposed method is consistently accurate when the sample size is small. Extensive simulations were carried out to illustrate the accuracy of our proposed method.

Rare genetic variants favor complex diseases. The analysis of genetic heterogeneity in families remains a big challenge. Mixed-effects models are a powerful tool to analyze the effect of rare variants after accounting for familial dependencies. However, classical methods for testing heterogeneity in families are often biased. We propose here a higher-order likelihood based method that significantly reduces the bias. This method is accurate when the sample size is small. Extensive simulations were carried out to illustrate the accuracy of the proposed method.

YING WU & RICHARD COOK (University of Waterloo)

[Wednesday May 29 / mercredi 29 mai, 15:45]

Selection of Genetic Risk Factors for Times of Disease Progression Subject to Interval-censoring

Sélection des facteurs de risque génétique pour les durées de progression de maladie censurées par intervalles

Disease progression times are interval-censored when progression status is only known at a series of assessment times. We consider the problem of selecting important biomarkers prognostics for progression from a high dimensional set of covariates when the progression time is interval-censored. We extend the idea of penalized regression (e.g. LASSO, adaptive LASSO and SCAD) for interval-censored data via failure time models with flexible piecewise constant hazard functions. An EM algorithm is developed which is shown to perform well empirically. An application involving the selection of important haplotype variables in predicting disease progression in psoriatic arthritis is given for illustration.

Disease progression times are interval-censored when the progression status is only known at a series of assessment times. We consider the problem of selecting important biomarkers prognostics for progression from a high dimensional set of covariates when the progression time is interval-censored. We extend the idea of penalized regression (for example: Lasso, adaptive Lasso and SCAD) for interval-censored data via failure time models with flexible piecewise constant hazard functions. An EM algorithm is developed which is shown to perform well empirically. An application involving the selection of important haplotype variables in predicting disease progression in psoriatic arthritis is given for illustration.

3D-C3: Competing Risks and Disease Progression

Risques concurrents et progression des maladies

la polyarthrite psoriasique, est donnée comme illustration.

3D-C4: Mixture Models Modèles de mélange

Chair/Présidente: Anne-Sophie Charest (Université Laval)

Room/Salle: EdSouth 213

Abstracts/Résumés

KATHERINE MORRIS & PAUL MCNICHOLAS (University of Guelph)

[Wednesday May 29 / mercredi 29 mai, 15:00]

Non-Gaussian Clustering and Dimension Reduction

Regroupement non gaussien et réduction de dimension

A dimension reduction method for model-based clustering via a finite mixture of non-Gaussian distributions – specifically the t , shifted asymmetric Laplace and generalized hyperbolic distributions – is introduced. The approach is analogous to existing work within the Gaussian paradigm and relies on identification of a reduced subspace. This subspace contains linear combinations of the original data, ordered by importance using the associated eigenvalues. Our clustering approaches are illustrated on simulated and real data, and compared to each other as well as their Gaussian counterpart.

Nous présentons une méthode de réduction de la dimension pour le regroupement à l'aide de modèles de mélange fini de distributions non gaussiennes, nommément les distributions de t , asymétrique de Laplace décalée et hyperbolique généralisée. Analogue à des travaux existant dans le cadre du paradigme gaussien, l'approche se fonde sur l'identification d'un sous-espace réduit. Ce sous-espace contient des combinaisons linéaires des données originales classées par ordre d'importance selon les valeurs propres associées. Les approches de regroupement proposées sont illustrées à l'aide de données réelles et simulées et comparées entre elles et à leur contrepartie gaussienne.

JEFFREY L. ANDREWS, PAUL MCNICHOLAS & RYAN BROWNE (University of Guelph)

[Wednesday May 29 / mercredi 29 mai, 15:15]

Robust Latent Variable Mixture Models for Clustering Data of Mixed Type

Modèles de mélanges robustes à variables latentes pour des données en grappes de type mixte

Data of mixed type arise when there are two or more variable forms, such as categorical and interval. A framework for mixture model-based clustering was recently proposed to handle this type of data that uses latent Gaussian variables (akin to factor analysis in the continuous case and latent trait analysis in the categorical case). In this presentation, we extend this work to introduce more robust models by utilizing t -distributed, instead of normally distributed, latent variables. Model fitting is performed via an EM algorithm. Our approach is compared with the Gaussian approach using real and simulated data.

Les données de type mixte sont présentes lorsqu'il y a deux ou plusieurs formes de variables, comme des variables de catégorie et d'intervalle. Un cadre pour le regroupement à l'aide de modèles de mélange a récemment été proposé pour traiter ce type de données qui utilisent des variables latentes gaussiennes (apparentées à l'analyse factorielle dans le cas continu et à l'analyse de trait latent dans le cas catégorique). Dans cet exposé, nous étendons ces travaux en proposant des modèles plus robustes qui utilisent des variables latentes suivant une distribution de Student plutôt que normale. Nous ajustons le modèle au moyen d'un algorithme EM. Nous comparons notre méthode à l'approche gaussienne à l'aide de données réelles et simulées.

3D-C4: Mixture Models Modèles de mélange

MOHAMMAD JAFARI JOZANI & ARMIN HATEFI (University of Manitoba), **OMER OZTURK** (Ohio State University)

[Wednesday May 29 / mercredi 29 mai, 15:30]

Mixture Model Analysis of Partially Rank Ordered Set Samples: Estimating the Age-Groups of Fish from Length-Frequency Data

Analyse de modèles de mélanges pour des échantillons d'ensembles partiellement ordonnés: estimation des groupes d'âge de poissons à partir de la fréquence des longueurs

We study inference for finite mixture models (FMM) based on partially rank-ordered set (PROS) sampling design. The PROS samples have a unique data structure different from simple random samples (SRS) and the traditional EM-algorithm is not applicable. We propose an EM algorithm for estimation based on PROS samples. We study classification of PROS samples into the components of the FMMs. The maximum likelihood estimators based on PROS design lead to more efficient estimates compared to their SRS counterparts. The results are used to estimate the age-group of a fish species in the Chesapeake Bay area using the length-frequency data.

Nous étudions l'inférence pour des modèles de mélanges finis (MMF) basée sur un plan d'échantillonnage d'ensembles partiellement ordonnés (EPO). Les échantillons d'EPO présentent une structure de données unique différente de celle des échantillons aléatoires simples (EAS) et l'algorithme EM classique ne s'applique pas. Nous proposons un algorithme EM pour l'estimation basée sur des échantillons d'EPO. Nous étudions la classification des échantillons d'EPO dans les composantes des MMF. Les estimateurs du maximum de vraisemblance basés sur des plans d'EPO sont plus efficaces que leurs homologues avec EAS. Les résultats sont utilisés pour estimer les groupes d'âge d'une espèce de poisson dans la région de la baie de Chesapeake à l'aide de données sur la fréquence des longueurs.

TASNIMA ABEDIN & JINGJING WU (University of Calgary)

[Wednesday May 29 / mercredi 29 mai, 15:45]

A Study on Two-Sample Mixture Model

Étude sur un modèle de mélange à deux échantillons

Suppose there is a sample from mixture population $(1 - \lambda)F + \lambda G$ and, independently, another sample from the first component F , where λ is an unknown mixing proportion such that $0 < \lambda < 1$ and F and G are two different unknown distribution functions. The goal is to estimate λ and G . This model arises very frequently in biostatistics, such as when detecting differentially expressed genes under two conditions (eg. healthy tissues vs. diseased tissues). To estimate λ and G , we are considering a log linear ratio model.

Supposons un échantillon d'un mélange de populations $(1 - \lambda)F + \lambda G$ et un autre échantillon, indépendant du premier, provenant de la première composante F , où λ est une proportion de mélange inconnue telle que $0 < \lambda < 1$ et F et G sont deux fonctions de distribution inconnues. L'objectif est d'estimer λ et G . Ce modèle est très fréquent en biostatistique, par exemple pour la détection de gènes exprimés de façon différentielle sous deux conditions (par exemple des tissus sains par rapport à des tissus malades). Pour estimer λ et G , nous considérons un modèle log-linéaire de rapport.

3D-C5: Screening and Diagnostic Tests Examens préventifs et tests de diagnostic

Chair/Présidente: Cecilia Cotton (University of Waterloo)

Room/Salle: EdSouth 206

Abstracts/Résumés

AMAY CHEAM & PAUL MCNICHOLAS (University of Guelph)

[Wednesday May 29 / mercredi 29 mai, 15:00]

Modelling Receiver Operating Characteristics (ROC) Curve Using Gaussian Mixture

Modélisation de la courbe des caractéristiques receveur-opérateur (ROC) avec le modèle de mélange gaussien

The receiver operating characteristic curve is widely applied in measuring performance of diagnostic tests. Many direct and indirect approaches have been proposed for modelling the ROC curve. Because its ease of manipulation, the Gaussian distribution has been used to model both populations. Instead, we propose a Gaussian mixture, leading to a more flexible model that accounts for 'atypical' data. Monte Carlo simulation is used to circumvent the issue of absence of a closed-form and to obtain confidence bands for the derived ROC. We show that our method performs favourably when compared to the binormal curve.

La courbe des caractéristiques receveur-opérateur est un outil statistique transcendant pour décrire la performance des tests diagnostiques. Plusieurs approches directes et indirectes ont été proposées pour modéliser la courbe ROC. Grâce à sa simplicité, la distribution gaussienne a été utilisée pour la modélisation des deux populations. Alternativement, nous proposons un modèle de mélange gaussien qui procurera plus de flexibilité avec les données atypiques. Puisque nous n'avons pas de formule explicite, la méthode de Monte Carlo est utilisée afin d'obtenir des intervalles de confiance. Comparativement à la binormale, notre méthode ajuste efficacement la courbe empirique.

JAMES HANLEY & ZHIHUI (AMY) LIU (McGill University)

[Wednesday May 29 / mercredi 29 mai, 15:15]

Statistical Animations in Understanding Mortality Reductions Due to Cancer Screening

Animations statistiques dans la compréhension des réductions de mortalité grâce au dépistage du cancer

Our website provides animations that we have used to illustrate statistical concepts involving: time, event counts, and variability: the Poisson distribution, life-table functions, connections between tails of discrete and continuous distributions, Cox regression, epidemics, and effects of collinearity and measurement error. The talk describes our use of animations to understand how repeated rounds of screening halt/alter the time-course of a cancer; why screening-induced reductions in age-specific mortality rates might exhibit a bathtub-shaped rate ratio curve; and how a 3-parameter model determining this curve's location, depth and duration can be used to characterize the effect of a single round of screening.

Notre site Web offre des animations que nous avons utilisées pour illustrer des concepts statistiques relatifs au temps, aux dénombrements d'événements et à la variabilité : la distribution de Poisson, les fonctions de table de mortalité, les connexions entre les queues des distributions discrètes et continues, la régression de Cox, les épidémies et les effets de colinéarité, ainsi que l'erreur de mesure. Cet article décrit l'utilisation que nous faisons des animations pour comprendre comment des cycles répétés de dépistage arrêtent ou modifient le décours temporel d'un cancer. Il explique comment les réductions du taux de mortalité spécifiques à l'âge grâce aux dépistages pourraient montrer une courbe de rapport des taux en forme de baignoire. Cet exposé décrit aussi comment un modèle à trois paramètres déterminant l'endroit, la profondeur et la durée de

3D-C5: Screening and Diagnostic Tests

Examens préventifs et tests de diagnostic

cette courbe peut être utilisé pour caractériser l'effet d'un seul cycle de dépistage.

KHOKAN C. SIKDAR, YUTAKA YASUI & MARCY WINGET (University of Alberta)

[Wednesday May 29 / mercredi 29 mai, 15:30]

Diagnostic Intervals of Colorectal Cancer in Alberta: A Population-Based Study

Intervalles diagnostiques pour le cancer colorectal en Alberta: une étude basée sur la population

We conducted a study to evaluate factors associated with the length of the diagnostic interval in patients diagnosed with colorectal cancer (CRC) and the impact of the length of the interval on patient stage at diagnosis and survival. Methods were applied to both real data as well as in a simulation study. The real data included all patients diagnosed with CRC in years 2000-2005; population-based administrative data were used to identify dates of interaction with the healthcare system. Cox-regression analysis was performed to assess if length of the diagnostic interval was independently predictive of disease stage and patient survival.

Nous avons mené une étude visant à évaluer les facteurs liés à la durée de l'intervalle diagnostique chez les patients atteints du cancer colorectal et l'impact de ce délai sur l'état du patient au moment du diagnostic et sur la survie. Les méthodes ont été appliquées à des données réelles et simulées. Les données réelles comprenaient tous les patients ayant reçu un diagnostic de cancer colorectal entre 2000 et 2005 ; des données administratives sur la population ont permis de déterminer les dates des interactions avec le système de santé. Une régression de Cox a été effectuée pour évaluer si la durée de l'intervalle diagnostique était prédictif du stade de la maladie et de la survie du patient.

WANHUA SU (Grant MacEwan University), **MU ZHU** (University of Waterloo), **YAN YUAN** (University of Alberta)

[Wednesday May 29 / mercredi 29 mai, 15:45]

Average Precision Analysis for Screening and Diagnostic Tests

Analyse de précision moyenne pour les tests de dépistage et de diagnostic

The average precision (AP) is a popular performance metric for comparing ranking systems in information retrieval, but its statistical properties are not well understood, which has limited its application elsewhere. We show that both the AUC (area under the Receiver Operating Characteristic curve) and the AP are weighted averages of precisions with a series of decreasing weights. The AP discriminates better than the AUC because its weights are data-driven, whereas those for the AUC are pre-determined. Using both simulation and real-data examples, we illustrate the use and advocate the inclusion of the AP when comparing screening and diagnostic tests.

Utilisée couramment dans la comparaison de systèmes de classement en recherche documentaire, la précision moyenne (PM) est une mesure de performance dont les propriétés statistiques sont mal comprises, ce qui a limité son application dans d'autres domaines. Nous montrons que l'aire sous la courbe ROC (« AUC ») et la PM sont des moyennes pondérées de précision comportant une série de poids décroissants. La PM permet une meilleure discrimination que l'AUC, parce que ses poids sont axés sur les données, alors que ceux de l'AUC sont préétablis. À l'aide d'exemples basés sur des simulations et des données réelles, nous illustrons l'utilisation de la PM et préconisons son inclusion dans la comparaison de tests de dépistage et de diagnostic.

3D-C6: Statistics: Applications and Case Studies Statistique : applications et études de cas

Chair/Présidente: Ying Zhang (Acadia University)

Room/Salle: EdSouth 265

Abstracts/Résumés

DAVID ROSS BRILLINGER (University of California Berkeley)

[Wednesday May 29 / mercredi 29 mai, 15:00]

Aligning Some Nicholson Sheep-Blowfly Data Sets with System Input

Aligner quelques jeux de données de Nicholson sur la mouche à viande aux intrants du système

During the 1950s the Australian entomologist Alexander Nicholson studied an insect pest, the sheep-blowfly. In laboratory experiments populations were set up in cages. In eight of the experiments the input food was varied in saw tooth fashion, each experiment having a given period. However which data set went which which period of application became unknown. In this study use is made of periodograms, spectrograms and seasonal adjustment to infer a correspondence. A variety of nonlinear dynamical systems have been fit to an experiment involving constant food input.

Au cours des années 1950, l'entomologiste australien Alexander Nicholson a étudié un insecte ravageur, la viande à mouche. Les populations expérimentales étaient installées dans des cages en laboratoire. Dans huit expériences, les apports en nourriture n'étaient pas constants, chaque expérience se déroulant sur une période donnée. Cependant, la période d'application rattachée à chaque jeu de données était inconnue. Les périodogrammes, les spectrogrammes et l'ajustement saisonnier sont employés dans cette étude pour inférer une correspondance. Différents systèmes dynamiques non-linéaires ont été ajustés lors d'une expérience comportant un apport constant en nourriture.

THEODORO KOULIS & SAMAN MUTHUKUMARANA (University of Manitoba)

[Wednesday May 29 / mercredi 29 mai, 15:15]

A Bayesian Stochastic Model for Batting Performance Evaluation in One-Day Cricket

Un modèle stochastique bayésien pour l'évaluation de la performance au bâton dans le cricket d'une journée

We consider a Bayesian ergodic Markov chain to model individual batting performance in one-day cricket. A cut-point vector is used to create a discrete state space for batting performance. By choosing an appropriate cut-point vector, the next performance state only depends on the current state, so that an overall prediction of batting performance of a batsman can be made. One advantage of this Markov model is that it does not require the consideration of unforeseen factors. This is important since cricket has gone through several rule changes in recent years that have further induced unforeseen dynamic factors to the game.

Nous considérons une chaîne de Markov bayésienne pour modéliser la performance individuelle au bâton dans le cricket d'une journée. Un vecteur de coupure est utilisé pour créer un espace d'états discret de performance. En choisissant un vecteur de coupure approprié, l'état prochain dépend seulement de l'état courant et une prédiction de performance peut être effectuée. L'avantage de ce modèle de Markov est que nous ne sommes pas obligés de considérer les facteurs imprévus pour modéliser la performance. Ceci est important car le cricket a subi plusieurs changements de règles qui ont provoqué des facteurs imprévus et dynamiques sur le jeu.

3D-C6: Statistics: Applications and Case Studies Statistique : applications et études de cas

TULSI PAUDEL (University of Guelph), **DAWN M. VANLEEUWEN** & **BETSY BENDER STRINGAM** (New Mexico State University)

[Wednesday May 29 / mercredi 29 mai, 15:30]

Iterative Model Building Using Regression Diagnostics and Mean Shift Parameters in Logistic Regression for the Analysis of a Large Hotel Rating Data Set

Construction d'un modèle itératif par des diagnostics de régression et de paramètres de décalage moyen dans la régression logistique pour l'analyse d'un grand ensemble de données de cotation d'hôtels

When performing logistic regression, extremely large data sets have increased sensitivity to lack of fit that can complicate data analysis. However, large data sets also allow a rich exploration of relationships between covariate patterns and the response variable. This study explores an iterative process of model fitting, residual analysis, and creating mean-shift variables according to patterns observed among high magnitude residuals for inclusion in the next iteration. This process based on the coefficients from the iterative model fits and mean-shift parameters are demonstrated using a data set consisting of 60648 consumer hotel ratings from an online travel website.

Lors d'une régression logistique, de très grands jeux de données deviennent plus sensibles au manque d'adéquation, ce qui peut compliquer l'analyse des données. Cependant, de grands jeux de données facilitent l'exploration des relations entre les modèles de covariables et les variables-réponses. Cette étude examine un processus itératif de l'adéquation de modèle, une analyse résiduelle et la création de variables moyennes de changement selon les modèles observés parmi les résidus très importants à inclure dans la prochaine itération. On démontre ce processus basé sur les coefficients provenant des adéquations de modèle itératives et des paramètres de décalage moyen par l'utilisation d'un jeu de données de 60 648 cotations de clients d'hôtels sur un site Web de voyage en ligne.

MARK REESOR & ALMAS NASEEM (University of Western Ontario)

[Wednesday May 29 / mercredi 29 mai, 15:45]

Risk Analysis of the Smith Manoeuvre for Re-advanceable Canadian Mortgages

Analyse des risques de la manœuvre de Smith pour les prêts sur la valeur nette des maisons canadiennes

According to Canadian tax law one way of transforming from non-tax deductible (personal mortgage) to tax deductible interest expenses is to borrow against home equity to make investments. A re-advanceable mortgage is a product in which the mortgagor immediately re-borrows principal as it is paid. We assume the re-borrowed funds are invested into a single risky asset. We study the risk associated with this strategy to provide a better description of the mortgagor's position. Factors such as income (tax rate), interest rate and housing prices are considered.

D'après le droit fiscal canadien, l'une des façons de transformer des intérêts non déductibles (emprunt hypothécaire personnel) en intérêts déductibles est de s'en servir pour garantir la valeur nette du logement et de faire des investissements. Un prêt sur la valeur nette d'une maison est le produit dans lequel le débiteur hypothécaire réemprunte immédiatement le capital une fois qu'il est payé. Nous assumons que les fonds réempruntés sont investis dans un seul actif risqué. Nous étudions le risque associé à cette stratégie pour mieux décrire la position du débiteur hypothécaire. Des facteurs comme le revenu (taux d'imposition), le taux d'intérêt et les prix du logement sont pris en compte.

Author List • Liste des auteurs

- Abdelrazeq, Ibrahim, 35, 104 (1C-C4)
 Abedin, Tasnima, 63, 228 (3D-C4)
 Abraham, Bovas, 60, 213 (3C-C1)
 Adcock, C. J., 50, 168 (2C-I4)
 Adjogou, Folly, 60, 215 (3C-C2)
 Afkanpour, Arash, 45, 149 (2B-I3)
 Afzal, Nasrin, 36, 110 (1C-C6)
 Agarwal, Reshu, 29, 82 (1B-C3)
 Ahmed, Rashid, 47, 157 (2B-C3)
 Ahmed, Syed Ejaz, 57, 199 (3B-C2)
 Ajavon, Ayi, 41, 135 (1D-C3)
 Akhtar-Danesh, Noori, 49, 163 (2C-I1)
 Albert-Green, Alisha, 34, 102 (1C-C3)
 Alobaidan, Khatoon, 29, 82 (1B-C3)
 Alsubie, Abdelaziz, 29, 79 (1B-C2)
 Amezziane, Mohamed, 57, 199 (3B-C2)
 Andrews, Jeffrey L., 63, 227 (3D-C4)
 Araujo, Cynthia, 41, 131 (1D-C2)
 Arbuckle, Tye, 49, 164 (2C-I2)
 Armstrong, Paul W., 40, 128 (1D-C1)
 Asgharian, Masoud, 46, 155 (2B-C2)
 Ashta, Shiva, 36 (1C-D1)
 Assa, Hirbod, 51, 174 (2C-C1)
 Atenafu, Eshetu G., 47, 157 (2B-C3)
 Athey, Taryn, 37, 113 (1C-P)
 Azzalini, Adelchi, 50, 168 (2C-I4)
- Bae, Taehan, 46 (2B-I4), 60 (3C-C1) 151 (2B-I4), 214 (3C-C1)
 Bagnato, Luca, 57, 197 (3B-C1)
 Bai, Sheldon, 38, 119 (1D-I1)
 Bakal, Jeff, 40, 128 (1D-C1)
 Balakrishnan, Narayanaswamy, 33 (1C-C1) 96 (1C-C1)
 Balasooriya, Uditha, 62, 222 (3D-C2)
 Bayne, Erin, 51, 177 (2C-C2)
 Beezley, Jonathan, 46, 153 (2B-C1)
 Beg, Mirza Faisal, 32, 89 (1C-I3)
 Behseta, Sam, 27, 70 (1B-I2)
- Bell, William, 35, 107 (1C-C5)
 Bergeron, Pierre-Jérôme, 40 (1D-C1) 53 (2D-I3), 128 (1D-C1), 185 (2D-I3)
 Besbeas, Takis, 34, 102 (1C-C3)
 Bickis, Mikelis, 33, 97 (1C-C1)
 Bien, Jacob, 50, 170 (2C-I5)
 Bingham, Derek, 55 (3A-A), 59 (3C-I2) 188 (3A-A), 207 (3C-I2)
 Bleuer, Susana Rubin, 28, 74 (1B-I4)
 Bohrmann, Thomas, 31, 87 (1C-I2)
 Bolker, Benjamin, 31, 87 (1C-I2)
 Bornn, Luke, 61, 219 (3D-A2)
 Bosch, Ronald J., 62, 220 (3D-C1)
 Bouhaddioui, Chafik, 42, 136 (1D-C3)
 Bowling, Michael, 45, 149 (2B-I3)
 Branscum, Adam, 32, 91 (1C-I4)
 Braun, W. John, 34 (1C-C3), 44 (2B-I1) 102, 103 (1C-C3), 145 (2B-I1)
 Bremner, Karen, 41, 133 (1D-C2)
 Brillinger, David Ross, 64 (3D-C6) 231 (3D-C6)
 Briollais, Laurent, 63, 225 (3D-C3)
 Brooks, Mollie, 31, 87 (1C-I2)
 Brown, Steve, 47, 157 (2B-C3)
 Browne, Richard, 61, 216 (3C-C2)
 Browne, Ryan, 36 (1C-C6), 38 (1C-P) 63 (3D-C4), 110 (1C-C6), 117 (1C-P), 227 (3D-C4)
 Bulaevskaya, Vera, 59, 207 (3C-I2)
 Bull, Shelley B., 33, 96 (1C-C1)
 Burke, Murray, 29, 80 (1B-C2)
 Burr, Wesley, 34, 102 (1C-C3)
- Cabilio, Paul, 59 (3C-I3), 62 (3D-C2) 209 (3C-I3), 222 (3D-C2)
 Cadigan, Noel, 31, 87 (1C-I2)
 Cai, Jun, 31, 85 (1C-I1)
 Cai, Song, 57, 199 (3B-C2)
 Cai, Yanling, 42, 135 (1D-C3)
 Cameron, Erin, 51, 177 (2C-C2)

- Campbell, David, 45 (2B-I2), 51 (2C-C2) 147 (2B-I2), 177 (2C-C2)
- Cantoni, Eva, 39 (1D-I4), 47 (2B-C3) 124 (1D-I4), 159 (2B-C3)
- Carriere, Keumhee C., 43, 141 (1D-C5)
- Cenny, Taslim, 38, 119 (1D-I1)
- Chaieb-Lakhal, M'hamed Lajmi, 56 (3B-C1) 196 (3B-C1)
- Chang, Hsing-Ming, 46, 153 (2B-C1)
- Chapman, Judy-Anne, 49, 164 (2C-I2)
- Charest, Anne-Sophie, 61, 215 (3C-C2)
- Chatrchi, Golshid, 42, 138 (1D-C4)
- Cheam, Amay, 63, 229 (3D-C5)
- Chen, Hao (Nelson), 27, 72 (1B-I3)
- Chen, Huijun, 31, 85 (1C-I1)
- Chen, Jiahua, 57, 199 (3B-C2)
- Chen, Min, 35, 107 (1C-C5)
- Chen, Shiyi, 37 (1C-D1)
- Chenouri, Shoja'eddin, 27 (1B-I2), 59 (3C-I1) 71 (1B-I2)
- Chipman, Hugh, 29 (1B-C3), 34 (1C-C2) 38 (1C-P), 82 (1B-C3), 100 (1C-C2), 118 (1C-P)
- Chkrebti, Oksana, 51, 177 (2C-C2)
- Choi, Yun-Hee, 62, 224 (3D-C3)
- Chough, Kimmie C., 58, 203 (3B-C3)
- Choulli, Tahir, 27, 69 (1B-I1)
- Christman, Mary, 31, 87 (1C-I2)
- Chu, Maggie, 37, 113 (1C-P)
- Cobb, Loren, 46, 153 (2B-C1)
- Cohen, Doug, 36 (1C-D1)
- Colin, Bernard, 57, 198 (3B-C1)
- Cook, Richard, 24, 29 (1B-C2) 33 (1C-I6), 40 (1D-C1), 56 (3B-C1), 63 (3D-C3), 81 (1B-C2), 94 (1C-I6), 130 (1D-C1), 196 (3B-C1), 225 (3D-C3)
- Corkum, Amber, 62, 222 (3D-C2)
- Cossette, H el ene, 31 (1C-I1) 85 (1C-I1)
- Cote, Marie-Pier, 37 (1C-D2)
- Cotton, Cecilia, 53, 184 (2D-I3)
- Coull, Jesse, 51, 172 (2C-I6)
- Coulombe, Janie, 36 (1C-D1)
- Cowen, Laura, 34, 102 (1C-C3)
- Creed, Irena, 40, 127 (1D-I5)
- Csorgo, Miklos, 55, 190 (3B-I1)
- Dalessandro, Brian, 60, 211 (3C-I4)
- Daly, Caitlin, 29, 79 (1B-C2)
- Dang, Sanjeena, 29, 82 (1B-C3)
- Dang, Utkarsh, 63, 224 (3D-C3)
- Dansu, Bolanle, 37, 113 (1C-P)
- Datta, Gauri Sankar, 35, 107 (1C-C5)
- Davies, Katherine, 59 (3C-I1)
- Dawson, Don, 40, 126 (1D-I5)
- Dean, Charmaine, 32, 89 (1C-I3)
- Deardon, Rob, 46, 153 (2B-C1)
- De Capitani, Lucio, 57, 197 (3B-C1)
- de Leon, Alexander R., 42 (1D-C3) 135 (1D-C3)
- Dendukuri, Nandini, 33, 97 (1C-C1)
- de Souza, Camila, 51, 177 (2C-C2)
- de Tibeiro, Jules Joseph S., 57 (3B-C1) 198 (3B-C1)
- Dewanji, Anup, 42, 138 (1D-C4)
- Diao, Liqun, 56, 196 (3B-C1)
- Diaz, Mireya, 58, 202 (3B-C3)
- Ding, Jing, 59, 209 (3C-I3)
- Ding, Keyue, 40, 128 (1D-C1)
- Donner, Allan, 43, 141 (1D-C5)
- Dubin, Joel A., 32 (1C-I5)
- Duchesne, Pierre, 52 (2C-C2), 57 (3B-C2) 178 (2C-C2), 200 (3B-C2)
- Duchesne, Thierry, 39, 124 (1D-I4)
- Dumais, Jean, 56, 192 (3B-I2)
- Dunham, Bruce, 44, 145 (2B-I1)
- Dunne, James, 61, 215 (3C-C2)
- Durocher, Stephane, 46, 155 (2B-C2)
- Elghafghuf, Adel, 40, 129 (1D-C1)
- Eliasziw, Misha, 47, 158 (2B-C3)
- Elktaibi, Farid, 35, 104 (1C-C4)
- Elmasri, Mohamad, 37 (1C-D2)
- Eltinge, John, 58, 204 (3C-A)
- Escarela, Gabriel, 40, 129 (1D-C1)
- Esterby, Sylvia, 41 (1D-C2), 49 (2C-I3) 131 (1D-C2), 166 (2C-I3)
- Farrell, Patrick, 42, 138 (1D-C4)
- Farsinezhad, Sepideh, 46, 155 (2B-C2)
- Feng, Zeny, 41, 132 (1D-C2)
- Ferraty, Frederic, 45, 147 (2B-I2)
- Fick, Gordon, 47, 158 (2B-C3)
- Field, Chris, 47 (2B-C3), 52 (2D-I1) 159 (2B-C3), 181 (2D-I1)
- Fillion, Jean-Marc, 28, 74 (1B-I4)
- Firouzi, Hassan Omid, 51 (2C-C1) 174 (2C-C1)
- Fishbane, Nicholas, 36 (1C-D1)
- Fisher, Mandy, 49, 164 (2C-I2)
- Flemming, Joanna Mills, 39 (1D-I4) 47 (2B-C3), 124 (1D-I4), 159 (2B-C3)
- Fokianos, Konstantinos, 55 (3B-I1) 190 (3B-I1)
- Forbes, Peter, 56, 194 (3B-I3)
- Fortin, Daniel, 39, 124 (1D-I4)

- Franczak, Brian, 36, 110 (1C-C6)
 Friendly, Michael, 39, 121 (1D-I2)
- Gadag, Veeresh, 62, 222 (3D-C2)
 Gagné, Christian, 36 (1C-C5) 108 (1C-C5)
 Gao, Xin, 41, 132 (1D-C2)
 Garrido, Jose, 51, 176 (2C-C1)
 Ge, Tian, 32, 89 (1C-I3)
 Gee, Steven, 27, 70 (1B-I2)
 Genest, Christian, 31, 85 (1C-I1)
 Genton, Marc G., 50, 168 (2C-I4)
 Ghahramani, Melody, 57, 199 (3B-C2)
 Ghosh, Palash, 42, 138 (1D-C4)
 Gibbs, Alison, 53 (2D-I2)
 Gilbert, Paul, 39, 121 (1D-I2)
 Gill, Paramjit, 41 (1D-C2), 51 (2C-C1) 131 (1D-C2), 175 (2C-C1)
 Golchi, Shirin, 52 (2C-C2), 59 (3C-I2) 178 (2C-C2), 207 (3C-I2)
 Gombay, Edit, 55, 190 (3B-I1)
 Gravel, Christopher A., 42 (1D-C4) 138 (1D-C4)
 Gray, Brian, 49, 166 (2C-I3)
 Gray, Darren, 36, 108 (1C-C5)
 Grebennikov, Ievgenii, 37 (1C-D2)
 Greenwood, Priscilla, 53, 186 (2D-I4)
 Grey, Vijay, 29, 79 (1B-C2)
 Groparu, Ionica, 51, 176 (2C-C1)
 Gruber, Susan, 60, 211 (3C-I4)
 Gustafson, Paul, 32 (1C-I4), 41 (1D-C2) 91 (1C-I4), 131 (1D-C2)
 Gyorgy, Andras, 45, 149 (2B-I3)
- Halder, Avik, 28, 76 (1B-C1)
 Haley, Charlotte, 37 (1C-D1)
 Hamid, Jemila, 29 (1B-C2), 33 (1C-C1) 62 (3D-C1), 79 (1B-C2), 96 (1C-C1), 221 (3D-C1)
 Hamid, Pezeshk, 37, 115 (1C-P)
 Han, Spectrum, 37, 114 (1C-P)
 Hanley, James, 33 (1C-C1), 63 (3D-C5) 97 (1C-C1), 229 (3D-C5)
 Hanner, Robert, 37, 113 (1C-P)
 Hanook, Sharoon, 38, 119 (1D-I1)
 Hanson, Tim, 32, 91 (1C-I4)
 Hasler, Caren, 42, 138 (1D-C4)
 Hatefi, Armin, 63, 227 (3D-C4)
 Hauff, Kristin, 29, 79 (1B-C2)
 He, Feng, 42, 139 (1D-C4)
 He, Wenqing, 37, 113 (1C-P)
 He, Zhoushanyue, 40, 128 (1D-C1)
- Heckman, Nancy, 51, 177 (2C-C2)
 Hein, Andrew, 31, 87 (1C-I2)
 Henderson, Nicholas, 55, 189 (3B-V)
 Henrey, Andrew, 62, 223 (3D-C2)
 Hidioglou, Mike, 36, 108 (1C-C5)
 Hoegg, Tanja, 41, 131 (1D-C2)
 Honerkamp-Smith, Gordon, 33 (1C-I6) 94 (1C-I6)
 Hooker, Giles, 45, 147 (2B-I2)
 Hossain, Shakhawat, 57, 199 (3B-C2)
 Hu, X. Joan, 42, 139 (1D-C4)
 Huang, Celia, 56, 192 (3B-I2)
 Huang, Longlong, 36, 110 (1C-C6)
 Huang, Mei Ling, 62, 223 (3D-C2)
 Huang, Qianyi, 37 (1C-D1)
 Huang, WeiLiang, 62, 221 (3D-C1)
 Hugh, Chipman, 59, 207 (3C-I2)
 Hung, Kelvin, 36 (1C-D1)
 Hussein, Abdulkadir, 55, 190 (3B-I1)
 Huston, Carolyn, 50, 167 (2C-I3)
 Hyndman, Cody, 27, 68 (1B-I1)
 Hyslop, Terry, 32 (1C-I5)
- Ingrassia, Salvatore, 29, 82 (1B-C3)
 Ingster, Yuri, 54, 187 (2D-I4)
 Iscoe, Ian, 60, 214 (3C-C1)
 Ismaila, Afisi, 49, 164 (2C-I2)
 Ivanoff, Gail, 35 (1C-C4), 40 (1D-I5) 104 (1C-C4), 126 (1D-I5)
- Jana, Sayantee, 33, 96 (1C-C1)
 Jang, Gun Ho, 46, 155 (2B-C2)
 Jiang, Depeng, 29 (1B-C2), 38 (1C-P) 80 (1B-C2), 118 (1C-P)
 Jin, Zhiyong, 37, 114 (1C-P)
 Johnson, Timothy D., 32, 89 (1C-I3)
 Johnson, Wes, 32, 91 (1C-I4)
 Joncas, Marc, 56, 192 (3B-I2)
 Joy, Ruth, 37 (1C-D2)
 Jozani, Mohammad Jafari, 63 (3D-C4) 227 (3D-C4)
 Julien, Pierre-Olivier, 36 (1C-C5) 108 (1C-C5)
 Jun, Seong-Hwan, 36 (1C-D1)
- Kainulainen, Craig, 35, 104 (1C-C4)
 Kalitsi, Clarence D., 51, 174 (2C-C1)
 Karim, Mohammad Ehsanul, 41 (1D-C2) 131 (1D-C2)
 Karsten, Richard, 59, 208 (3C-I2)
 Keen, Kevin, 61, 215 (3C-C2)
 Keim, Jonah, 52, 181 (2D-I1)
 Keles, Sunduz, 29, 80 (1B-C2)

- Keller, Sallie, 26, 66 (1A-A2)
 Kendzioriski, Christina, 38 (1D-I1) 119 (1D-I1)
 Khadraoui, Khader, 28, 76 (1B-C1)
 Khalili, Abbas, 45, 149 (2B-I3)
 Kim, Changki, 60, 214 (3C-C1)
 Kim, Mijeong, 50, 168 (2C-I4)
 Kingwell, Elaine, 41, 131 (1D-C2)
 Klesov, Oleg, 35, 105 (1C-C4)
 Kondo, Yumi, 34, 99 (1C-C2)
 Konigorski, Stefan, 33 (1C-C1), 37 (1C-D1) 96 (1C-C1)
 Kopciuk, Karen A., 32 (1C-I5), 36 (1C-C6) 110 (1C-C6)
 Kopp, Gregory, 34, 103 (1C-C3)
 Koren, Yehuda, 52, 180 (2D-A)
 Koulis, Theodoro, 64, 231 (3D-C6)
 Krahn, Murray, 41, 133 (1D-C2)
 Krewski, Daniel, 42, 138 (1D-C4)
 Krishnamurthy, Ashok, 46, 153 (2B-C1)
 Krzeminski, Mark, 34, 99 (1C-C2)
 Kulik, Rafal, 35, 104 (1C-C4)
 Kulperger, Reg, 35 (1C-C4), 36 (1C-C6) 40 (1D-I5), 52 (2C-C2), 105 (1C-C4), 111 (1C-C6), 127 (1D-I5), 178 (2C-C2)
 Kutney, Tracey, 59, 208 (3C-I2)
 Kwong, Pui Sze, 46, 153 (2B-C1)
- Labadi, Luai Al, 28, 76 (1B-C1)
 Labuda, Aleksander, 47, 160 (2B-C4)
 Lafaye de Micheaux, Pierre, 52 (2C-C2) 178 (2C-C2)
 Landriault, David, 45, 151 (2B-I4)
 Lansimaki, Evi, 37 (1C-D1)
 LaRoche, Sylvie, 56, 192 (3B-I2)
 Lawless, Jerry, 24, 47 (2B-C3) 53 (2D-I3), 157 (2B-C3), 184 (2D-I3)
 Lawryshyn, Yuri, 51, 172 (2C-I6)
 Lawson, Andrew, 32, 89 (1C-I3)
 Leblanc, Alexandre, 46, 155 (2B-C2)
 Lee, Chel Hee, 33, 97 (1C-C1)
 Lee, Jin Hyung, 36 (1C-D1)
 Lee, Jonathan, 52, 178 (2C-C2)
 Lee, Ker-Ai, 56, 196 (3B-C1)
 Léger, Christian, 26 (1A-A1)
 Lele, Subhash, 39 (1D-I2), 39 (1D-I4) 52 (2D-I1), 59 (3C-I3), 121 (1D-I2), 124 (1D-I4), 181 (2D-I1), 209 (3C-I3)
 Lemieux, Christiane, 45, 151 (2B-I4)
 Leng, Ning, 38, 119 (1D-I1)
 Leon, Carlos A., 36, 108 (1C-C5)
 Lesperance, Mary, 32, 89 (1C-I3)
- Li, Deli, 35, 105 (1C-C4)
 Li, Hao, 60, 213 (3C-C1)
 Li, Jingya, 51, 174 (2C-C1)
 Li, Weiming, 57, 200 (3B-C2)
 Li, Yi, 33, 94 (1C-I6)
 Li, Yin, 43, 141 (1D-C5)
 Liang, Kun, 29, 80 (1B-C2)
 Liang, Yuanyuan, 43, 141 (1D-C5)
 Lim, Hyun, 41, 132 (1D-C2)
 Lin, Chunfang Devon, 27, 73 (1B-I3)
 Lin, Daming, 37 (1C-D1)
 Lin, Shili, 38, 119 (1D-I1)
 Liu, Juxin, 28, 77 (1B-C1)
 Liu, Lihui, 59, 209 (3C-I3)
 Liu, Weiwei, 35, 105 (1C-C4)
 Liu, Xiaofeng, 29, 79 (1B-C2)
 Liu, Yang, 37, 115 (1C-P)
 Liu, Zhihui (Amy), 33 (1C-C1), 63 (3D-C5) 97 (1C-C1), 229 (3D-C5)
 Liu, Zi Zhen, 36, 111 (1C-C6)
 Lix, Lisa M., 37, 113 (1C-P)
 Lockhart, Richard, 38 (1C-P), 52 (2C-C2) 59 (3C-I1), 117 (1C-P), 178 (2C-C2)
 Loepky, Jason, 27, 72 (1B-I3)
 Lok, Judith J., 62, 220 (3D-C1)
 Lorenzi, Maria, 42, 139 (1D-C4)
 Lou, Wendy, 47, 157 (2B-C3)
 Loughin, Tom, 34 (1C-C2), 62 (3D-C2) 100 (1C-C2), 223 (3D-C2)
 Lu, Xuwen, 29 (1B-C2), 36 (1C-C6) 80 (1B-C2), 110 (1C-C6), 112 (1C-C6)
 Lu, Yi, 34, 100 (1C-C2)
 Lu, Zihang, 37 (1C-D1), 37 (1C-D2) 42 (1D-C3), 136 (1D-C3)
 Lucas, Don, 59, 207 (3C-I2)
 Luo, Xi (Rossi), 27, 70 (1B-I2)
 Lysy, Martin, 47, 160 (2B-C4)
 Lyubchich, Vyacheslav, 49 (2C-I3) 166 (2C-I3)
- Ma, Junfeng, 27, 69 (1B-I1)
 Ma, Wenkai (Kevin), 34, 103 (1C-C3)
 Ma, Yanyuan, 50, 168 (2C-I4)
 MacDonald, Blake, 34, 100 (1C-C2)
 Mach, Lenka, 28, 74 (1B-I4)
 Mailhot, Mélina, 57 (3B-C1) 196 (3B-C1)
 Makvandi, Ewa, 40, 128 (1D-C1)
 Malik, Rajat, 46, 153 (2B-C1)
 Mancina, Maria Elena Rivera, 47 (2B-C4) 160 (2B-C4)

- Mandel, Jan, 46, 153 (2B-C1)
 Marceau, Étienne, 31 (1C-I1) 57 (3B-C1), 85 (1C-I1), 196 (3B-C1)
 Mariaca Hajducek, Dagmar, 56 (3B-I2) 192 (3B-I2)
 Maroufy, Vahed, 51, 175 (2C-C1)
 Marriott, Paul, 27 (1B-I2), 51 (2C-C1) 71 (1B-I2), 175 (2C-C1)
 Martell, David, 34, 102 (1C-C3)
 Martsynyuk, Yuliya, 53 (2D-I4), 57 (3B-C2) 186 (2D-I4), 201 (3B-C2)
 Matus, Maciak, 61, 216 (3C-C2)
 McBride, Mary, 42, 139 (1D-C4)
 McCanny, Angela, 53 (2D-I2)
 McDonnell, Mark, 53, 186 (2D-I4)
 McIsaac, Michael, 40, 130 (1D-C1)
 McLeish, Don, 26 (1B-I1), 62 (3D-C1) 68 (1B-I1), 220 (3D-C1)
 McLeod, Ian, 38, 117 (1C-P)
 McNicholas, Paul, 29, 30 (1B-C3), 36 (1C-C6) 37, 38 (1C-P), 61 (3C-C2), 63 (3D-C3), 63 (3D-C4), 63 (3D-C5), 82–84 (1B-C3), 110 (1C-C6), 113 (1C-P), 117 (1C-P), 216, 217 (3C-C2), 224 (3D-C3), 227 (3D-C4), 229 (3D-C5)
 Melnikov, Alexander, 60, 213 (3C-C1)
 Mesfioui, Mhamed, 57, 196 (3B-C1)
 Metzler, Adam, 51 (2C-C1), 51 (2C-I6) 172 (2C-I6), 174 (2C-C1)
 Miller, Wayne, 59, 207 (3C-I2)
 Mitsakakis, Nicholas, 41, 133 (1D-C2)
 Mizera, Ivan, 61, 216 (3C-C2)
 Mondal, Prosanta, 41, 132 (1D-C2)
 Morales, Manuel, 51, 174 (2C-C1)
 Morgan, Byron, 34, 102 (1C-C3)
 Morris, Katherine, 63, 227 (3D-C4)
 Mulayath Variyath, Asokan, 62 (3D-C2) 222 (3D-C2)
 Murdoch, Duncan, 38, 116 (1C-P)
 Murray, Paula, 61, 216 (3C-C2)
 Murua, Alejandro, 29 (1B-C3), 60 (3C-C2) 83 (1B-C3), 215 (3C-C2)
 Muthukumarana, Saman, 28 (1B-C1), 64 (3D-C6) 77 (1B-C1), 231 (3D-C6)
 Muthuswamy, Lakshmi, 46, 155 (2B-C2)
 Nadeem, Khurram, 39 (1D-I2), 39 (1D-I4) 121 (1D-I2), 124 (1D-I4)
 Naik, Shanoja, 60, 213 (3C-C1)
 Nambeu, Christian Olivier, 36 (1C-C5) 108 (1C-C5)
 Nasari, Masoud, 55, 191 (3B-I1)
 Naseem, Almas, 64, 232 (3D-C6)
 Nathoo, Farouk, 32, 89 (1C-I3)
 Neuhaus, John M., 43, 142 (1D-C5)
 Newton, Michael, 55, 189 (3B-V)
 Nguyen, Jennifer, 56, 194 (3B-I3)
 Nichols, Thomas E., 32, 89 (1C-I3)
 Nie, Yunlong, 36 (1C-D1)
 Nielsen, Anders, 31, 88 (1C-I2)
 Obeidat, Mohammed, 28, 77 (1B-C1)
 Oldford, Wayne, 57, 197 (3B-C1)
 Osenberg, Craig, 31, 87 (1C-I2)
 Overstall, Antony, 27, 72 (1B-I3)
 Oyet, Alwell, 48, 162 (2B-C4)
 Oyono Ngou, Polynice, 27, 68 (1B-I1)
 Ozturk, Omer, 63, 227 (3D-C4)
 Palumbo, Francesco, 30, 83 (1B-C3)
 Partovi Nia, Vahid, 40, 128 (1D-C1)
 Paudel, Tulsii, 64, 231 (3D-C6)
 Payman, Nickchi, 37, 115 (1C-P)
 Payne, Nathaniel, 34, 100 (1C-C2)
 Peck, Roxy, 53 (2D-I2)
 Peng, Jianan, 29, 79 (1B-C2)
 Pennell, Beth-Ellen, 56, 192 (3B-I2)
 Pequeno, Priscila, 37 (1C-D2)
 Perera, Harsha, 51, 175 (2C-C1)
 Perlich, Claudia, 60, 211 (3C-I4)
 Perron, François, 41 (1D-C3) 135 (1D-C3)
 Petkau, A. John, 33 (1C-I5), 41 (1D-C2) 131 (1D-C2)
 Petrcich, William, 61, 215 (3C-C2)
 Picka, Jeffrey, 47, 160 (2B-C4)
 Pintilie, Melania, 42, 136 (1D-C3)
 Pirvu, Ciprian, 38, 116 (1C-P)
 Platt, Robert, 58, 202 (3B-C3)
 Pohlkamp-Hardt, Joshua, 37 (1C-D1)
 Polagye, Brain, 59, 208 (3C-I2)
 Poon, Jimmy, 34, 100 (1C-C2)
 Pordeli, Pooneh, 29, 80 (1B-C2)
 Poulin-Costello, Melanie, 40 (1D-C1) 130 (1D-C1)
 Provost, Foster, 60, 211 (3C-I4)
 Pullenayegum, Eleanor M, 62 (3D-C1) 221 (3D-C1)
 Punzo, Antonio, 29 (1B-C3), 57 (3B-C1) 82 (1B-C3), 197 (3B-C1)
 Qiao, Shan, 35, 103 (1C-C3)
 Qin, Jing, 61, 218 (3D-A1)
 Qin, Yingli, 57, 200 (3B-C2)

- Rad, Narges Nazeri, 47, 157 (2B-C3)
 Rahman, Adam, 38, 116 (1C-P)
 Rajapakshe, Rasika, 41, 131 (1D-C2)
 Ramezan, Reza, 27, 71 (1B-I2)
 Ramsay, James O., 45, 148 (2B-I2)
 Ramsay, Tim, 49, 164 (2C-I2)
 Ranjan, Pritam, 27 (1B-I3), 29 (1B-C3) 34 (1C-C2), 73 (1B-I3), 82 (1B-C3), 100 (1C-C2)
 Rao, J.N.K., 58, 204 (3C-A)
 Raymond-Belzile, Léo, 46 (2B-C2) 156 (2B-C2)
 Reesor, Mark, 51 (2C-C1), 64 (3D-C6) 174 (2C-C1), 232 (3D-C6)
 Regoli, Giuliana, 50, 168 (2C-I4)
 Rémillard, Bruno, 24
 Resende, Camila Casquilho, 36 (1C-D1)
 Ribereau, Pierre, 28, 76 (1B-C1)
 Riegert, David, 37 (1C-D1)
 Rivest, Louis-Paul, 39 (1D-I4), 56 (3B-C1) 124 (1D-I4), 196 (3B-C1)
 Rivière, Pascal, 28 (1B-I4) 74 (1B-I4)
 Romdhani, Hela, 56, 196 (3B-C1)
 Rosychuk, Rhonda J., 26 (1A-A1), 46 (2B-C1) 153 (2B-C1)
 Rotondi, Michael A, 43, 141 (1D-C5)
 RoyChoudhury, Arindam, 47 (2B-C3) 158 (2B-C3)
 Rutledge, Angela, 29, 79 (1B-C2)
- Saarela, Olli, 46, 156 (2B-C2)
 Sadeghi, Samira, 35, 106 (1C-C4)
 Sajobi, Tolulope, 37, 113 (1C-P)
 Salibián-Barrera, Matías, 37 (1C-P) 115 (1C-P)
 Sallier, Kenza, 36 (1C-D1)
 Sango, Joel, 57, 200 (3B-C2)
 Sattar, Abdus, 38, 119, 120 (1D-I1)
 Saunders, David, 24
 Schiopu-Kratina, Ioana, 28 (1B-I4) 74 (1B-I4)
 Schnitzer, Mireille, 62, 220 (3D-C1)
 Schwarz, Carl, 34, 102 (1C-C3)
 Scott, Alastair J., 43, 142 (1D-C5)
 Shen, Xiaotong, 56, 194 (3B-I3)
 Sheng, Susan Zhe, 47, 161 (2B-C4)
 Shi, Tianji, 38, 116 (1C-P)
 Shin, Hwashin H., 34, 102 (1C-C3)
 Shirani, Afsaneh, 41, 131 (1D-C2)
 Shokoohi, Farhad, 28, 78 (1B-C1)
 Shrier, Ian, 53, 185 (2D-I3)
 Sigal, Matthew, 39, 121 (1D-I2)
 Sikdar, Khokan C., 64, 230 (3D-C5)
- Simpson, Matthew, 59, 207 (3C-I2)
 Sinha, Sanjoy, 38, 120 (1D-I1)
 Skala, Matthew, 46, 155 (2B-C2)
 Small, Dylan, 27, 70 (1B-I2)
 Sohal, Vikaas, 27, 70 (1B-I2)
 Solymos, Peter, 39, 121 (1D-I2)
 Song, Peter X.-K., 29, 80 (1B-C2)
 Soudabeh, Shemehsavar, 37, 115 (1C-P)
 Spencer, Neil, 34, 101 (1C-C2)
 Spiegelman, Donna, 44, 144 (2B-V)
 Springford, Aaron, 37 (1C-D1)
 Stallard, Jim, 44, 145 (2B-I1)
 Starbuck, Bob, 39 (1D-I3)
 Steele, Russell, 46 (2B-C2), 53 (2D-I3) 61 (3C-C2), 155 (2B-C2), 185 (2D-I3), 215 (3C-C2)
 Stepanova, Natalia, 54, 187 (2D-I4)
 Stephens, David A., 31 (1C-I1), 46 (2B-C2) 85 (1C-I1), 156 (2B-C2)
 Stephens, Michael, 38, 117 (1C-P)
 Stevens, Forrest, 31, 87 (1C-I2)
 Stitelman, Ori, 60, 211 (3C-I4)
 Stoica, George, 35, 105 (1C-C4)
 Stone, Eric, 60, 210 (3C-I3)
 Stringam, Betsy Bender, 64 (3D-C6) 231 (3D-C6)
 Strumpf, Erin, 33, 97 (1C-C1)
 Struthers, Cynthia, 62, 220 (3D-C1)
 Stryhn, Henrik, 40, 129 (1D-C1)
 Su, Wanhua, 64, 230 (3D-C5)
 Sun, Zheng, 37 (1C-D2), 38 (1C-P) 117 (1C-P)
 Swartz, Tim, 51 (2C-C1), 52 (2C-C2) 175 (2C-C1), 179 (2C-C2)
 Sykes, Jenna, 42, 136 (1D-C3)
 Szafron, Duane, 26 (1A-A1)
 Szepesvari, Csaba, 45, 149 (2B-I3)
- Takahara, Glen, 28, 76 (1B-C1)
 Tang, Yang, 38, 117 (1C-P)
 Tate, Robert, 29, 80 (1B-C2)
 Tatsinkou, Joseph Francois Tagne, 52 (2C-C2) 178 (2C-C2)
 Taylor, Jonathan, 50, 170 (2C-I5)
 Tekougang, Thierry Chekouo, 29 (1B-C3) 83 (1B-C3)
 Teterina, Anastasia, 37 (1C-D1), 37 (1C-D2)
 Thabane, Lehana, 33 (1C-I5)
 Thompson, Mary, 35 (1C-C5), 56 (3B-I2) 59 (3C-A), 107 (1C-C5), 192 (3B-I2), 205 (3C-A)
 Thompson, Steve, 50 (2C-I3), 53 (2D-I1) 167 (2C-I3), 182 (2D-I1)

- Thorne, Jeffrey, 60, 210 (3C-I3)
 Thorpe, Lilian, 41, 132 (1D-C2)
 Tibshirani, Rob, 44 (2A-A), 50 (2C-I5) 143 (2A-A), 170 (2C-I5)
 Tibshirani, Ryan, 50, 170 (2C-I5)
 Tillé, Yves, 42, 138 (1D-C4)
 Tomal, Jabed Hossain, 36, 111 (1C-C6)
 Torabi, Mahmoud, 28 (1B-C1), 46 (2B-C1) 78 (1B-C1), 154 (2B-C1)
 Tortora, Cristina, 30, 83 (1B-C3)
 Trejo, Maria Esther Perez, 58 (3B-C3) 202 (3B-C3)
 Tremlett, Helen, 41, 131 (1D-C2)
 Tuzov, Ekaterina, 57, 201 (3B-C2)
- Vallée, Audrey-Anne, 36 (1C-D1)
 van der Laan, Mark, 60 (3C-I4) 211, 212 (3C-I4)
 van Eeden, Stephan, 61, 215 (3C-C2)
 VanLeeuwen, Dawn M., 64, 231 (3D-C6)
 Vieu, Philippe, 45, 147 (2B-I2)
 von Rosen, Dietrich, 33, 96 (1C-C1)
 Vrbik, Irene, 30, 84 (1B-C3)
 Vrbik, Jan, 51, 174 (2C-C1)
- Waddell, Adrian, 37 (1C-D1)
 Walker, Stephen, 32, 91 (1C-I4)
 Wang, Derek, 57, 197 (3B-C1)
 Wang, Huijing, 42, 139 (1D-C4)
 Wang, Jing, 43, 141 (1D-C5)
 Wang, Ruodu, 60, 213 (3C-C1)
 Wang, Xiaogang, 41, 132 (1D-C2)
 Ward, Lawrence, 53, 186 (2D-I4)
 Weiss, Lorelle, 49, 164 (2C-I2)
 Welch, William, 27 (1B-I3), 36 (1C-C6) 38 (1C-P), 42 (1D-C3), 72 (1B-I3), 111 (1C-C6), 116 (1C-P), 135 (1D-C3)
 Welsh, Alan, 39 (1D-I4), 52 (2D-I1) 124 (1D-I4), 181 (2D-I1)
 Westerhout, Cynthia M., 40 (1D-C1) 128 (1D-C1)
 Wharton, Sonia, 59, 207 (3C-I2)
 Wieczorek, Jerzy, 35, 107 (1C-C5)
 Wikle, Christopher, 26, 67 (1B-V)
 Wild, Chris, 43 (1D-C5), 53 (2D-I2) 142 (1D-C5)
 Williamson, Tyler, 47, 158 (2B-C3)
 Willmot, Gordon, 45, 151 (2B-I4)
 Winget, Marcy, 64, 230 (3D-C5)
 Withanage, Niroshan, 42, 135 (1D-C3)
 Wolfson, Julian, 50, 170 (2C-I5)
 Wong, Augustine, 63, 225 (3D-C3)
 Wong, Monica H.T., 61, 217 (3C-C2)
- Woods, Dave, 27, 72 (1B-I3)
 Woolford, Douglas, 34, 102 (1C-C3)
 Wu, Changbao, 28 (1B-I4), 35 (1C-C5) 75 (1B-I4), 107 (1C-C5)
 Wu, Jingjing, 63, 228 (3D-C4)
 Wu, Sicheng, 41, 132 (1D-C2)
 Wu, Yan Yan, 63, 225 (3D-C3)
 Wu, Ying, 63, 225 (3D-C3)
- Xiao, Jinkun, 38, 117 (1C-P)
 Xin, Lu, 38, 118 (1C-P)
 Xing, Nana, 47, 161 (2B-C4)
 Xu, Jianfeng, 40, 128 (1D-C1)
 Xu, Peirong, 33, 94 (1C-I6)
 Xu, Ronghui (Lily), 33, 94 (1C-I6)
 Xu, Xiaojian, 34, 99 (1C-C2)
 Xu, Ximing, 47, 159 (2B-C3)
 Xu, Yawen, 41, 132 (1D-C2)
 Xue, Lin, 38, 118 (1C-P)
- Yan, Xiang, 62, 223 (3D-C2)
 Yan, Ying, 58, 203 (3B-C3)
 Yang, Xiaojian, 41, 132 (1D-C2)
 Yapa, Gaitri, 44, 145 (2B-I1)
 Yasui, Yutaka, 64, 230 (3D-C5)
 Ye, Changqing, 56, 194 (3B-I3)
 Yi, Grace, 42 (1D-C4), 58 (3B-C3) 139 (1D-C4), 203 (3B-C3)
 Yi, Yanqing, 43, 142 (1D-C5)
 Yilmaz, Yildiz E., 33, 96 (1C-C1)
 Younger, Jaime, 40, 128 (1D-C1)
 Yousefi, Kasra, 52, 179 (2C-C2)
 Yu, Hao, 35 (1C-C4), 36 (1C-C6) 52 (2C-C2), 105 (1C-C4), 111 (1C-C6), 178 (2C-C2)
 Yuan, Yan, 64, 230 (3D-C5)
 Yuen, Wai Kong, 51, 174 (2C-C1)
- Zamar, Ruben H., 36 (1C-C6), 37 (1C-P) 111 (1C-C6), 115 (1C-P)
 Zarepour, Mahmoud, 28, 76 (1B-C1)
 Zhang, Biao, 61, 218 (3D-A1)
 Zhang, Xuechen, 41, 133 (1D-C2)
 Zhang, Ying, 59 (3C-I3), 62 (3D-C2) 209 (3C-I3), 222 (3D-C2)
 Zhang, Yukun, 36, 112 (1C-C6)
 Zhao, Yinshan, 41, 131 (1D-C2)
 Zheng, Yinngan, 40, 128 (1D-C1)
 Zhong, Guan, 61, 218 (3D-A1)
 Zhong, Yujie, 29, 81 (1B-C2)

Zhou, Xiaowen, 45 (2B-I4), 51 (2C-C1) 151 (2B-I4), 176
(2C-C1)

Zhu, Lixing, 33, 94 (1C-I6)

Zhu, Mu, 38 (1C-P), 56 (3B-I3) 64 (3D-C5), 118 (1C-P),
194 (3B-I3), 230 (3D-C5)

Zhu, Qiaohao, 58, 203 (3B-C3)

Zhu, Yunzhang, 56, 194 (3B-I3)

Zidek, James V., 34 (1C-C2), 37, 38 (1C-P) 42 (1D-C3),
99 (1C-C2), 115, 116 (1C-P), 135 (1D-C3)

Zolghadr, Maryam, 33, 98 (1C-C1)

Zou, Liwen, 60, 210 (3C-I3)

Zuyev, Sergei, 33, 98 (1C-C1)