

SSC Statistical Society of Canada
Société Statistique du Canada

40th Annual Meeting
of the
Statistical Society of Canada

40^e Congrès annuel
de la
Société statistique du Canada

June 3-6, 2012

3 au 6 juin 2012

University of Guelph

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	13
Social and Information Events • Événements sociaux et informatifs	15
List of Sessions • Liste des sessions	17
Program • Programme	21
Abstracts • Résumés	62
Author List • Liste des auteurs	253

Welcome • Bienvenue

The Department of Mathematics and Statistics, University of Guelph, takes great pride in welcoming you to the 40th Annual Meeting of the Statistical Society of Canada. The Department previously hosted the meetings in 1984, on the first occasion that the Society met independently of the Congress of Canadian Learned Societies.

The University of Guelph was founded in 1964, but the Ontario Veterinary College traces its history to 1862, making this its 150th anniversary. Our open, friendly campus, on a 330-hectare site, is characterized by green lawns and a blend of modern and traditional architecture. When you have a few moments, take a walk around the campus. The Arboretum on the eastern edge of the campus has extensive walking trails featuring a wide variety of flora and fauna. Or enjoy the galleries at the Macdonald Stewart Art Centre.

The City, with a population of 125,000, takes pride in having been named by Macleans as Canada's Safest City in 2009 and by MSN as one of Canada's most livable cities. It has also been named one of Canada's smartest cities and as Canada's volunteer capital. The Guelph Arts Council will offer the historical walking tour, *Slopes of the Speed*, Sunday June 3, at 2:00 pm, beginning at the Guelph Public Library (\$5 fee). This tour explores the slopes of the Speed River between Norwich and Macdonell Streets, where many of Guelph's largest and most distinguished historic homes are located. Up to the minute information about what's going on in the City can be found at <http://www.visitguelphwellington.ca/Modules/Pages/>

C'est un grand honneur pour le Département de mathématiques et de statistique de l'University of Guelph de vous accueillir au 40^e congrès annuel de la Société statistique du Canada. Le Département avait été l'hôte du congrès en 1984, la première année que la Société s'était réunie indépendamment du Congrès des sociétés savantes du Canada.

L'University of Guelph a été fondée en 1964, mais le Collège de médecine vétérinaire de l'Ontario, qui date de 1862, célèbre cette année son 150^e anniversaire. Ouvert et accueillant, notre campus de 330 hectares se distingue par ses grandes pelouses vertes et un mélange d'architecture moderne et traditionnelle. Lorsque vous aurez quelques instants de liberté, faites-y donc un tour. L'arboretum qui longe le côté est du campus propose de nombreux sentiers balisés d'où admirer la flore et la faune. Vous pourrez également visiter les galeries du musée Macdonald Stewart Art Centre.

La ville, forte de 125 000 habitants, se vante d'avoir été couronnée ville la plus sûre du Canada par Macleans en 2009 et l'une des villes canadiennes où il fait le mieux vivre par MSN. On l'a aussi couronnée l'une des villes les plus intelligentes du Canada et sa capitale du bénévolat. Le dimanche 3 juin, à 14 heures, le Guelph Arts Council propose une promenade historique, *Slopes of the Speed*, au départ de la bibliothèque municipale (frais : 5 \$). Cette visite explore les côtes de la rivière Speed entre les rues Norwich et Macdonell, où sont situées un grand nombre des plus grandes et plus illustres demeures historiques de la ville. Pour plus d'informations en temps réel sur les événements à Guelph, consultez <http://www.visitguelphwellington.ca/Modules/Pages/>

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 :

- Atlantic Association for Research in the Mathematical Sciences
- Centre de recherches mathématiques
- City of Guelph
- College of Physical and Engineering Science, University of Guelph
- Department of Mathematics and Statistics, University of Guelph
- Fields Institute for Research in Mathematical Sciences
- MINITAB
- Mitacs
- Pacific Institute for the Mathematical Sciences
- University of Guelph

Exhibitors • Exposants

The exhibitors, providing displays for examination and purchase, can be found in the Science Complex Atrium, near the morning and afternoon coffee breaks.

Les kiosques des exposants seront situés dans l'Atrium du Science Complex, près des pauses-café du matin et de l'après-midi.

- W. H. Freeman
- Minitab
- Pearson Canada
- SAS Institute Canada Inc
- Taylor and Francis
- Wiley

Organizers • Organismes

Local Arrangements Committee • Comité des arrangements locaux

Brian Allen (Chair • Président), Gary Umphrey (Vice-Chair • Vice-président)

Ayesha Ali, Jeremy Balka, Gerarda Darlington, Rob Deardon, Zeny Feng, Dan Gillis, Julie Horrocks, Paul McNicholas.

It would be impossible to organize an event of this magnitude without the help of many individuals. At the University of Guelph, Kathy Nahwegahbow and Lisa Tersigni-Holt of Conference Services have been indispensable over the past 5 years. Gertrud Jeewanjee and Alan Kelm, at the SSC Office, have provided the support, continuity and expertise we lacked in managing the electronic services, conference website, registration, exhibitors and sponsors. The 2011 local arrangements chair, Hugh Chipman, has been a fountain of information for us. The SSC Meetings Coordinator, Tim Swartz, his predecessor, Duncan Murdoch, and other executive members of the Society, have provided a great deal of valuable information, opinion and wisdom. Catherine Cox has provided precise translation for us, always in a very timely fashion. Finally, our crew of graduate students are on the ground providing AV support, manning the registration desk and providing information, all of this in bright Gryphon gold tee shirts.

Il est impossible d'organiser un événement de cette envergure sans l'aide de nombreux individus. À l'University of Guelph, Kathy Nahwegahbow et Lisa Tersigni-Holt des Services de conférences nous ont été d'une aide inestimable ces cinq dernières années. Gertrud Jeewanjee et Alan Kelm, du bureau de la SSC, nous ont apporté le soutien, la continuité et l'expertise dont nous manquons pour gérer les services électroniques, le site Web du congrès, l'inscription, les exposants et les commanditaires. Le président des arrangements locaux de 2011, Hugh Chipman, a été pour nous une mine de renseignements. Le coordonnateur des congrès, Tim Swartz, son prédécesseur, Duncan Murdoch, et les autres membres de l'exécutif de la Société nous ont offert des informations, des opinions et énormément de sagesse. Catherine Cox nous a fourni des traductions précises et en temps opportun. Enfin, notre équipe d'étudiants de cycle supérieur sont sur le terrain pour le soutien audiovisuel, assurer la permanence au bureau des inscriptions et renseigner les délégués : vous les reconnaîtrez à leurs chemises or Gryphon.

Program Committee • Comité du programme

Chair • Président: Alexandre Leblanc, *University of Manitoba*

Members • Membres:

Biostatistics Section • Groupe de biostatistique

Yutaka Yasui, *University of Alberta*

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

Boxin Tang, *Simon Fraser University*

Probability Section • Group de probabilité

Hao Yu, *University of Western Ontario*

Survey Methods Section • Groupe de méthodologie d'enquête





David Haziza, *Université de Montréal*





An important feature of our meetings is the presentation of the abstracts and of the plenary session visual aids in both official languages. This translation was very ably carried out under the supervision of the Bilingualism Committee (chaired by Jean-François Quessy) and the translators Anne-Sophie Charest, Catherine Cox, Pascale Giroux-Denis, Caroline Gras and Caroline Pelletier. Special thanks to Jean-François Plante (member of the bilingualism committee) and Pascale Giroux-Denis for their help with assembling and proofreading the program.

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 (présidé par Jean-François Quessy) par les traductrices Anne-Sophie Charest,

Catherine Cox, Pascale Giroux-Denis, Caroline Gras et Caroline Pelletier. Un grand merci à Jean-François Plante (membre du comité du bilinguisme) et à Pascale Giroux-Denis pour leur aide à la préparation et à la relecture du programme.

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 from Saturday at 4:00 pm, according to the following schedule:

- Saturday, June 2: 4:00 to 6:00 pm in the East Residence Lobby
- Sunday, June 3: 8:00 am to 4:00 pm in the Rozanski Hall Foyer
- Sunday, June 3: 6:00 pm to 8:00 pm in Science Complex Atrium
- Monday, June 4: 7:30 am to 5:00 pm in Science Complex Atrium
- Tuesday, June 5: 8:15 am to 5:00 pm in Science Complex Atrium
- Wednesday, June 6: 8:15 am to 3:00 pm in Science Complex Atrium

Vous pourrez vous inscrire sur place à partir du samedi à 16 heures, en respectant l'horaire suivant :

- Samedi, 2 juin : 16h00 à 18h00 dans le hall d'entrée de la résidence East
- Dimanche 3 juin : 8h00 à 16h00 dans l'entrée du Rozanski Hall
- Dimanche 3 juin : 18h00 à 20h00 dans l'Atrium du Science Complex
- Lundi 4 juin : 7h30 à 17h00 dans l'Atrium du Science Complex
- Mardi 5 juin : 8h15 à 17h00 dans l'Atrium du Science Complex
- Mercredi 6 juin : 8h15 à 15h00 dans l'Atrium du Science Complex

On-campus accommodation • Logement sur le campus

Those staying in campus residences or townhouses may check-in at East Residence Desk. The desk will be open 24 hours per day and may be reached by phone at 519-824-4120 ext 58121. For guests staying in residence, parking and breakfast are included in accommodation fees.

Les participants qui logent en résidence universitaire ou en appartement peuvent s'enregistrer au bureau de la résidence East. Ce bureau sera ouvert 24 heures sur 24 ; vous pouvez le joindre par téléphone au 519-824-4120, poste 58121. Pour les délégués qui logent en résidence, votre séjour inclut le stationnement et le petit-déjeuner.

Directions • Emplacement

Campus maps are available in the conference bag and on the inside back cover of this program. Those staying at the Best Western Royal Brock, Days Inn or Delta Guelph Hotel are encouraged to walk to campus. The Holiday Inn is a pleasant 25 minute walk to the centre of campus.

Des cartes du campus sont disponibles dans le sac du congressiste et en troisième de couverture de ce programme. Les délégués séjournant au Best Western Royal Brock, au Days Inn ou au Delta Guelph sont encouragés à venir à pied au campus. Le Holiday Inn est à 25 minutes de marche du centre du campus.

Parking on Campus • Stationnement sur le campus

Parking permits are required to park on campus Monday to Friday 8:00 to 17:00. Daily permits can be purchased for \$9.00 per day. These permit you to park in lots P13, P14, P15 or P19 and must be displayed on the dashboard. For your convenience upon initial arrival, permits will be available on Monday June 4 until 3:00 pm at lot P15. Permits will also be available at the Registration desk at all times. For those staying in Residence, parking passes are included in the accommodation rates.

Un permis de stationnement est obligatoire pour stationner sur le campus du lundi au vendredi, entre 8 heures et 17 heures. Vous pouvez vous procurer un permis quotidien au bureau des inscriptions pour 9 \$ par jour. Celui-ci vous autorise à stationner dans les stationnements P13, P14, P15 et P19 ; vous devez le placer bien en évidence sur votre tableau de bord. Pour votre commodité, les permis seront disponibles le lundi 4 juin, jusqu'à 15 heures, au stationnement P15. Vous pourrez aussi vous en procurer à tout moment au bureau des inscriptions. Pour les délégués logeant en résidence universitaire, le prix de la chambre inclut un permis de stationnement.

Campus Security • Sécurité sur le campus

The University Police can be reached at 519-824-4120, ext 52245. In an emergency, call 519-824-4120, ext 2000 (emergencies only).

Emergency poles: These poles are bright yellow and scattered across campus. If there is a medical emergency or you feel unsafe, all you need to do is press the button and the police dispatcher will respond. The blue emergency phones may also be used in the event of car trouble.

Parking Lots 13 and 19 are equipped with video surveillance 24-hours a day and linked to the University Police Dispatcher. These lots are also equipped with microphones to detect any noise that is out of the ordinary.

Vous pouvez joindre la Police de l'université au 519-824-4120, poste 52245. En cas d'urgence, composez le 519-824-4120, poste 2000 (urgences uniquement).

Poteaux d'urgence : Ces poteaux jaune vif sont dispersés à travers le campus. En cas d'urgence médicale ou si vous avez le sentiment d'être en danger, appuyez simplement sur le bouton et l'opérateur de la police vous répondra. Les téléphones bleus d'urgence peuvent aussi être utilisés en cas de panne de voiture.

Les stationnements 13 et 19 sont sous vidéosurveillance 24 heures sur 24 et reliés à l'opérateur de la police de l'université. Ils sont aussi équipés de microphones permettant de détecter tout bruit anormal.

Rooms • Salles

Scientific sessions will occur in Rozanski Hall, Alexander Hall, MacNaughton Building and the Thornbrough Building. Coffee breaks, exhibits and posters are found in the Science Complex Atrium, as well as the Welcome Reception. Lunches, which are included in registration Monday to Wednesday, are served in Creelman Hall. Most noontime business meetings will occur in the MacKinnon Building. The Monday evening barbeque will occur on Creelman Green, near Creelman Hall.

Les séances scientifiques se dérouleront au Rozanski Hall, au Alexander Hall, au MacNaughton Building et au Thornbrough Building. Vous trouverez les pauses-café, les exposants et les affiches dans l'atrium du Science Complex, tout comme la réception de bienvenue. Les repas de midi, qui sont inclus dans les frais d'inscription (du lundi au mercredi) seront servis au Creelman Hall. La plupart des réunions d'affaires de midi se dérouleront au MacKinnon Building. Lundi soir, le barbecue se tiendra sur la place Creelman, près du Creelman Hall.

Internet Access • Accès internet

A login and password for internet access over the wireless and wired University of Guelph networks can be found in your conference registration envelope. Desktop computers, available in the Library, do not require a login to access the internet. Printing and copying are available at Design & Print in the basement of the University Centre. Detailed instructions on accessing the networks and printing and copying, can be found in your conference bag.

Vous trouverez un nom d'utilisateur et un mot de passe pour l'accès Internet sur les réseaux wifi et câblés de l'University of Guelph dans votre enveloppe d'inscription au congrès. Les ordinateurs de bureau disponibles dans la bibliothèque ne nécessitent aucune connexion particulière. Un service d'impression et de copie est disponible chez Design & Print, dans le sous-sol de l'University Centre. Pour plus de détails sur l'accès aux réseaux et l'impression et la copie, consultez la fiche à ce sujet dans votre trousse du délégué.

Smoking Prohibited • Interdiction de fumer

Smoking is prohibited in all campus buildings and within nine 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 neuf 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

Lunch Monday to Wednesday is provided with registration and must be taken at Creelman Hall. Other venues throughout the campus, including Centre Six in the University Centre, provide food on a cash basis. The City boasts a wide selection of restaurants in all price ranges, representing many ethnic origins. The Visit Guelph and Area Guide 2012, found in your registration bag, provides a guide to many of these restaurants. Several restaurants, within walking distance of the University campus, may be found near the corner of Stone Rd and Edinburgh Rd and in the Campus Estates Plaza, at Harvard Road and Gordon Street. There is also a grocery store in the Harvard Estates Plaza.

Les repas de midi, qui sont inclus dans les frais d'inscription (du lundi au mercredi) doivent être pris au Creelman Hall. Vous pouvez aussi acheter à manger ailleurs sur le campus, notamment au Centre Six à l'University Centre. La ville regorge de restaurants dans toutes les catégories de prix et de toutes ethnicités. Le guide Visit Guelph and Area Guide 2012, que vous trouverez dans votre trousse de délégué, vous en présente un certain nombre. Vous en trouverez plusieurs à distance de marche du campus, près du coin des rues Stone et Edinburgh et dans le centre commercial Campus Estates, au coin des rues Harvard et Gordon. Un supermarché est situé au centre commercial Harvard Estates.

Use of Athletics Facilities • Utilisation des installations sportives

Visitors to the University of Guelph campus have access to the athletics facilities, including a swimming pool and weight room, with lockers and showers. The facilities are open Sunday 8:00 am to 8:00 pm and weekdays 6:30 am to 10:00 pm (weight room closes a half hour earlier). To use the facilities, go to the Front Desk of the Athletics Centre. The daily rate ranges from \$4 to \$7, depending on which facilities you wish to use.

Les visiteurs de l'University of Guelph ont accès aux installations sportives, notamment à la piscine et à la salle de musculation, ainsi qu'aux casiers et aux douches. Les installations sont ouvertes dimanche de 8 heures à 20 heures et les jours de semaine de 6 heures 30 à 22 heures (la salle de musculation ferme une demi-heure plus tôt). Pour utiliser les installations, présentez-vous à la réception du centre athlétique. Le tarif quotidien varie entre 4 \$ et 7 \$, selon les installations que vous souhaitez utiliser.

Workshops • Ateliers

Sunday, June 3, 9:00 - 4:00 pm • Dimanche 3 juin, 9h00 - 16h00 Rozanski Hall

Workshops organized by the sections will be held in Rozanski Hall, Sunday 9:00 am to 4:00 pm. Lunch is provided in the foyer of Rozanski. The Biostatistics Section workshop will be held in room 105, the Business and Industrial Statistics Section in room 107 and the Survey Methods Section in room 108.

Les ateliers organisés par les groupes se tiendront au Rozanski Hall, dimanche de 9h00 à 16h00. Le repas de midi sera servi dans le foyer du Rozanski Hall. L'atelier du groupe de biostatistique se tiendra dans la salle 105, celui du groupe de statistique industrielle et de gestion dans la salle 107 et celui du groupe des méthodes d'enquête dans la salle 108.

Blue Jays Game • Match des Blue Jays

Sunday, June 3, 10:30 am • Dimanche 3 juin, 10:30 Rozanski Hall

Those who have purchased a ticket for this event will find it in their registration envelope. See the Boston Red Sox play the Toronto Blue Jays at Rogers Centre. The package includes coach transportation, Premium 200 level seats and a premium deli lunch. Departure is at 10:30 am from Rozanski Hall. The bus will return immediately after the game in order to be back in Guelph in time for the welcome mixer.

Les délégués qui ont acheté un billet pour cet événement le trouveront dans leur enveloppe d'inscription. Assistez au match entre les Red Sox de Boston et les Blue Jays de Toronto au centre Rogers. L'excursion inclut le trajet en autocar, une place de choix au niveau 200 et un casse-croûte de qualité. Le car partira à 10h30 du Rozanski Hall et repartira dès la fin du match, si bien que vous serez de retour à Guelph pour la réception d'ouverture.

Information Session: NSERC Long-Range Planning Exercise in Mathematics and Statistics/Canadian Statistical Institute Development Committee • Exercice de planification à long terme en mathématiques et en statistique du CRSNG / Comité de développement d'un institut canadien de statistique

Sunday, June 3, 4:00 - 6:00 pm • Dimanche 3 juin, 16h00 - 18h00 Rozanski Hall, room/local 102

At this information session, two draft reports will be discussed, both related to the funding of statistical research in Canada. One is a report from NSERC's Steering Committee for the Long Range Plan for the Mathematical and Statistical Sciences, chaired by Nancy Reid. The other is a report from the Canadian Statistical Institute Development Committee, chaired by Mary Thompson. Nancy and Mary will give short presentations summarizing their reports and then open the floor for questions and feedback. Everyone is invited.

Lors de cette séance d'information, deux rapports préliminaires concernant le financement de la recherche statistique au Canada seront discutés. Il s'agit d'une part du rapport du comité de direction du plan à long terme pour les sciences mathématiques et statistiques du CRSNG, présidé par Nancy Reid. L'autre est un rapport du Comité de développement d'un institut canadien de statistique, présidé par Mary Thompson. Nancy et Mary donneront de courtes présentations résumant leurs rapports, puis inviteront les participants à formuler des questions et des commentaires. Tous les délégués sont les bienvenus.

Welcome Reception • Réception de bienvenue

Sunday, June 3, 6:00 - 8:00 pm • Dimanche 3 juin, 18h00 - 20h00 Science Complex Atrium

All conference participants are cordially invited to attend a welcome reception in the Atrium of the Science Complex, from 6:00 to 8:00 pm, on Sunday June 3. Tickets for two complementary beverages will be found in your registration envelope.

Les participants au congrès sont tous cordialement invités à une réception de bienvenue dans l'atrium du Science Complex, dimanche 3 juin de 18 heures à 20 heures. Vous trouverez des coupons donnant droit à deux consommations gratuites dans votre enveloppe d'inscription.

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, June 3, 7:30 pm • Dimanche 3 juin, 19h30 Shakespeare's Arms Restaurant

The SSC Committee on Women in Statistics and the Caucus for Women in Statistics have organized an informal supper at 7:30 pm, Sunday June 3, 2012 at Shakespeare's Arms (<http://www.shakespearearms.ca/about.htm>). Interested women and men are asked to email jhorrock@uoguelph.ca by May 28. 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 3 juin 2012 à 19h30 au Shakespeare's Arms (<http://www.shakespearearms.ca/about.htm>). Toutes les intéressées et tous les intéressés sont demandés d'envoyer un courriel à jhorrock@uoguelph.ca avant le 28 mai. Cet événement n'est pas inclus dans les frais d'inscription. En espérant vous y voir !

Lunch • Déjeuner

Lunch on Monday, Tuesday and Wednesday, is included in registration. You will receive a swipe card when you register, with credit for three meals. Lunch is served from 11:45 to 1:15 each day in Creelman Hall. Lunch-time committee meetings will occur in the scheduled rooms in the MacKinnon Building.

Les repas de midi du lundi, mardi et mercredi sont inclus dans les frais d'inscription. Lors de votre inscription, vous recevrez une carte magnétique créditée de trois repas. Le dîner est servi de 11h45 à 13h15 chaque jour au Creelman Hall. Les réunions d'affaires de midi se dérouleront dans les salles indiquées du MacKinnon Building.

Poster Sessions • Séances d'affichage

All poster sessions will take place in the Science Complex Atrium. Case Studies I and II and Poster Session I will occur in Monday, June 4 from 1:15 to 3:15 pm. Poster session II will take place on Tuesday, June 5 from 1:15 to 3:15 pm. The authors will be with their posters until the end of coffee break, at 3:15. Instructions for poster presenters can be found on the meeting website.

Les séances d'affichage se tiendront dans l'atrium du Science Complex. Les études de cas I et II et la séance d'affiches I se tiendront lundi 4 juin de 13h15 à 15h15. La séance d'affiches II se tiendra mardi 5 juin de 13h15 à 15h15. Les auteurs seront présents jusqu'à la fin de la pause café à 15h15. Les présentateurs trouveront plus d'instructions sur le site Web du congrès.

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

Monday, June 4, 3:15 - 4:45 • Lundi 4 juin, 15h15 - 16h45 MacKinnon Building, room/local 120

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.

Crawford Lake, Iroquois Longhouses and Mohawk Racetrack • Lac Crawford, maisons longues iroquoises et circuit de courses Mohawk

Monday, June 4, 5:50 • Lundi 4 juin, 17h50 Rozanski Hall

Those who have purchased tickets for this event will find them in their registration envelope. The bus will depart from Rozanski Hall at 5:50 pm. At Crawford Lake, you will be given a guided tour of the reconstructed Iroquois longhouses. Then its on to Mohawk racetrack where you will enjoy a premium buffet dinner while watching harness racing in action. To test your stochastic and risk management skills, you will receive a race program and a \$5 gaming coupon for the slots.

Les délégués qui ont acheté des billets pour cet événement les trouveront dans leur enveloppe d'inscription. Le bus quittera Rozanski Hall à 17h50. Au lac Crawford, vous participerez à une visite guidée des maisons longues iroquoises reconstruites. Ensuite vous vous rendrez au circuit de courses Mohawk où vous dégusterez un buffet d'înatore tout en regardant les courses attelées. Pour tester vos connaissances des probabilités et de la gestion du risque, vous recevrez le programme de courses et un coupon de jeu de 5 \$ pour les machines à sous.

Barbeque • Barbecue

Monday, June 4, 6:00 - 8:00 pm • Lundi 4 juin, 18h00 - 20h00 Creelman Green

A barbeque will be held on Creelman Green, near Creelman Hall (rain location Creelman Hall), from 6:00 pm until 8:00 pm, Monday, June 4. All graduate students, and anyone else who has paid for this meal, will find a ticket in their registration envelope.

Un barbecue est organisé sur la place Creelman près du Creelman Hall (ou dans ce hall en cas de pluie), lundi 4 juin 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.

UWaterloo Math Alumni Reception • Réception des anciens étudiants en maths de la UWaterloo

Monday, June 4, 6:00 - 8:00 pm • Lundi 4 juin, 18h00 - 20h00 The Brass Taps

A reception for all Faculty of Mathematics alumni will be held in The Brass Taps from 6 - 8 pm Monday, June 4. To register, please contact Kristine McGlynn at kmcglynn@uwaterloo.ca or call 519-888-4567 ext. 37747.

Une réception est organisée pour tous les anciens étudiants de la Faculté de mathématiques aux Brass Taps de 18 heures à 20 heures, lundi 4 juin. Pour vous inscrire, veuillez contacter Kristine McGlynn à kmcglynn@uwaterloo.ca ou par téléphone au 519-888-4567, poste 37747.

University of Guelph Alumni Reception • Réception des anciens étudiants de l'University of Guelph

Monday, June 4, 7:00 pm • Lundi 4 juin, 19h00 Bullring Pub

The Department of Mathematics and Statistics and the College of Physical and Engineering Science will host a reception for all University of Guelph Alumni and friends at the Bullring Pub, Monday, June 4, beginning at 7:00 pm.

Le Département de mathématiques et de statistique et le Collège de sciences physiques et de génie organisent une réception à l'intention des anciens étudiants de l'University of Guelph et de leurs amis au Bullring Pub, lundi 4 juin à partir de 19 heures.

Information on 2012 NSERC Discovery Grant Competition • Information sur le concours 2012 des subventions à la découverte du CRSNG

Tuesday, June 5, 4:50 - 5:50 pm • Mardi 5 juin, 16h50 - 17h50 Rozanski Hall, room/local 102

A representative from NSERC will make a presentation to inform researchers on the peer review process for NSERC Discovery Grants. Summary statistics on the 2012 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 2012 seront fournies. Tous les chercheurs sont les bienvenus.

Banquet

Tuesday, June 5, 6:30 pm • Mardi 5 juin, 18h30 W. F. Mitchell Athletic Centre

The banquet will be held in the Athletics Centre, with cocktails and a cash bar at 6:30 pm and dinner service beginning at 7:15 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.

Le banquet se tiendra au centre athlétique, avec un coquetel et bar payant à partir de 18h30, le souper étant servi à partir de 19h15. Tous les participants au congrès qui n'ont pas décliné leur participation au banquet retrouveront dans leur enveloppe d'inscription un ticket qu'ils devront placer devant eux à leur table.

Committees and Meetings • Comités et réunions

Saturday June 2

samedi 2 juin

18:00 - 22:00 Executive Committee 1 / Comité exécutif 1
Off Campus - Delta Hotel / Hors campus - Hôtel Delta

Sunday June 3

dimanche 3 juin

9:30 - 12:00 Meeting of Heads, Chairs, Directors and Representatives of Departments / Rencontre des directeurs et représentants des départements
Rozanski Hall, room/local 109

10:30 - 12:00 Publications Committee / Comité des publications
University Centre, room/local 334

12:00 - 16:00 Board of Directors 1 / Conseil d'administration 1
University Club (University Centre, 5th floor) / University Club (University Centre, 5e étage)

Monday June 4

lundi 4 juin

11:45 - 13:00 Business and Industrial Statistics Executive Committee / Comité exécutif du groupe de statistique industrielle et de gestion
MacKinnon Building, room/local 311

11:45 - 13:00 Biostatistics Section Executive Committee 1 / Comité exécutif du groupe de biostatistique 1
MacKinnon Building, room/local 310

11:45 - 13:00 Canadian Statistical Institute Development Committee / Comité de développement d'un institut canadien de statistique
MacKinnon Building, room/local 312

11:45 - 13:00 Probability Section Executive Committee 1 / Comité exécutif du groupe de probabilité 1
MacKinnon Building, room/local 309

11:45 - 13:00 Statistical Education Committee / Comité d'éducation en statistique
MacKinnon Building, room/local 308

11:45 - 13:00 Committee on Women in Statistics / Comité de promotion de la femme en statistique
MacKinnon Building, room/local 307

16:50 - 17:50 SSC Annual General Meeting / Assemblée générale annuelle de la SSC
Rozanski Hall, room/local 104

Tuesday June 5

mardi 5 juin

11:45 - 13:00 CJS Editorial Board / Comité de rédaction de la RCS
MacKinnon Building, room/local 311

11:45 - 13:00 Bilingualism Committee / Comité du bilinguisme
MacKinnon Building, room/local 309

11:45 - 13:00 Public Relations Committee / Comité des relations publiques
MacKinnon Building, room/local 310

11:45 - 13:00 Research Committee / Comité de la recherche
MacKinnon Building, room/local 307

11:45 - 13:00 Accreditation General Meeting / Assemblée générale des membres accrédités
MacKinnon Building, room/local 312

12:00 - 18:00 Award for Case Studies in Data Analysis Committee / Comité du prix pour les études de cas en analyse de données
Science Complex, room/local 1301

16:00 - 18:30	Student Research Presentation Award Committee / Comité du prix pour la meilleure présentation de recherche étudiante <i>MacKinnon Building, room/local 308</i>
16:50 - 17:50	Business and Industrial Statistics Section General Meeting / Assemblée générale du groupe de statistique industrielle et de gestion <i>MacKinnon Building, room/local 311</i>
16:50 - 17:50	Biostatistics Section General Meeting / Assemblée générale du groupe de biostatistique <i>MacKinnon Building, room/local 310</i>
16:50 - 17:50	Probability Section General Meeting / Assemblée générale du groupe de probabilité <i>MacKinnon Building, room/local 309</i>
16:50 - 17:50	Survey Methods Section General Meeting / Assemblée générale du groupe de méthodes d'enquête <i>MacKinnon Building, room/local 312</i>

Wednesday June 6**mercredi 6 juin**

11:45 - 13:00	Accreditation Committee / Comité d'accréditation <i>MacKinnon Building, room/local 312</i>
11:45 - 13:00	Biostatistics Section Executive Committee 2 / Comité exécutif du groupe de biostatistique 2 <i>MacKinnon Building, room/local 310</i>
11:45 - 13:00	Probability Section Executive Committee 2 / Comité exécutif du groupe de probabilité 2 <i>MacKinnon Building, room/local 309</i>
11:45 - 13:00	Program Committee / Comité des congrès <i>MacKinnon Building, room/local 311</i>
15:00 - 16:30	Board of Directors 2 / Conseil d'administration 2 <i>University Club (University Centre, 5th floor) / University Club (University Centre, 5e étage)</i>
16:30 - 17:30	Executive Committee 2 / Comité exécutif 2 <i>University Club (University Centre, 5th floor) / University Club (University Centre, 5e étage)</i>

Social and Information Events • Événements sociaux et informatifs

Sunday June 3

dimanche 3 juin

10:30 - 17:30	Blue Jays vs. Red Sox at Rogers Centre - Fee event / Blue Jays / Red Sox au Centre Rogers - Événement payant <i>Off campus / Hors campus</i>
16:00 - 18:00	NSERC Long-Range Planning Exercise in Mathematics and Statistics/Canadian Statistical Institute Development Committee / Exercice de planification à long terme en mathématiques et en statistique du CRSNG/Comité de développement d'un institut canadien de statistique <i>Rozanski Hall, room/local 102</i>
18:00 - 20:00	Welcome Reception / Réception de bienvenue <i>Science Complex Atrium</i>
19:30	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>Off campus - Shakespeare's Arms / Hors campus - Shakespeare's Arms</i>

Monday June 4

lundi 4 juin

9:45 - 10:15	Coffee Break / Pause-café <i>Science Complex Atrium</i>
11:45 - 13:15	Lunch / Dîner <i>Creelman Hall</i>
14:45 - 15:15	Coffee Break / Pause-café <i>Science Complex Atrium</i>
15:15 - 16:45	NSERC Discovery Grant Workshop / Atelier des subventions à la découverte du CRSNG <i>MacKinnon Building, room/local 120</i>
17:50	Crawford Lake, Iroquois Village and Mohawk Raceway - Fee event / Lac Crawford, village Iroquois et circuit de courses Mohawk - Événement payant <i>Off campus / Hors campus</i>
18:00 - 20:00	Student BBQ - Free for graduate students / Barbecue des étudiants - Gratuit pour les étudiants de cycle supérieur <i>Creelman Green</i>
18:00 - 20:00	UWaterloo Math Alumni Reception / Réception des anciens étudiants en maths de UWaterloo <i>The Brass Taps - Registration required / The Brass Taps - Inscription nécessaire</i>
19:00	University of Guelph Alumni Evening Reception / Réception en soirée pour les anciens étudiants de l'University of Guelph <i>Bullring Pub</i>

Tuesday June 5

mardi 5 juin

9:45 - 10:15	Coffee Break / Pause-café <i>Science Complex Atrium</i>
11:45 - 13:15	Lunch / Dîner <i>Creelman Hall</i>
14:45 - 15:15	Coffee Break / Pause-café <i>Science Complex Atrium</i>
16:50 - 17:50	Information on 2012 NSERC Discovery Grant Competition / Information sur le concours 2012 des subventions à la découverte du CRSNG <i>Rozanski Hall, room/local 102</i>

18:30 - 19:15 Banquet - Cocktails and Cash Bar / Banquet - Bar payant
W. F. Mitchell Athletic Centre

19:15 Banquet - Dinner Service / Banquet - Service du repas
W. F. Mitchell Athletic Centre

Wednesday June 6**mercredi 6 juin**

9:45 - 10:15 Coffee Break / Pause-café
Science Complex Atrium

11:45 - 13:15 Lunch / Dîner
Creelman Hall

List of Sessions • Liste des sessions

- 1A-A1 Inaugural Session
Séance inaugurale
- 1A-A2 SSC Presidential Invited Address
Allocution de l'invité du Président de la SSC
- 1B-V Nonparametric Bayes Methods: an Overview
Statistique bayésienne nonparamétrique: un survol
- 1B-I1 Extending Mixtures of Factor Analyzers
Généralisations des mélanges avec covariances d'analyse factorielle
- 1B-I2 Responsive Designs
Conception adaptée de plans d'échantillonnage
- 1B-I3 Statistical Inference for Differential Equation Models
Inférence statistique pour les modèles d'équations différentielles
- 1B-I4 Uncertainty Quantification in Analysis of Computer Models
Quantifier l'incertitude dans l'analyse des modèles informatiques
- 1B-C1 Biostatistics: Modelling and Study Design
Biostatistique: modélisation et planification d'étude
- 1B-C2 Frailty Models and the Analysis of Recurrent Events
Modèles de fragilité et analyse d'événements récurrents
- 1B-C3 Predictive Models and Survey Methods
Modèles de prédiction et méthodes d'enquête
- 1B-C4 Times Series Models
Modèles de séries chronologiques
- 1C-A SSC Impact Award Address
Allocution du récipiendaire du prix pour impact de la SSC
- 1C-I1 Asymptotic Methods in Stochastics: a Session in Honour of Miklós Csörgő's 80th Birthday
Séance en l'honneur du 80e anniversaire de Miklós Csörgő
- 1C-I2 Inference and Applications using Estimating Functions
Inférence à l'aide de fonctions d'estimation et applications
- 1C-I3 Modern Multiple Imputation Approaches to Missing Data Problems
Approches modernes à l'imputation multiple pour les problèmes de données manquantes
- 1C-I4 Reflections on 40 Years of Teaching Statistics
Réflexions sur 40 ans d'enseignement de la statistique
- 1C-C1 Analysis of Longitudinal Data
Analyse de données longitudinales
- 1C-C2 Estimating Population Size and Growth
Estimation de la taille et de la croissance des populations
- 1C-C3 Generalized Linear Models, Overdispersion and Zero-Inflation
Modèles linéaires généralisés, inflation de la variance et surabondance de zéros
- 1C-C4 Methods for High-Dimensional Data/Data Reduction
Analyse de données en plusieurs dimensions
- 1C-C5 Modern Regression Methods
Méthodes modernes de régression
- 1C-D1 Case Study I: Mark-Recapture on Atlantic Cod
Études de cas I: Marquage-recapture de la morue de l'Atlantique
- 1C-D2 Case Study II: Risk Factors for Osteoporotic Fracture
Étude de cas II: Facteurs de risque de fracture ostéoporotique
- 1C-P Poster Session I
Séance d'affichage I
- 1D-I1 Analysis of Interval-censored Data
Analyse de données censurées par intervalles

- 1D-I2 Celebrating Our 40th Anniversary
Célébrons nos 40 ans
- 1D-I3 Linear and Nonlinear Time Series
Modèles linéaires et non linéaires de séries chronologiques
- 1D-I4 Recent Advances in Small Area Estimation
Progrès récents en estimation pour les petits domaines
- 1D-I5 Statistical Methods for Clinical Trials and Causal Inference
Méthodes statistiques pour les essais cliniques et l'inférence causale
- 1D-C1 Density Approximation and Estimation
Approximation et estimation de densité
- 1D-C2 Impact and Comparison of Cancer Screening Strategies
Impact et comparaison des examens préventifs pour le cancer
- 1D-C3 Monte Carlo Methods and Inference for Stochastic Processes
Méthodes de Monte Carlo et inférence pour les processus stochastiques
- 1D-C4 Optimal Design
Planification optimale d'expériences
- 1D-C5 Survival Analysis and Accelerated Life Testing
Analyse de survie et tests accélérés de durée de vie
- 2A-A Gold Medal Address
Allocution du récipiendaire de la médaille d'or
- 2B-V An Overview of Models and Methods for Spatio-temporal Data Analysis
Un survol des modèles et méthodes d'analyse de données spatio-temporelle
- 2B-I1 Applications of Statistics in Medicine
Applications de la statistique à la médecine
- 2B-I2 On Assessing Liking and Intensity in Sensory Science
Évaluation du goût et de l'intensité en sciences sensorielles
- 2B-I3 Outlier-Robust Estimation in Survey Sampling
Estimation robuste et méthodes d'échantillonnage
- 2B-I4 Trend Analysis of Time Series Data and Applications
Analyse de tendances en séries chronologiques et applications
- 2B-C1 Biostatistics: Inference Methods
Biostatistique: méthodes inférentielles
- 2B-C2 Hazard Models and Censored Data
Modèles de risque et données censurées
- 2B-C3 Hypothesis Testing and R Statistical Software
Tests d'hypothèses et logiciel statistique R
- 2B-C4 Model and Variable Selection
Sélection de modèle et de variables
- 2B-C5 Theory and Applications of Copulas
Théorie et applications des copules
- 2C-A Isobel Loutit Invited Address
Allocution Isobel Loutit
- 2C-I1 Absolute Penalty Estimation and Variable Selection
Estimation à l'aide de la perte absolue et sélection de variables
- 2C-I2 New Statistical Design and Analysis for Epidemiological Investigations
Nouvelles méthodes de conception et d'analyse pour les enquêtes épidémiologiques
- 2C-I3 Session in Honour of Jon N.K. Rao's 75th Birthday
Séance en l'honneur du 75e anniversaire de Jon N.K. Rao
- 2C-I4 Statistical Analysis of Network Graph Data
Analyse statistique de réseaux et graphes
- 2C-I5 Stochastic Methods in Finance
Méthodes stochastiques en finance
- 2C-C1 Models Accounting for Measurement Error
Modèles avec erreurs de mesure

- 2C-C2 Nonparametric Methods
Méthodes non-paramétriques
- 2C-C3 Spatial Statistics
Statistique spatiale
- 2C-C4 Studying Gene Expression/Association
Étude de l'expression et de l'association des gènes
- 2C-P Poster Session II
Séance d'affichage II
- 2D-I1 Nonparametric Bayes Methods
Méthodes bayésiennes non paramétriques
- 2D-I2 Recent Advances in Analysis of Survival Data from Cross Sectional Sampling
Progrès récents en analyse de survie à partir de données transversales
- 2D-I3 Teaching and Learning Statistics with Technology
Enseignement et apprentissage de la statistique à l'aide des nouvelles technologies
- 2D-I4 The Census of Population in Canada: Looking Back at 2011 and Into the Future
Le Recensement de la population au Canada : Retour sur 2011 et regard vers l'avenir
- 2D-I5 Theory and Applications of Self-Normalized Processes
Théorie et application des processus auto-normalisés
- 2D-C1 Business and Industrial Statistics and Quality Control
Statistique industrielle et de gestion et contrôle de la qualité
- 2D-C2 Financial Applications
Applications en finance
- 2D-C3 Mixture Models, Clustering and Classification
Modèles de mélanges, partitionnement et classification
- 2D-C4 Modelling the Spread of Infectious Diseases
Modélisation de la propagation des maladies infectieuses
- 3A-A1 CJS Award Address
Allocution du récipiendaire du Prix de la RCS
- 3A-A2 CRM-SSC Prize in Statistics Address
Allocution du récipiendaire du Prix CRM-SSC
- 3A-A3 Pierre Robillard Award Address
Allocution du récipiendaire du Prix Pierre-Robillard
- 3B-I1 Confidentiality
Confidentialité
- 3B-I2 Current Issues in the Design and Analysis of Cluster Randomization Trials
Problèmes actuels dans la conception et l'analyse d'essais randomisés par grappes
- 3B-I3 Spatial and Environmental Statistics
Statistique spatiale et environnementale
- 3B-I4 Stochastic Models and Applications
Modèles stochastiques et applications
- 3B-I5 Teaching Statistical Consulting with a Mixture of Classroom and Practicum
Enseignement de la consultation statistique: un mélange d'exposés magistraux et d'expériences pratiques
- 3B-C1 Clustering and Classification
Méthodes de partitionnement et de classification
- 3B-C2 Mixed Effects Models
Modèles à effets mixtes
- 3B-C3 Models with Binary Responses
Modèles à variables indépendantes binaires
- 3B-C4 Spectral Methods in Time series
Analyse spectrale de séries chronologiques
- 3C-A Survey Methods Section Presidential Invited Address: Empirical Likelihood in Survey Sampling
Allocution de l'invité du Président du Groupe des méthodes d'enquêtes: Vraisemblance empirique et échantillonnage

- 3C-I1 Causality and Graphical Markov Models in High Dimensions
Causalité et modèles graphiques markoviens en plusieurs dimensions
- 3C-I2 Copula Models
Modèles de copules
- 3C-I3 Design and Analysis of Experiments in Modern Agriculture
Conception et analyse d'expériences en agriculture moderne
- 3C-I4 Robustness for Complex Data
Robustesse et données complexes
- 3C-C1 Financial Applications and Heavy-Tailed Distributions
Applications en finance et distributions à ailes relevées
- 3C-C2 Functional Data, Principal Components and Clustering
Données fonctionnelles, composantes principales et méthodes de partitionnement
- 3C-C3 Methods for Genetic Association
Méthodes d'association génétique

Program • Programme

Sunday June 3

dimanche 3 juin

9:00 - 16:00

Rozanski Hall, room/local 105

Workshop of the Biostatistics Section: Emerging Issues in Analysis of Longitudinal Data

Atelier du Groupe de biostatistique : Problématiques émergentes dans l'analyse des données longitudinales

Organizer and Chair / Responsable et président: Yutaka Yasui (University of Alberta)

9:00 - 16:00

Paul Rathouz (University of Wisconsin)

9:00 - 16:00

Rozanski Hall, room/local 107

Workshop of the Business and Industrial Statistics Section: Introduction to Bayesian Statistics and Use of WinBUGS

Atelier du Groupe de statistique industrielle et de gestion : Introduction à la statistique bayésienne et à l'utilisation de WinBUGS

Organizer and Chair / Responsable et président: Boxin Tang (Simon Fraser University)

9:00 - 16:00

Paramjit Gill (University of British Columbia, Okanagan)

9:00 - 16:00

Tim Swartz (Simon Fraser University)

9:00 - 16:00

Rozanski Hall, room/local 108

Workshop of the Survey Methods Section: A Primer on Capture-Recapture Models

Atelier du Groupe des méthodes d'enquête : Une introduction aux modèles de capture-recapture

Organizer and Chair / Responsable et président: David Haziza (Université de Montréal)

9:00 - 16:00

Louis-Paul Rivest (Université Laval)

16:00 - 18:00

Rozanski Hall, room 102

NSERC Long-Range Planning Exercise in Mathematics and Statistics/Canadian Statistical Institute Development Committee

Exercice de planification à long terme en mathématiques et en statistique du CRSNG/Comité de développement d'un institut canadien de statistique

Monday June 4**lundi 4 juin****8:15 - 8:30 Session 1A-A1 Rozanski Hall, room/local 104****Inaugural Session
Séance inaugurale**

Special Session / Conférence spéciale

Chair/Président: Brian Allen (University of Guelph)

Organizers/Responsables: Brian Allen (University of Guelph) and/et John Brewster (University of Manitoba)

8:15 - 8:20 Kevin Hall (Vice-President Research, University of Guelph / Vice-recteur à la recherche, University of Guelph)

8:20 - 8:25 John Brewster (SSC President / Président de la SSC)

8:25 - 8:30 Brian Allen (SSC2012 Local Arrangements Chair / Président du comité des arrangements locaux, SSC 2012)

8:30 - 9:45 Session 1A-A2 (abstract/résumé - page 63) Rozanski Hall, room/local 104**SSC Presidential Invited Address****Allocution de l'invité du Président de la SSC**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: John Brewster (University of Manitoba)

8:30 - 9:45 James O. Berger (Duke University)
Reproducibility and Multiplicity / Reproductibilité et multiplicité**9:45 - 10:15 Science Complex Atrium****Coffee Break
Pause-café****10:15 - 11:45 Session 1B-V (abstract/résumé - page 64) Rozanski Hall, room/local 104****Nonparametric Bayes Methods: an Overview****Statistique bayésienne nonparamétrique : un survol**

Introductory Overview Lecture / Conférence de vue d'ensemble



Organizer and Chair / Responsable et président: Alexandre Leblanc (University of Manitoba)

10:15 - 11:45 Steve MacEachern (Ohio State University)
Nonparametric Bayes Methods: An Overview / Méthodes bayésiennes non-paramétriques :
une vue d'ensemble**10:15 - 11:45 Session 1B-I1 (abstracts/résumés - page 65) Alexander Hall, room/local 200****Extending Mixtures of Factor Analyzers****Généralisations des mélanges avec covariances d'analyse factorielle**

Invited Paper Session / Présentations sur invitation

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

10:15 - 10:45 Jeffrey Andrews & Paul McNicholas (University of Guelph)
Mixtures of Modified t -Factor Analyzers / Mélanges de lois de Student modifiées avec covariance d'analyse factorielle10:45 - 11:15 Jangsun Baek (Chonnam National University, South Korea)
Applications of Various Mixtures of Common Factor Analyzers to Clustering of High-Dimensional Data / Application de divers mélanges à covariance d'analyse factorielle pour l'analyse de regroupement de données de grande dimension

- 11:15 - 11:45 Salvatore Ingrassia (Università di Catania, Italy), Paul McNicholas (University of Guelph), Antonio Punzo (Università di Catania, Italy), Sanjeena Subedi (University of Guelph)
Cluster Weighted Factor Analyzers / Lois à covariance d'analyse factorielle pour modèles de grappes pondérées  

10:15 - 11:45 Session 1B-I2 (abstracts/résumés - page 67) **Rozanski Hall, room/local 103**

Responsive Designs

Conception adaptée de plans d'échantillonnage





Invited Paper Session / Présentations sur invitation

Chair/Président: David Haziza (Université de Montréal)

Organizer/Responsable: Wesley Yung (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 10:15 - 10:45 Pete Brodie (Office for National Statistics, U.K.)
Developing a Response Chasing Strategy for Business Surveys within the Office for National Statistics / Développement d'une stratégie de rappel pour les enquêtes auprès des entreprises à l'Office for National Statistics  
- 10:45 - 11:15 Daniell Toth & Polly Phipps (U.S. Bureau of Labor Statistics), Nick Horton (Smith College)
Analyzing Establishment Nonresponse Using an Interpretable Regression Tree Model with Linked Administrative Data / Analyse de non-réponse d'établissements à l'aide d'un arbre de régression interprétable avec des données administratives liées  
- 11:15 - 11:45 Mike Hidiroglou, Discussant (Statistics Canada / Statistique Canada)







10:15 - 11:45 Session 1B-I3 (abstracts/résumés - page 69) **Alexander Hall, room/local 100**

Statistical Inference for Differential Equation Models

Inférence statistique pour les modèles d'équations différentielles

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Jiguo Cao (Simon Fraser University)

- 10:15 - 10:45 Michael Dowd (Dalhousie University)
Data Assimilation for Large-Scale Spatio-Temporal Dynamic Models / Assimilation de données pour modèles dynamiques spatio-temporels à grande échelle  
- 10:45 - 11:15 Giles Hooker & Liefur Thorbergsson (Cornell University), Kevin Lin (Arizona State University), Bruce Rogers (SAMSU)
Control Theory and the Design of Dynamic Experiments / Théorie de la régulation et conception d'expériences dynamiques  
- 11:15 - 11:45 Aaron A. King (University of Michigan)
POMP: Statistical Inference for Stochastic Dynamical Systems / Processus markovien partiellement observable : l'inférence statistique pour les systèmes stochastiques dynamiques  

10:15 - 11:45 Session 1B-I4 (abstracts/résumés - page 71) **Rozanski Hall, room/local 101**

Uncertainty Quantification in Analysis of Computer Models







Quantifier l'incertitude dans l'analyse des modèles informatiques

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: William Welch (University of British Columbia)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 10:15 - 10:45 Dorin Drignei (Oakland University)
Sensitivity Analysis of a Dynamical Computer Model for Vehicle Suspension Systems / Analyse de sensibilité d'un modèle informatique dynamique pour les systèmes de suspension des véhicules  
- 10:45 - 11:15 Joslin Goh & Derek Bingham (Simon Fraser University), James Paul Holloway, Carolyn C. Kuranz, Michael J. Grosskopf, Erica Rutter & Forrest W. Doss (University of Michigan)
Tuning with Multiple Emulators of Different Fidelities / Ajustement au moyen de plusieurs émulateurs de niveau de fidélité différent  
- 11:15 - 11:45 Jason Loeppky (University of British Columbia, Okanagan), William Welch (University of British Columbia), Derek Bingham (Simon Fraser University)
Quantification of Computer Model Bias / Quantification du biais d'un modèle informatique  

10:15 - 11:45 Session 1B-C1 (abstracts/résumés - page 73) **MacNaughton Building, room/local 105**

Biostatistics: Modelling and Study Design

Biostatistique : modélisation et planification d'étude

Contributed Paper Session / Séance de communications libres











Chair/Présidente: Yanqing Yi (Memorial University of Newfoundland)

- 10:15 - 10:30 Lehana Thabane (McMaster University), Janusz Kaczorowski (Université de Montréal), Lisa Dolovich (McMaster University), Larry Chambers (University of Ottawa), on behalf of CHAP Investigators
Dealing with Criticisms and Controversies of Pragmatic Trials / Traitement des critiques et des controverses entourant les essais pragmatiques  
- 10:30 - 10:45 Shiyuan (William) Zhang & Lehana Thabane (McMaster University)
Critical Appraisal and Application of Different Statistical Methods of Analyzing Continuous Outcomes in RCTs / Appréciation critique et application de différentes méthodes d'analyse de données continues d'ECR  
- 10:45 - 11:00 Meaghan S. Cuerden, Cecilia Cotton & Richard J. Cook (University of Waterloo)
Directed Donor Transfusions in Infants Admitted to the Neonatal Intensive Care Unit / Transfusions par don dirigé pour nourrissons admis à l'unité de soins intensifs néonataux  
- 11:00 - 11:15 Olli Saarela, Erica Moodie & David Stephens (McGill University)
On Bayesian Estimation of Marginal Structural Models / De L'estimation bayésienne des modèles structurels marginaux  
- 11:15 - 11:30 Ramon Clèries & Josepa Ribes (Cancer Registry of Catalonia and University of Barcelona, Spain / Cancer Registry of Catalonia et University of Barcelona, Spain), Jose Miguel Martinez (CISAL / GREDS-EMCONET - University Pompeu Fabra, Spain), Yutaka Yasui (University of Alberta)
Decomposing Changes in Disease Incidence and Mortality Counts into Overall and Age-Specific Changes in Risk and Population Size / Décomposition des changements de l'incidence d'une maladie et de la mortalité en changements globaux et spécifiques à l'âge du risque et de la taille de la population  
- 11:30 - 11:45 Qing Guo, Lehana Thabane, Geoffrey Hall, Ron Goeree, Margaret McKinnon & Eleanor M. Pullenayegum (McMaster University)
The Reporting of Sample Size Calculations in Functional Magnetic Resonance Imaging Studies: A Systematic Review / Publication des calculs de la taille d'échantillon dans les études d'imagerie par résonance magnétique fonctionnelle : un examen systématique  

10:15 - 11:30 Session 1B-C2 (abstracts/résumés - page 77) **MacNaughton Building, room/local 113**
Frailty Models and the Analysis of Recurrent Events**Modèles de fragilité et analyse d'événements récurrents**

Contributed Paper Session / Séance de communications libres











Chair/Présidente: Katherine Davies (University of Manitoba)

- 10:15 - 10:30 Avik Halder & Glen Takahara (Queen's University)
Bayesian Semiparametric Inference of Frailty Model Using Lévy Process Priors / Inférence bayésienne semi-paramétrique d'un modèle de la fragilité utilisant un processus de Lévy comme loi a priori  
- 10:30 - 10:45 Christian Boudreau (University of Waterloo), Hugo Varet & Sandrine Katsahian (Hôpital Henri Mondor)
Multivariate Frailty Models for Competing Risks / Modèles de fragilité multivariés pour risques concurrents  
- 10:45 - 11:00 Balakumar Swaminathan & Yun-Hee Choi (University of Western Ontario)
Ascertainment Corrected Penetrance Estimate for Bivariate Event Time Outcomes using Frailty Model / Estimation de la pénétrance corrigée par constatation à l'aide d'un modèle de fragilité  
- 11:00 - 11:15 Liqun Diao & Richard J. Cook (University of Waterloo)
Analysis of Recurrent Exacerbations in Chronic Disease via Vine Copula Models / Analyse d'exacerbations récurrentes dans les maladies chroniques via des modèles de copules avec des vignes  
- 11:15 - 11:30 Hua Shen & Richard J. Cook (University of Waterloo)
A Sequential Mover-Stayer Model for Recurrent Events with "Burn-out" / Modèle mobile-stable séquentiel pour les événements récurrents avec épuisement  

10:15 - 11:30 Session 1B-C3 (abstracts/résumés - page 80) **Rozanski Hall, room/local 105**
Predictive Models and Survey Methods**Modèles de prédiction et méthodes d'enquête**

Contributed Paper Session / Séance de communications libres

Chair/Président: Christian Léger (Université de Montréal)

- 10:15 - 10:30 Christopher van Bommel, Angela Thompson, Laurie Wadsworth & Xu Wang (St. Francis Xavier University)
Keeping Pace: Physical Activity and Dietary Intake of Nova Scotia Students / Keeping Pace : Activité physique et apport alimentaire des étudiants de Nouvelle-Écosse  
- 10:30 - 10:45 Moshe Feder & Danny Pfeffermann (University of Southampton, U.K.)
Empirical Likelihood Approach to Non-Response / Approche par la vraisemblance empirique pour la non-réponse  
- 10:45 - 11:00 Isabel Molina & Nirian Martín (Universidad Carlos III de Madrid, Spain)
Empirical Best Estimation under a Nested Error Linear Regression Model with Log Transformation / Meilleure estimation empirique pour un modèle de régression linéaire avec erreurs emboîtées et transformation logarithmique  
- 11:00 - 11:15 Fatemeh Dorri & Ali Ghodsi (University of Waterloo)
Covariate Shift Adaptation by Metric Learning / Adaptation au décalage d'une covariable par apprentissage métrique  
- 11:15 - 11:30 Yan Yuan (University of Alberta)
Prediction Error Estimation with a Changing Covariates Distribution / Estimation de l'erreur de prévision lorsque la distribution des covariables est changeante  
-











10:15 - 11:30 Session 1B-C4 (abstracts/résumés - page 83) **Rozanski Hall, room/local 102**

Times Series Models

Modèles de séries chronologiques

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Yulia Gel (University of Waterloo)

- 10:15 - 10:30 Maciej Augustyniak (Université de Montréal)
Estimation of the Markov-switching GARCH Model by a Monte Carlo EM Algorithm / Estimation pour le modèle GARCH à changement de régime markovien par un algorithme EM Monte Carlo  
- 10:30 - 10:45 Weiwei Liu (University of Western Ontario)
General Markov Tools with Application to GARCH Models / Outils de Markov généralisés avec application aux modèles GARCH  
- 10:45 - 11:00 Vyacheslav Lyubchich & Yulia Gel (University of Waterloo)
ARCH/GARCH Models and Trends: From Finance to Medicine, Weather and Further / Modèles ARCH/GARCH et tendance : de la finance à la médecine, météorologie et plus encore  
- 11:00 - 11:15 Zi Zhen Liu, Reg Kulperger & Hao Yu (University of Western Ontario)
Variable Selection and Estimation in AR Modelling via Adaptive LASSO Procedure / Sélection de variables et estimation pour modèles AR via une procédure LASSO adaptative  
- 11:15 - 11:30 Wesley Burr & Glen Takahara (Queen's University), Hwashin H. Shin (Health Canada)
Towards Developing an Air Health Index using Spectral Decomposition of GAMs / Élaboration d'un indice de la qualité de l'air au moyen de la décomposition spectrale des MAG  

11:45 - 13:15
Lunch / Dîner

Creelman Hall



13:15 - 14:15 Session 1C-A (abstract/résumé - page 86) **Rozanski Hall, room/local 104**

SSC Impact Award Address

Allocution du récipiendaire du prix pour impact de la SSC

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: James Stafford (University of Toronto)

- 13:15 - 14:15 Francis Zwiers (Pacific Climate Impacts Consortium, University of Victoria)
On Bridging Two Solitudes – Examples of a Slow Diffusive Process / Sur la réunion de deux solitudes – exemples d'un processus de diffusion lent  

13:15 - 14:45 Session 1C-I1 (abstracts/résumés - page 87) **Rozanski Hall, room/local 103**



Asymptotic Methods in Stochastics: a Session in Honour of Miklós Csörgő's 80th Birthday





Séance en l'honneur du 80e anniversaire de Miklós Csörgő

Invited Paper Session / Présentations sur invitation

Chair/Président: Don McLeish (University of Waterloo)

Organizer/Responsable: Barbara Szyszkowicz (Carleton University)







- 13:15 - 13:45 Don Dawson (Carleton University)
Path Properties of Fifty Years of Research in Probability and Statistics: a Tribute to Miklós Csörgő / Les propriétés d'une trajectoire de cinquante ans de recherche en probabilités et en statistique : un hommage à Miklós Csörgő  

- 13:45 - 14:15 Don McLeish (University of Waterloo)
Simulation of some Continuous-time Models in Finance / Simulation de modèles en temps continu en finance  
- 14:15 - 14:45 Jon N.K. Rao (Carleton University)
Large-sample Theory for Sampling from Finite Populations: Selected Developments / Théorie asymptotique pour l'échantillonnage à partir de populations finies : quelques développements  

13:15 - 14:45 Session 1C-I2 (abstracts/résumés - page 89) **Alexander Hall, room/local 100**

Inference and Applications using Estimating Functions
Inférence à l'aide de fonctions d'estimation et applications







Invited Paper Session / Présentations sur invitation
Chair/Président: Aerambamoorthy Thavaneswaran (University of Manitoba)
Organizer/Responsable: Melody Ghahramani (University of Winnipeg)

- 13:15 - 13:45 Peter X.-K. Song, Fei Wang & Lu Wang (University of Michigan)
Estimating Equations Approach to Merging Longitudinal Data / Approche par équations d'estimation pour la fusion de données longitudinales  
- 13:45 - 14:15 Grace Y. Yi (University of Waterloo), Nancy Reid (University of Toronto)
Biased Estimating Functions and Applications / Fonctions d'estimation biaisées et applications  
- 14:15 - 14:45 Michelle Qian Zhou (Harvard University), Peter X.-K. Song (University of Michigan), Mary Thompson (University of Waterloo)
Information Ratio Test for Model Misspecification in Quasi-Likelihood Inference / Test du rapport d'information pour l'erreur de spécification de modèle dont l'inférence est basé sur la quasi-vraisemblance  

13:15 - 14:45 Session 1C-I3 (abstracts/résumés - page 91) **Rozanski Hall, room/local 101**

Modern Multiple Imputation Approaches to Missing Data Problems
Approches modernes à l'imputation multiple pour les problèmes de données manquantes

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

- 13:15 - 13:45 Ben Goodrich (Columbia University)
Evaluating the Missing-At-Random Assumption after Multiply Imputing Missing Data / Évaluation de l'hypothèse de « données manquantes au hasard » après l'imputation multiple de données manquantes  
- 13:45 - 14:15 Nathaniel Schenker (National Center for Health Statistics and Centers for Disease Control and Prevention)
Multiple Uses for Multiple Imputation at the U.S. National Center for Health Statistics / Diverses utilisations de l'imputation multiple au National Center for Health Statistics des États-Unis  
- 14:15 - 14:45 Yajuan Si & Jerome P. Reiter (Duke University)
Multiple Imputation in Panel Studies with Attrition and Refreshment Samples / Imputation multiple dans des études de panel avec érosion et échantillons ajoutés  

13:15 - 14:45 Session 1C-I4 (description - page 93) **Alexander Hall, room/local 200**
Reflections on 40 Years of Teaching Statistics
Réflexions sur 40 ans d'enseignement de la statistique

Panel Discussion / Table ronde

Organizer and Chair / Responsable et présidente: Alison Gibbs (University of Toronto)

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

- 13:15 - 14:45 John Brewster (University of Manitoba)
 13:15 - 14:45 Paul Corey (University of Toronto)
 13:15 - 14:45 Chris Field (Dalhousie University)
 13:15 - 14:45 Jerry Lawless (University of Waterloo)
 13:15 - 14:45 Larry Weldon (Simon Fraser University)

13:15 - 14:45 Session 1C-C1 (abstracts/résumés - page 94) **MacNaughton Building, room/local 105**
Analysis of Longitudinal Data
Analyse de données longitudinales

Contributed Paper Session / Séance de communications libres









Chair/Président: Paramjit Gill (University of British Columbia, Okanagan)

- 13:15 - 13:30 Hyun (June) Lim, Prosanta Mondal & Stuart Skinner (University of Saskatchewan)
 Joint Modelling of Repeated CD4+ Count Measurements and Survival Time in HIV Study
 / Modélisation conjointe de mesures répétées de dénombrement de CD4+ et de temps de
 survie dans une étude sur le VIH E E
- 13:30 - 13:45 Tasneem Zaihra (Concordia University), Sara Ahmed & Pierre Ernst (McGill University)
 Evaluating Predictive Validity of Patient Reported Asthma Control, Belief about Medications
 and Self-Efficacy / Évaluation de la validité prédictive du contrôle de l'asthme, des croyances
 à l'égard des médicaments et de l'autoefficacité déclarés par le patient E E
- 13:45 - 14:00 Abderazzak Mouiha (Centre de Recherche de l'Institut Universitaire de Santé Mentale de
 Québec), Simon Duchesne (Université Laval)
 The Dynamic of Alzheimer's Disease: a Longitudinal Study of Patients Progressing From MCI
 to AD / La dynamique de la maladie d'Alzheimer : une étude longitudinale des patients ayant
 progressé de MCI vers AD F E
- 14:00 - 14:15 Narges Nazeri Rad & Jerry Lawless (University of Waterloo)
 Multi-state Models for Viral Load Patterns / Modèles multi-états pour la charge virale E E
- 14:15 - 14:30 Depeng Jiang, Andrea E. Bombak & Qier Tan (University of Manitoba)
 Latent Class Growth Model for National Longitudinal Data / Modèle de croissance à classes
 latentes pour données longitudinales nationales E E
- 14:30 - 14:45 Chenglin Ye, Gary Foster, Lisa Dolovich & Lehana Thabane (McMaster University)
 Comparison of Methods to Model Diastolic and Systolic Blood Pressure Measurements over
 Time / Comparaison de méthodes de modélisation des mesures de la tension artérielle dias-
 tolique et systolique au fil du temps E E

13:15 - 14:15 Session 1C-C2 (abstracts/résumés - page 97) **Rozanski Hall, room/local 105**
Estimating Population Size and Growth
Estimation de la taille et de la croissance des populations

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Llwellyn Armstrong (Ducks Unlimited Canada)

- 13:15 - 13:30 Noel Cadigan (Marine Institute of Memorial University of Newfoundland)
A Nonparametric Approach to Fitting a Stock-Recruitment Relationship, and Implications for Fisheries Management Reference Points / Approche non-paramétrique d'ajustement d'une relation stock-recrutement et implications pour les points de référence de gestion des pêcheries  
- 13:30 - 13:45 Marylène Troupé & Jean Vaillant (Université des Antilles et de la Guyane), Léonide Celini (Université de Paris-Est-Créteil, France)
Sampling of Thinned Point Processes: applications to Aphis Gossypii mortality rates / Échantillonnage de processus ponctuels amincis : application aux taux de mortalité d'Aphis Gossypii  
- 13:45 - 14:00 Lawrence McCandless, Michelle Patterson, Lauren Currie, Akm Moniruzzaman & Julian Somers (Simon Fraser University)
Counting the Unsheltered Homeless in Edmonton Using Plant-Capture Studies from Toronto / Dénombrement des sans-abris sans refuge à Edmonton à l'aide d'une étude « placer-capturer » de Toronto  
- 14:00 - 14:15 B. Johannes Edinger (First Nations Statistical Institute)
A Stochastic Population Projection Model for Canada's First Nations / Un modèle stochastique de projection de la population pour les Premières Nations du Canada  

13:15 - 14:45 Session 1C-C3 (abstracts/résumés - page 99) **MacNaughton Building, room/local 113**

Generalized Linear Models, Overdispersion and Zero-Inflation

Modèles linéaires généralisés, inflation de la variance et surabondance de zéros

Contributed Paper Session / Séance de communications libres











Chair/Président: Thierry Duchesne (Université Laval)

- 13:15 - 13:30 Michael Evans & Zeynep Baskurt (University of Toronto)
Bayesian Goodness of Fit and g -Priors / Adéquation bayésienne et lois a priori de type g  
- 13:30 - 13:45 Niroshan Withanage & Alexander R. de Leon (University of Calgary)
Joint Estimation of Specificities and Sensitivities in Reader-based Diagnostic Studies / Estimation conjointe des spécificités et des sensibilités des études fondées sur la lecture du diagnostic  
- 13:45 - 14:00 Catherine Crea & Ayesha Ali (University of Guelph)
On the Robustness of Dirichlet-Multinomial Regression in the Context of Modeling Pollination Networks / De la robustesse de la régression Dirichlet-multinomiale dans le contexte de la modélisation des réseaux de pollinisation  
- 14:00 - 14:15 Xuemao Zhang (Lakehead University), Sudhir Paul (University of Windsor), Deli Li (Lakehead University)
Modeling Longitudinal Binomial Responses by Generalized Estimating Equations / Modélisation de réponses binomiales longitudinales par des équations d'estimation généralisées  
- 14:15 - 14:30 S. M. Khurshid Alam & Sudhir Paul (University of Windsor), Xuemao Zhang (Lakehead University)
Negative Binomial Behrens-Fisher Problem / Problème de Behrens-Fisher pour la binominale négative  
- 14:30 - 14:45 Connie Stewart (University of New Brunswick), Chris Field (Dalhousie University)
Statistical Methods in Diet Estimation Using Quantitative Fatty Acid Signature Analysis / Méthodes statistiques pour estimer le régime alimentaire via l'analyse quantitative de la signature des acides gras  
-

13:15 - 14:30 Session 1C-C4 (abstracts/résumés - page 102) **Thornbrough Building, room/local 1307**
Methods for High-Dimensional Data/Data Reduction**Analyse de données en plusieurs dimensions**

Contributed Paper Session / Séance de communications libres

Chair/Président: Renjun Ma (University of New Brunswick)



- 13:15 - 13:30 Xiaogang (Steven) Wang (York University)
Dimensionality Reduction for Categorical Data / Réduction de dimensionnalité pour des données catégorielles  
- 13:30 - 13:45 Babak Alipanahi Ramandi (University of Waterloo), Nathan Krislock (INRIA Grenoble Rhône-Alpes, France), Ali Ghodsi (University of Waterloo)
Large-Scale Manifold Learning by Semidefinite Facial Reduction / Apprentissage de variétés à grande échelle par réduction de faces semidéfinie  
- 13:45 - 14:00 Sun Makosso-Kallyth (Centre de Recherche du Centre Hospitalier Universitaire de Québec)
Symbolic Data Analysis of m Sets of Intervals / Analyse de données symboliques pour m ensembles d'intervalles  
- 14:00 - 14:15 Ashley Bonner (McMaster University)
Sparse Principal Component Analysis for High-Dimensional Data: A Comparative Study with Applications in Genetics. / Analyse parcimonieuse en composantes principales pour données de grande dimension : une étude comparative appliquée à la génétique  
- 14:15 - 14:30 Sathish Chandra Pichika & Joseph Beyene (McMaster University)
Sparse Canonical Correlation Analysis (SCCA): A Comparative Study / Analyse de corrélation canonique étendue : étude comparative  

13:15 - 14:30 Session 1C-C5 (abstracts/résumés - page 105) **Rozanski Hall, room/local 102**
Modern Regression Methods**Méthodes modernes de régression**

Contributed Paper Session / Séance de communications libres

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

- 13:15 - 13:30 Song Cai (University of British Columbia)
Two Fast Algorithms for L^1 -type Estimation in Regression / Deux algorithmes rapides pour l'estimation de type L^1 en régression  
- 13:30 - 13:45 Fuqi Chen & Sévérien Nkurunziza (University of Windsor)
Shrinkage Estimation in Some Model Selection by LASSO in Multivariate Regression with Structural Changes / Estimation par rétrécissement dans certains modèles de sélection par la méthode LASSO en régression multivariée avec changements structurels  
- 13:45 - 14:00 Nicholas Mitsakakis (Toronto Health Economics and Technology Assessment, University of Toronto), George Tomlinson (Department of Health Policy, Management and Evaluation, University of Toronto)
Bayesian Regression Models for Estimation of Disease-Specific Net Costs using Aggregate Data / Modèles de régression bayésiens pour l'estimation des coûts nets de certaines maladies à l'aide de données agrégées  
- 14:00 - 14:15 Vanda Lourenço (Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal), Ana Maria Pires (Instituto Superior Técnico da Universidade Técnica de Lisboa, Portugal)
Robust Outlier Detection in Genetic Association Studies of Quantitative Traits in the Context of M-regression / Détection robuste d'observations aberrantes dans les études d'association génétique avec modèles de régression de type M  

- 14:15 - 14:30 Salimah Ismail (University of Victoria), Wenqi Sun (Simon Fraser University), Farouk Nathoo & Arif Babul (University of Victoria), Alexander Moiseev (Down Syndrome Research Foundation of British Columbia), Naznin Virji-Babul (University of British Columbia), Faisal Beg (Simon Fraser University)
A Skew-t Space-varying Regression Model for the Spectral Analysis of Magnetoencephalography Readings / L'analyse spectrale des mesures de magnétoencéphalographie à l'aide d'un modèle de régression à espace variable et un modèle t asymétrique  

13:15 - 15:15 Session 1C-D1 (description - page 108) **Science Complex Atrium**

Case Study I: Mark-Recapture on Atlantic Cod

Études de cas I : Marquage-recapture de la morue de l'Atlantique

Case Study / Étude de case

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

Organizers/Responsables: John Bratley (Fisheries and Oceans Canada / Pêches et Océans Canada), Karla Fox (Statistics Canada / Statistique Canada) and/et Lajmi Lakhal-Chaieb (Université Laval)

- 13:15 - 15:15 Audrey Béliveau & Zheng Sun (Simon Fraser University)
13:15 - 15:15 Maggie Chen & Ying Qi (University of Toronto)
13:15 - 15:15 Paul Van Dam-Bates, Fan Wu & Dean Koch (University of Victoria)
13:15 - 15:15 Tingting Zhao, Andy Leung & Yiyang Pan (University of British Columbia)

13:15 - 15:15 Session 1C-D2 (description - page 109) **Science Complex Atrium**

Case Study II: Risk Factors for Osteoporotic Fracture

Étude de cas II : Facteurs de risque de fracture ostéoporotique

Case Study / Étude de case

Chair/Président: Lajmi Lakhal Chaieb (Université Laval)

Organizers/Responsables: Lajmi Lakhal Chaieb (Université Laval) and/et Lisa Lix (University of Saskatchewan)

- 13:15 - 15:15 Janie Coulombe, Marie-Christine Doucas & Kenza Sallier (Université de Montréal)
13:15 - 15:15 Shu Guo, Jialin Zou & Lingling Fan (University of Toronto)
13:15 - 15:15 Wei Lin, Ximing Xu, Thi Lam Uyen Hoang & Vu Thien Huong Le (University of Toronto)
13:15 - 15:15 Vachan Misir, Azadeh Tanha & Yang Chen (University of Toronto)
13:15 - 15:15 Ying Qi & Jingxiang Xu (University of Toronto)
13:15 - 15:15 Nahid Sultana, Mohammad Shahidul Islam, Tagreed Almalki & Huda Alsulami (McMaster University)
13:15 - 15:15 Yanhua Tian and Nishan Mudalige (York University)
13:15 - 15:15 Daoxing Wang, Xin Lu, Xin Chen & Xueshan Zhang (University of Waterloo)

13:15 - 15:15 Session 1C-P (abstracts/résumés - page 110) **Science Complex Atrium**

Poster Session I

Séance d'affichage I

Contributed Poster Session / Séance d'affichage libre

- 13:15 - 15:15 Prathayana Balendran & Hanna Jankowski (York University)
Estimation of Songbird Flight Paths Using a Bayesian Modelling Framework / Estimation des trajectoires de vol des oiseaux chanteurs au moyen d'un cadre de modélisation bayésien
13:15 - 15:15 Jessica Connolly (University of Guelph)
Assessing Accuracy and Error Trends in an Underwater Acoustic Telemetry System / Évaluation de l'exactitude et des tendances d'erreur en télémétrie acoustique sous-marine
13:15 - 15:15 L. Spectrum Han & John Braun (University of Western Ontario)
Diurnal Wind Speed Modelling / Modélisation de la vitesse diurne du vent

- 13:15 - 15:15 Shofiqul Islam, Joseph Beyene & Jemila Hamid (McMaster University)
Multivariate Growth Curve Models for the Analysis of Longitudinal Data / Modèles de courbe de croissance multivariés pour l'analyse de données longitudinales
- 13:15 - 15:15 Grace P.S. Kwong, Zvonimir Poljak, Rob Deardon & Cate E. Dewey (University of Guelph)
Bayesian Analysis of PRRS Virus Spread in Ontario Swine Herds / Analyse bayésienne de la propagation du virus SDRP dans les troupeaux porcins de l'Ontario
- 13:15 - 15:15 Wenkai Ma, John Braun & Gregory Kopp (University of Western Ontario), Mike Wotton (University of Toronto)
Spotfire Distance Distribution Modelling: Designing a Wind Tunnel Experiment / Modélisation de la distribution des distances de propagation des feux disséminés : conception d'un essai en soufflerie
- 13:15 - 15:15 Nagham Mohammad & Ian A. McLeod (University of Western Ontario)
Left-Censored Gamma / Gamma censurée à gauche
- 13:15 - 15:15 Melissa Morgan & Ashok Krishnamurthy (Mount Royal University)
Bayesian Tracking of Epidemic Infectious Diseases using Data Assimilation Methods / Suivi bayésien des épidémies de maladies infectieuses par des méthodes d'assimilation des données
- 13:15 - 15:15 Lori L. Murray & David R. Bellhouse (University of Western Ontario)
Early Smoothing Techniques: Halley's 1701 Map of Magnetic Declination / Techniques de lissage précoces : carte de la déclinaison magnétique de 1701 de Halley
- 13:15 - 15:15 Paul Nguyen (University of Toronto), Patrick Brown (Cancer Care Ontario and University of Toronto / Cancer Care Ontario et University of Toronto), James Stafford (University of Toronto)
Mapping Disease Risk with Changing Census Boundaries / Cartographie du risque de maladie avec des limites de recensement changeantes

14:45 - 15:15

Coffee Break
Pause-café

Science Complex Atrium

15:15 - 16:45

NSERC Discovery Grant Workshop
Atelier des subventions à la découverte du CRSNG

MacKinnon Building, room 120

15:15 - 16:45 **Session 1D-I1** (abstracts/résumés - page 115)

Alexander Hall, room/local 100

Analysis of Interval-censored Data
Analyse de données censurées par intervalles



Invited Paper Session / Présentations sur invitation



Organizer and Chair / Responsable et président: John Braun (University of Western Ontario)

15:15 - 15:45 James Stafford (University of Toronto), Patrick Brown (Cancer Care Ontario and University of Toronto / Cancer Care Ontario et University of Toronto), Chun-Po Steve Fan & Paul Nguyen (University of Toronto)

Local-EM and Penalized Likelihood / Algorithme EM local et vraisemblance pénalisée  

15:45 - 16:15 Joan Hu (Simon Fraser University), John Braun (University of Western Ontario)

Two Stage Modeling in Event History Analysis, with Application to Forest Fire Control / Modélisation en deux phases de l'historique des événements avec application au contrôle des feux de forêt  

- 16:15 - 16:45 Patrick Brown (Cancer Care Ontario and University of Toronto / Cancer Care Ontario et University of Toronto), Paul Nguyen & James Stafford (University of Toronto), Lennon Li (Public Health Ontario)
Disease Mapping with Spatially Aggregated Data / Cartographie des maladies à l'aide de données spatialement agrégées  

15:15 - 16:45 Session 1D-I2 (abstracts/résumés - page 117) **Rozanski Hall, room/local 104**




Celebrating Our 40th Anniversary

Célébrons nos 40 ans

Invited Paper Session / Présentations sur invitation

Chair/Président: John Brewster (University of Manitoba)

Organizers/Responsables: Christian Genest (McGill University) and/et Alexandre Leblanc (University of Manitoba)

- 15:15 - 16:00 David R. Bellhouse (University of Western Ontario), Christian Genest (McGill University)
Celebrating our 40th Anniversary / Célébrons nos 40 ans  
- 16:00 - 16:45 Christian Genest (McGill University), David R. Bellhouse (University of Western Ontario)
Celebrating our 40th Anniversary / Célébrons nos 40 ans  

15:15 - 16:45 Session 1D-I3 (abstracts/résumés - page 118) **Rozanski Hall, room/local 101**

Linear and Nonlinear Time Series







Modèles linéaires et non linéaires de séries chronologiques

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Reg Kulperger (University of Western Ontario)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

- 15:15 - 15:45 David Brillinger (University of California, Berkeley)
The Nicholson Blowfly Experiments: Some History and EDA, II / Les expériences avec les mouches de Nicholson : un historique et EDA, II  
- 15:45 - 16:15 Weibin Jiang (BMO Financial Group)
Asymptotic Theory of General Multivariate GARCH Models / Théorie asymptotique pour des modèles GARCH multivariés généraux  
- 16:15 - 16:45 Rafal Kulik (University of Ottawa), Philippe Soulier (Université Paris X, France)
Stochastic Volatility Models with Heavy Tails and Long Memory: Structure and Estimation / Modèles de volatilité stochastique à mémoire longue et à ailes lourdes : structure et estimation  



15:15 - 16:45 Session 1D-I4 (abstracts/résumés - page 120) **Rozanski Hall, room/local 103**





Recent Advances in Small Area Estimation

Progrès récents en estimation pour les petits domaines

Invited Paper Session / Présentations sur invitation

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

- 15:15 - 15:45 Gauri Datta (University of Georgia), Peter Hall (University of Melbourne, Australia), Lily Wang (University of Georgia)
Estimation of Small Area Means under Semi-Parametric Measurement Error Model / Estimation des moyennes de petits domaines en fonction d'un modèle d'erreur de mesure semi-paramétrique  

- 15:45 - 16:15 Jiming Jiang (University of California, Davis), Thuan Nguyen (Oregon Health and Science University), J. Sunil Rao (University of Miami)
Observed Best Prediction via Nested-Error Regression / Meilleure prévision observée par la régression à erreur emboîtée  
- 16:15 - 16:45 Partha Lahiri (University of Maryland), Santanu Pramanik (NORC, University of Chicago)
Estimation of Average Design-Based Mean Squared Error of Synthetic Small Area Estimators / Estimation de la moyenne des écarts quadratiques moyens selon le plan d'échantillonnage d'estimateurs synthétiques pour petits domaines  







15:15 - 16:45 Session 1D-I5 (abstracts/résumés - page 122) **Alexander Hall, room/local 200**

Statistical Methods for Clinical Trials and Causal Inference

Méthodes statistiques pour les essais cliniques et l'inférence causale

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Lawrence McCandless (Simon Fraser University)

- 15:15 - 15:45 Juli Atherton (Université du Québec à Montréal), David Wolfson (McGill University), Edens Thaddeus
An Overview for Bayesian Optimal Design for Clinical Trials with an Unknown Delay to Treatment Effect / Survol de la planification optimale bayésienne d'essais cliniques avec un délai inconnu à l'effet du traitement  
- 15:45 - 16:15 Geneviève Lefebvre & Juli Atherton (Université du Québec à Montréal)
On the Role of the Prior Distribution in the "Bayesian Adjustment for Confounding" Algorithm / Sur le rôle de la loi a priori de l'algorithme "Bayesian Adjustment for Confounding"  
- 16:15 - 16:45 Saman Muthukumarana (University of Manitoba)
Assessing Noninferiority in Multi Arm Clinical Trials / Analyse de non-infériorité dans des essais cliniques comportant plusieurs groupes  











15:15 - 16:30 Session 1D-C1 (abstracts/résumés - page 124) **Rozanski Hall, room/local 105**

Density Approximation and Estimation

Approximation et estimation de densité

Contributed Paper Session / Séance de communications libres









Chair/Président: Mikelis Bickis (University of Saskatchewan)

- 15:15 - 15:30 Anas Ramadan & Alan Ker (University of Guelph)
Empirical Bayes Nonparametric Density Estimation of Crop Yield Densities: Rating of Crop Insurance / Estimation non paramétrique bayésienne empirique de la densité des rendements des cultures  
- 15:30 - 15:45 Shyam Dadimuni (Wilfrid Laurier University)
Large Forest Fire Risk Density Estimation / Estimation de la densité du risque de grands feux de forêt  
- 15:45 - 16:00 Serge Provost (University of Western Ontario)
Polynomial Representations of Differentiated Logdensity Approximants / Sur l'approximation de la dérivée du logarithme d'une densité de probabilité par polynômes  
- 16:00 - 16:15 Aliakbar Mohsenipour & Serge Provost (University of Western Ontario)
The Distribution of Quadratic forms in Elliptically Contoured Vectors / La distribution des formes quadratiques pour des vecteurs aléatoires elliptiques  
- 16:15 - 16:30 Susan Zhe Sheng (University of Western Ontario)
Polynomially Adjusted Saddlepoint Approximations / Approximations des points de selle avec ajustement polynomial  

15:15 - 16:15 Session 1D-C2 (abstracts/résumés - page 127) **MacNaughton Building, room/local 113****Impact and Comparison of Cancer Screening Strategies****Impact et comparaison des examens préventifs pour le cancer**

Contributed Paper Session / Séance de communications libres









Chair/Présidente: Wendy Lou (University of Toronto)





- 15:15 - 15:30 Cecilia Cotton & Krithika Suresh (University of Waterloo), David Hodgson (University of Toronto), Crystal Hann & Linda Mayhew (Princess Margaret Hospital)
Breast Cancer Screening in Female Survivors of Childhood Hodgkin Lymphoma: A Simulation Study / Dépistage du cancer du sein chez les survivantes du lymphome hodgkinien de l'enfant : étude de simulation  
- 15:30 - 15:45 James A. Hanley, Zihui (Amy) Liu, Nandini Dendukuri & Erin Strumpf (McGill University)
Measuring Impact of Cancer Screening without Data-Holders Having to Share Trial-Specific Data / Mesure des répercussions du dépistage du cancer sans que les détenteurs de données doivent partager des informations spécifiques au sujet de leur essai clinique  
- 15:45 - 16:00 Amidu Olalekan Raifu & Stephen D. Walter (McMaster University), Eduardo Franco & Ramana-Kumar Agnihotram (McGill University)
Effects of Covariates on the Performance of Cervical Cancer Screening Tests: Logistic Regression and Latent Class Models / Effets des covariables sur la performance des tests de dépistage du cancer du col de l'utérus : régression logistique et modèles de catégorie latente  
- 16:00 - 16:15 Zihui (Amy) Liu, Erin Strumpf, Nandini Dendukuri, Maurice McGregor & James A. Hanley (McGill University)
Measuring the Mortality Reductions Produced by Breast Cancer Screening / Mesurer la réduction de mortalité due au dépistage du cancer du sein  

15:15 - 16:45 Session 1D-C3 (abstracts/résumés - page 129) **Rozanski Hall, room/local 102****Monte Carlo Methods and Inference for Stochastic Processes****Méthodes de Monte Carlo et inférence pour les processus stochastiques**

Contributed Paper Session / Séance de communications libres

Chair/Président: Rob Deardon (University of Guelph)

- 15:15 - 15:30 Christiane Lemieux (University of Waterloo)
A Study of Quasi-Monte Carlo Methods Via Dependence Concepts / Concepts de dépendance et méthodes quasi-Monte Carlo  
- 15:30 - 15:45 Jourdan Gold, Rob Deardon & Zeny Feng (University of Guelph)
An Investigation of Multi-stage MCMC Approaches for Sampling from Correlated, Discrete Target Distributions / Une étude de différentes approches dans la méthode de Monte Carlo par chaîne de Markov multi-étapes pour la simulation de distributions discrètes corrélées  
- 15:45 - 16:00 Lotfi Khribi & Marc Fredette (HEC Montréal), Brenda MacGibbon (Université du Québec à Montréal)
The Poisson Maximum Entropy Model for Homogeneous Poisson Processes / Application du principe d'entropie maximale dans la prédiction des événements récurrents dans le cas des processus de Poisson  
- 16:00 - 16:15 Larissa Valmy & Jean Vaillant (Université des Antilles et de la Guyane)
Statistical Inference for Point Processes Associated with a Dirichlet Process / Inférence statistique pour des processus ponctuels associés à un processus de Dirichlet  

- 16:15 - 16:30 Liwen Zou (North Carolina State University)
Fitting Nonstationary General-time-reversible Models to Obtain Edge-lengths and Frequencies for the Barry-Hartigan Model / Ajustement de modèles temporels réversibles généraux non stationnaires pour l'obtention des fréquences et longueurs des frontières dans le modèle Barry-Hartigan  
- 16:30 - 16:45 Annaliza McGillivray & Abbas Khalili (McGill University)
A Penalized Quasi-Likelihood Approach for Estimating the Number of States in a Hidden Markov Model / Une approche de la quasi-vraisemblance pénalisée pour estimer le nombre d'états dans une chaîne de Markov cachée  









15:15 - 16:15 Session 1D-C4 (abstracts/résumés - page 132) **Thornbrough Building, room/local 1307**

Optimal Design

Planification optimale d'expériences

Contributed Paper Session / Séance de communications libres

Chair/Président: François Pageau (General Dynamics Produits de défense et Systèmes tactiques - Canada Inc.)

- 15:15 - 15:30 Yin Li & Keumhee Chough Carriere (University of Alberta)
Optimal Designs in the N of 1 Trials / Plans optimaux pour essais N de 1  
- 15:30 - 15:45 Saumen Mandal & Chao Zhu (University of Manitoba)
Some Iterative Procedures for Constructing Optimal Designs / Des processus itératifs pour la construction de plans optimaux  
- 15:45 - 16:00 Yanqing Yi & Yuan Yuan (Memorial University of Newfoundland)
An Optimal Randomization Rule for Response Adaptive Designs / Règle de randomisation optimale pour les plans adaptatifs à la réponse  
- 16:00 - 16:15 Xuan Li & Xikui Wang (University of Manitoba)
Variance-penalized Response-adaptive Randomization with Mismeasurement / Randomisation adaptée à la réponse avec pénalisation de la variance et erreur de mesure  









15:15 - 16:15 Session 1D-C5 (abstracts/résumés - page 134) **MacNaughton Building, room/local 105**

Survival Analysis and Accelerated Life Testing

Analyse de survie et tests accélérés de durée de vie

Contributed Paper Session / Séance de communications libres



Chair/Président: Christian Boudreau (University of Waterloo)

- 15:15 - 15:30 Ammar Sarhan & Joseph Apaloo (St. Francis Xavier University)
Exponentiated Modified Weibull Extension Distribution / La distribution de Weibull modifiée généralisée et exponentiée  
- 15:30 - 15:45 Dr. Peter Adamic & Sylvain Caron (Laurentian University)
A Model for Time-Dependent Competing Risks / Un modèle sujet aux risques concurrents lié au temps  
- 15:45 - 16:00 Scott Hunt & Xiaojian Xu (Brock University)
Robust Designs for an Accelerated Life Testing Model with Simple Step-Stress Plans / Plans d'expériences robustes pour un modèle d'essai de vie accéléré avec un plan simple en contraintes échelonnées  
- 16:00 - 16:15 Longlong Huang (University of Calgary), Karen Kopciuk (Alberta Health Services), Xuewen Lu (University of Calgary)
Jackknife Empirical Likelihood for Smoothed Weighted Rank Regression with Censored Data / Vraisemblance empirique avec jackknife pour régression lisse pondérée basée sur les rangs avec données censurées  

Tuesday June 5**mardi 5 juin****8:45 - 9:45 Session 2A-A (abstract/résumé - page 136) Rozanski Hall, room/local 104****Gold Medal Address****Allocution du récipiendaire de la médaille d'or**

Special Session / Conférence spéciale



Organizer and Chair / Responsable et président: Roman Viveros-Aguilera (McMaster University)

8:45 - 9:45 Christian Genest (McGill University)
 How to Model Dependence if You Must / Comment modéliser la dépendance s'il le faut  

9:45 - 10:15 Science Complex Atrium**Coffee Break****Pause-café****10:15 - 11:45 Session 2B-V (abstract/résumé - page 137) Rozanski Hall, room/local 104****An Overview of Models and Methods for Spatio-temporal Data Analysis****Un survol des modèles et méthodes d'analyse de données spatio-temporelle**

Introductory Overview Lecture / Conférence de vue d'ensemble

Organizer and Chair / Responsable et président: Alexandre Leblanc (University of Manitoba)

10:15 - 11:45 Jim Zidek (University of British Columbia)
 An Overview of Models and Methods for Spatio-temporal Data Analysis / Un survol des modèles et méthodes d'analyse pour données spatio-temporelles  



10:15 - 11:45 Session 2B-I1 (abstracts/résumés - page 138) Rozanski Hall, room/local 101**Applications of Statistics in Medicine****Applications de la statistique à la médecine**



Invited Paper Session / Présentations sur invitation



Organizer and Chair / Responsable et présidente: Juli Atherton (Université du Québec à Montréal)

Session sponsored by the Committee on New Investigators

Séance parrainée par le Comité des nouveaux chercheurs

10:15 - 10:45 Jiguo Cao (Simon Fraser University), Wei Luo, Marcus Gallagher & Janet Wiles (University of Queensland, Australia)
 Estimating the Intensity of Ward Admission and its Effect on Emergency Department Access Block / Estimation de l'intensité des admissions en unité de soins et son effet sur le blocage d'accès aux services d'urgence  

10:45 - 11:15 Abbas Khalili, Yeting Du, Johanna Nešlehová & Russell Steele (McGill University)
 The Use of Mixture of Linear Mixed Effect Models in identifying Clinical Characteristics of Systemic Sclerosis Disease / L'utilisation d'un mélange de modèles linéaires à effets mixtes pour l'identification des caractéristiques cliniques de la sclérose systémique  

11:15 - 11:45 Mahmoud Torabi (University of Manitoba)
 Likelihood Inference in Spatio-Temporal Modeling / Inférence à l'aide de la vraisemblance en modélisation spatio-temporelle  

10:15 - 11:35 Session 2B-I2 (abstracts/résumés - page 140) **Alexander Hall, room/local 200**

On Assessing Liking and Intensity in Sensory Science

Évaluation du goût et de l'intensité en sciences sensorielles









Invited Paper Session / Présentations sur invitation

Chair/Président: Paul McNicholas (University of Guelph)

Organizers/Responsables: Paul McNicholas (University of Guelph) and/et 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

- 10:15 - 10:35 Chris Findlay (Compusense Inc.)
You Know What You Like, But What About Everyone Else? A Case Study on Incomplete Block Segmentation of White-Bread Consumers. / Vous savez ce que vous aimez, mais qu'en est-il des autres? Une étude de cas sur la segmentation en blocs incomplets de consommateurs de pain blanc.  
- 10:35 - 10:55 Ryan Browne & Paul McNicholas (University of Guelph)
Design and Analysis of Incomplete Block Designs / Conception et analyse de plans en blocs incomplets  
- 10:55 - 11:15 John Castura (Compusense Inc.)
Sweet and Powerful: A Two-attribute Time Intensity Study of Fruit Liqueur / Doux et puissant : une étude de deux attributs de l'intensité temporelle de la liqueur de fruits  
- 11:15 - 11:35 Moyi Li, Ryan Browne & Paul McNicholas (University of Guelph), John Castura (Compusense Inc.)
Analysis of Time Intensity Data: a Fruit Liqueur Study / Analyse de données temporelles d'intensité : étude d'une liqueur de fruits  

10:15 - 11:45 Session 2B-I3 (abstracts/résumés - page 142) **Alexander Hall, room/local 100**

Outlier-Robust Estimation in Survey Sampling

Estimation robuste et méthodes d'échantillonnage







Invited Paper Session / Présentations sur invitation

Chair/Présidente: Julie Trépanier (Statistics Canada / Statistique Canada)

Organizer/Responsable: Jean-François Beaumont (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 10:15 - 10:45 Malay Gosh (University of Florida)
Outlier Adjusted Robust Bayes Estimators in Survey Sampling / Estimateurs bayésiens robustes ajustés pour les valeurs aberrantes pour les méthodes d'enquête  
- 10:45 - 11:15 David Haziza (Université de Montréal), Jean-François Beaumont (Statistics Canada / Statistique Canada), Anne Ruiz-Gazen (Université de Toulouse)
A Unified Approach to Robust Estimation in Finite Population Sampling / Une approche unifiée d'estimation robuste en échantillonnage  
- 11:15 - 11:45 Nikolas Tzavidis (University of Southampton, U.K.), Sabine Krieg (Netherlands Central Bureau of Statistics), Chiara Bocci (IRPET and University of Florence, Italy) / IRPET et University of Florence, Italy), Marc Smeets & Virginie Raymond-Blaess (Netherlands Central Bureau of Statistics)
Outlier Robust Small Domain Estimation for Business Surveys in the Netherlands / Estimation robuste aux valeurs aberrantes pour petits domaines dans les enquêtes auprès des entreprises aux Pays-Bas  
-

10:15 - 11:45 Session 2B-I4 (abstracts/résumés - page 144) **Rozanski Hall, room/local 103**

Trend Analysis of Time Series Data and Applications

Analyse de tendances en séries chronologiques et applications







Invited Paper Session / Présentations sur invitation

Chair/Présidente: Debbie Dupuis (HEC Montréal)

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

- 10:15 - 10:45 Xin Gao (York University), Grace Y. Yi (University of Waterloo)
 Simultaneous Model Selection and Estimation for Mean and Association Structures with Clustered Binary Data / Sélection de modèle et estimation simultanés pour la moyenne et les structures d'associations avec données binaires regroupées  
- 10:45 - 11:15 Yulia Gel (University of Waterloo), Abdel El Shaarawi (National Water Research Institute), Vyacheslav Lyubchich (University of Waterloo)
 Nonparametric Tests to Detect Parametric Patterns in Time Series with ARCH Effects / Tests non-paramétriques pour détecter des tendances paramétriques dans les séries chronologiques avec effets ARCH  
- 11:15 - 11:45 Ying Zhang & Paul Cabilio (Acadia University)
 A Generalized Jonckheere Test for Testing Trend in Repeated Measures in Randomized Blocks and Single Group Short Time Series / Un test généralisé de Jonckheere pour tester la tendance dans des mesures répétées en blocs aléatoires ou pour un seul groupe de séries chronologiques courtes  









10:15 - 11:15 Session 2B-C1 (abstracts/résumés - page 146) **MacNaughton Building, room/local 105**

Biostatistics: Inference Methods

Biostatistique : méthodes inférentielles

Contributed Paper Session / Séance de communications libres











Chair/Présidente: Lisa Lix (University of Saskatchewan)

- 10:15 - 10:30 Lajmi Lakhal-Chaieb & Thierry Duchesne (Université Laval)
 Association Measures in the Presence of a Bivariate Cure Fraction / Mesure d'association en présence d'une fraction de guérison bivariable  
- 10:30 - 10:45 Jianan Peng (Acadia University)
 Simultaneous Inference for Low Dose Risk Estimation with Nonquantal Data in Benchmark Analysis / Inférence simultanée pour l'estimation du risque à faible dose avec données continues pour l'analyse d'un banc d'essais  
- 10:45 - 11:00 Wanhua Su (Grant MacEwan University), Mu Zhu (University of Waterloo), Hugh Chipman (Acadia University)
 Precision-Recall Curve or ROC Curve? / Courbe précision-rappel ou courbe ROC ?  
- 11:00 - 11:15 Lihui Zhao (Northwestern University), Lu Tian (Stanford University), Tianxi Cai, Brian Claggett & Lee-Jen Wei (Harvard University)
 Effectively Selecting a Target Population with Desired Treatment Benefit Using the Data from a Randomized Comparative Study / Choix efficace d'une population cible sensible à l'effet bénéfique voulu d'un traitement au moyen des données provenant d'une étude comparative randomisée  
-

10:15 - 11:30 Session 2B-C2 (abstracts/résumés - page 148) **MacNaughton Building, room/local 113**
Hazard Models and Censored Data**Modèles de risque et données censurées**

Contributed Paper Session / Séance de communications libres

Chair/Président: Pierre-Jérôme Bergeron (University of Ottawa)

- 10:15 - 10:30 Xin Xin, Julie Horrocks & Gerarda Darlington (University of Guelph)
Ties Between Event Times and Jump Times in the Cox Model / Égalités entre les temps d'événements et les temps de sauts dans le modèle de Cox  
- 10:30 - 10:45 Tharshanna Nadarajah & Asokan Mulayath Variyath (Memorial University of Newfoundland)
Penalized Empirical Likelihood Based Variable Selection for Cox's Proportional Hazards Model / Sélection de variables pour le modèle de Cox à risque proportionnel basée sur la vraisemblance empirique pénalisée  
- 10:45 - 11:00 Katrina Tomanelli, Abdulkadir Hussein & Sévérien Nkurunziza (University of Windsor)
Improved Estimation of Aalen's Additive Hazards Model / Estimation raffinée dans le modèle de risque additif d'Aalen  
- 11:00 - 11:15 Pordeli Pooneh, Xuewen Lu & Murray Burke (University of Calgary)
Estimation in Partially Linear Single-index Additive Hazards Regression with Current Status data / Estimation dans la régression des risques additive à index unique partiellement linéaire avec des données d'état actuel  
- 11:15 - 11:30 Xuewen Lu (University of Calgary), Heng Lian (Nanyang Technological University, Singapore), Hua Liang (University of Rochester)
Variable Selection for Censored Partially Linear Single-Index Models / Sélection de variables pour modèles à un index partiellement linéaires et censurés  

10:15 - 11:45 Session 2B-C3 (abstracts/résumés - page 151) **Rozanski Hall, room/local 105**
Hypothesis Testing and R Statistical Software**Tests d'hypothèses et logiciel statistique R**

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Erica Moodie (McGill University)

- 10:15 - 10:30 Marta Padilla & David R. Bickel (University of Ottawa)
Empirical Bayes Method Corrected for Small Number of Tests / Méthode bayésienne empirique corrigée pour un petit nombre de tests  
- 10:30 - 10:45 Min Chen, Changbao Wu & Mary Thompson (University of Waterloo)
Empirical Likelihood Method for Pretest-Posttest Studies / Méthode de vraisemblance empirique pour les études prétest/post-test  
- 10:45 - 11:00 G. Arthur Mihram (Princeton, NJ), Danielle Mihram (University of Southern California)
Confronting Quinn: Our Regression and Hypothesis-testing Methodologies in Modern Science / Face à Quinn : nos méthodologies, régression et tests d'hypothèses en science moderne.  
- 11:00 - 11:15 Eric Cai (University of Toronto)
A Pathological Monte Carlo Simulation of the Power Function of the Pearson Correlation Test / Simulation de Monte Carlo problématique pour la fonction puissance du test de corrélation de Pearson
- 11:15 - 11:30 Adrian Waddell & R. Wayne Oldford (University of Waterloo)
Extending RnavGraph to Use the R Plotting Facility / Extension de RnavGraph pour utiliser la fonctionnalité graphique de R  
- 11:30 - 11:45 Duncan Murdoch (University of Western Ontario)
tables: An R Package to Produce Tables / tables : Une librairie R pour produire des tableaux  

10:15 - 11:45 Session 2B-C4 (abstracts/résumés - page 154) **Rozanski Hall, room/local 102**
Model and Variable Selection**Sélection de modèle et de variables**

Contributed Paper Session / Séance de communications libres







Chair/Président: Patrice Mathieu (Statistics Canada / Statistique Canada)





- 10:15 - 10:30 Denis Talbot, Geneviève Lefebvre & Juli Atherton (Université du Québec à Montréal)
Bayesian Adjustment for Confounding : A Look at its Theoretical Basis / Ajustement bayésien pour la confusion : un regard sur la base théorique de la méthode  
- 10:30 - 10:45 Shinpei Imori, Hirokazu Yanagihara & Hirofumi Wakaki (Hiroshima University, Japan)
Theoretical Comparisons with AIC and TIC / Comparaisons théoriques de l'AIC et du TIC  
- 10:45 - 11:00 Ahmed Farahat, Ali Ghodsi & Mohamed Kamel (University of Waterloo)
An Efficient Greedy Method for Unsupervised Variable Selection / Une méthode gloutonne efficace de sélection de variables non supervisée  
- 11:00 - 11:15 Christian Léger & Pierre Lafaye de Micheaux (Université de Montréal)
A Law of the Single Logarithm for Weighted Sums of Arrays Applied to Bootstrap Model Selection in Regression / Loi du simple logarithme de sommes pondérées d'une suite triangulaire avec application au bootstrap en sélection de modèle  
- 11:15 - 11:30 Hirokazu Yanagihara, Hirofumi Wakaki & Yasunori Fujikoshi (Hiroshima University, Japan)
A Consistency Property of AIC for Multivariate Linear Models when the Dimension and the Sample Size are Large / Propriété de convergence de l'AIC pour les modèles linéaires multivariés de grandes dimensions et avec une grande taille d'échantillon  
- 11:30 - 11:45 Mariko Yamamura & Hirokazu Yanagihara (Hiroshima University, Japan), Muni Srivastava (University of Toronto)
Variable Selection by C_p Statistic in Multiple Responses Regression with Fewer Sample Size than the Dimension / Sélection de variables avec la statistique C_p dans des régressions à réponses multiples dont la taille d'échantillons est inférieure à la dimension  

10:15 - 11:30 Session 2B-C5 (abstracts/résumés - page 157) **Thornbrough Building, room/local 1307**
Theory and Applications of Copulas**Théorie et applications des copules**

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Johanna Nešlehová (McGill University)

- 10:15 - 10:30 Elif Fidan Acar (McGill University), Radu V. Craiu & Fang Yao (University of Toronto)
Statistical Testing for Conditional Copulas / Tests statistiques pour copules conditionnelles  
- 10:30 - 10:45 Tounkara Fode, Louis-Paul Rivest & Belkacem Abdous (Université Laval)
A Copula-based Estimator for the Intracluster Correlation Coefficient / Estimation du coefficient de corrélation intra-classe pour des données binaires basée sur des copules  
- 10:45 - 11:00 Héla Romdhani, Louis-Paul Rivest & Lajmi Lakhel-Chaieb (Université Laval)
Measuring and Testing the Intracluster Dependence for Non-normally Distributed Clustered Data / Mesurer et tester la dépendance intra-classe pour des données groupées non normales  

- 11:00 - 11:15 Bin Dong & Judy-Anne W. Chapman (NCIC Clinical Trials Group, Queen's University), Paul E. Goss (Massachusetts General Hospital Cancer Center, Harvard University), Michael N. Pol-lak (Lady Davis Institute, Jewish General Hospital, McGill University), Lois E. Shepherd (NCIC Clinical Trials Group, Queen's University)
Copula Estimation for Censored Data via Empirical Likelihood: NCIC CTG MA.27 and MA.14 Endocrine Breast Cancer Trials / Estimation par copules pour données censurées à l'aide d'une vraisemblance empirique : essais endocriniens NCIC CTG MA.27 et MA.14 sur le cancer du sein  
- 11:15 - 11:30 Beilei Wu & Alexander R. de Leon (University of Calgary)
Flexible Random Effects Copula Models for Clustered Mixed Outcomes – Application in Toxicology / Modèles de copules flexibles à effets aléatoires pour des résultats mixtes regroupés – application en toxicologie  

11:45 - 13:15
Lunch / Dîner

Creelman Hall

13:15 - 14:45 Session 2C-A (abstract/résumé - page 160) **Rozanski Hall, room/local 103**

Isobel Loutit Invited Address



Allocution Isobel Loutit

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Boxin Tang (Simon Fraser University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 13:15 - 14:45 Roger Hoerl (GE Global Research)
The Next Frontier in Business and Industrial Statistics: Attacking Complex, Unstructured Problems / La prochaine frontière de la statistique industrielle et d'entreprise : s'attaquer aux problèmes complexes et non structurés  







13:15 - 14:45 Session 2C-I1 (abstracts/résumés - page 161) **MacNaughton Building, room/local 105**

Absolute Penalty Estimation and Variable Selection

Estimation à l'aide de la perte absolue et sélection de variables

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Ejaz Ahmed (Brock University)

- 13:15 - 13:45 Yang Feng (Columbia University), Jianqing Fan & Xin Tong (Princeton University)
A Nonparametric Optimal Decision Rule in High Dimensional Space / Une règle de décision non-paramétrique optimale dans un espace à plusieurs dimensions  
- 13:45 - 14:15 M. Shakhawat Hossain (University of Winnipeg), Ejaz Ahmed (Brock University), Kjell A. Doksum (University of Wisconsin-Madison)
Shrinkage and Penalty Estimators in Generalized Linear Models / Estimateurs par rétrécissement et pénalité pour les modèles linéaires généralisés  
- 14:15 - 14:45 Paul McNicholas & Sakyajit Bhattacharya (University of Guelph)
A Modified BIC for Mixture Model Selection / Un CIB modifié pour la sélection d'un modèle de mélange  

13:15 - 14:45 Session 2C-I2 (abstracts/résumés - page 163) **Rozanski Hall, room/local 101**

New Statistical Design and Analysis for Epidemiological Investigations







Nouvelles méthodes de conception et d'analyse pour les enquêtes épidémiologiques

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Yutaka Yasui (University of Alberta)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 13:15 - 13:45 Masoud Asgharian, David Wolfson & Christina Wolfson (McGill University)
Prevalent Cohort Design and Analysis: A Case Study / Plan et analyse de cohorte prévalente :
une étude de cas  
- 13:45 - 14:15 Juxin Liu (University of Saskatchewan), Paul Gustafson (University of British Columbia)
Interactions and Average Predictive Comparison / Interactions et comparaison prédictive
moyenne  
- 14:15 - 14:45 Jose Miguel Martinez (Universitat Pompeu Fabra, Spain), José Maria Velarde (CiSAL, Univer-
sitat Pompeu Fabra, Spain), Ramon Clèries (Cancer Registry of Catalonia and University of
Barcelona, Spain / Cancer Registry of Catalonia et University of Barcelona, Spain), Joan Be-
nach (Health Inequalities Research Group, GREDS-EMCONET, Spain), Fernando G. Bena-
vides (CiSAL, Universitat Pompeu Fabra and CIBERESP, Spain / CiSAL, Universitat Pompeu
Fabra et CIBERESP, Spain), Yutaka Yasui (University of Alberta)
Improving Multilevel Analyses: the Integrated Epidemiologic Design and its Extension / Amé-
lioration des analyses multiniveaux : le plan d'expérience épidémiologique intégré et son ex-
tension  

13:15 - 14:45 Session 2C-I3 (abstracts/résumés - page 165) **Rozanski Hall, room/local 104**







Session in Honour of Jon N.K. Rao's 75th Birthday

Séance en l'honneur du 75e anniversaire de Jon N.K. Rao

Invited Paper Session / Présentations sur invitation

Chair/Président: Louis-Paul Rivest (Université Laval)

Organizer/Responsable: David Haziza (Université de Montréal)

- 13:15 - 13:45 Mike Hidioglou & Yong You (Statistics Canada / Statistique Canada)
Confidence Intervals for Unit and Area Level Small Area Estimators / Estimation par intervalles
de confiance pour petits domaines au niveau de l'unité ou du petit domaine  
- 13:45 - 14:15 Jae-Kwang Kim (Iowa State University), Seunghwan Park (Seoul National University, South Ko-
rea)
Fractional Imputation for Regression Analysis Combining Two Independent Surveys / Impu-
tation fractionnaire pour des analyses de régression dans le cas de deux enquêtes indépen-
dantes  
- 14:15 - 14:45 Mary Thompson (University of Waterloo)
Bootstrap Methods in Complex Surveys / Méthodes bootstrap dans les enquêtes com-
plexes  







13:15 - 14:45 Session 2C-I4 (abstracts/résumés - page 167) **Alexander Hall, room/local 100**

Statistical Analysis of Network Graph Data

Analyse statistique de réseaux et graphes

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Shojaeddin Chenouri (University of Waterloo)

- 13:15 - 13:45 Chris Groendyke (Robert Morris University), David Welch (University of Auckland, New Zealand), David R. Hunter (Pennsylvania State University)
A Network-Based Analysis of the 1861 Hagelloch Measles Data / Une analyse réseau des données sur l'épidémie de rougeole d'Hagelloch de 1861  
- 13:45 - 14:15 Patrick Perry & Nikolay Kulmatitskiy (New York University, Stern), Patrick Wolfe (University College London, U.K.)
A Parametric View of Some Nonparametric Network Algorithms / Une vision paramétrique de certains algorithmes réseau non paramétriques  
- 14:15 - 14:45 Ali Shojaie (University of Washington), Basu Sumanta & Michailidis George (University of Michigan)
Estimation of Network Granger Causal Models with Inherent Group Structures / Estimation des modèles de réseaux de liens de causalité au sens de Granger avec des structures de groupe inhérentes  

13:15 - 14:45 Session 2C-I5 (abstracts/résumés - page 169) **Alexander Hall, room/local 200**







Stochastic Methods in Finance
Méthodes stochastiques en finance

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Cody Hyndman (Concordia University)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité







- 13:15 - 13:45 Alexandru Badescu (University of Calgary), Juan-Pablo Ortega (Université de Franche-Comté, France)
Hedging GARCH Options with Generalized Innovations / Couverture d'options GARCH avec innovations généralisées  
- 13:45 - 14:15 Sebastian Jaimungal (University of Toronto), Alvaro Cartea (University College London, U.K.)
Risk Measures and Fine Tuning of High Frequency Trading Strategies / Mesures de risque et ajustement fin de stratégies de négociation à haute fréquence  
- 14:15 - 14:45 Adam Metzler, R. Mark Reesor & Jingya Li (University of Western Ontario)
Information, Ambiguity and the Cost of Contingent Capital / Information, ambiguïté et coût du capital conditionnel  





13:15 - 14:30 Session 2C-C1 (abstracts/résumés - page 171) **Rozanski Hall, room/local 102**

Models Accounting for Measurement Error
Modèles avec erreurs de mesure

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Michelle Qian Zhou (Harvard School of Public Health)

- 13:15 - 13:30 Ying Yan & Grace Y. Yi (University of Waterloo)
Methods for Additive Hazards Models with Covariate Measurement Error / Méthodes pour les modèles à risques additifs avec erreur de mesure dans les covariables  
- 13:30 - 13:45 Xue Yao, Lisa Lix & Liu Junxin (University of Saskatchewan)
Multiple Imputation to Correct Measurement Error: Application to Chronic Disease Case Ascertainment in Administrative Databases / Imputation multiple pour la correction de l'erreur de mesure : application à la confirmation des cas de maladie chronique dans les bases de données administratives  
- 13:45 - 14:00 Taraneh Abarin, Brajendra Sutradhar & Guang Sun (Memorial University of Newfoundland)
Gene-Environment Interaction Studies with Measurement Error / Études d'interaction gène-environnement avec erreur de mesure  









- 14:00 - 14:15 Wei Liu (York University), Lang Wu (University of British Columbia)
Simultaneous Inference for Semiparametric Nonlinear Mixed-effects Models with Covariate Measurement Errors and Missing Responses / Inférence simultanée pour modèles semi-paramétriques non-linéaires à effets mixtes avec erreur de mesure sur les covariables et données manquantes  
- 14:15 - 14:30 Anne Michele Millar & Rebecca Romo (Mount Saint Vincent University), Bethany White (University of Western Ontario), Candace Schau (CS Consultants, LLC, New Mexico)
Measuring Students' Attitudes to Statistics in the Presence of Regression-to-the-Mean / Mesurer l'attitude d'étudiants envers la statistique en présence de régression à la moyenne  

13:15 - 14:15 Session 2C-C2 (abstracts/résumés - page 174) **Rozanski Hall, room/local 105**

Nonparametric Methods

Méthodes non-paramétriques

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Elif Fidan Acar (McGill University)







- 13:15 - 13:30 Muhammad Islam (Grant MacEwan University)
Comparing Distributions Using Data Driven Statistics / Comparaison de distributions à l'aide de tests adaptatifs  
- 13:30 - 13:45 Thierry Duchesne, Lajmi Lakhal-Chaieb & Belkacem Abdous (Université Laval)
Estimation of the Conditional Survival Function Under Dependent Censoring / Estimation de la fonction de survie conditionnelle en présence de censure dépendante  
- 13:45 - 14:00 Mehrdad Gangeh, Ali Ghodsi & Mohamed Kamel (University of Waterloo)
Kernelized Supervised Dictionary Learning / Méthode du noyau pour l'apprentissage supervisé par dictionnaire  
- 14:00 - 14:15 Mohammad Nourmohammadi, Mohammad Jafari Jozani & Brad Johnson (University of Manitoba)
Nonparametric Confidence Intervals of Quantiles of a Finite Population using Randomized Nomination Sampling / Intervalles de confiance non-paramétriques pour quantiles dans une population finie à l'aide de l'échantillonnage randomisé par nomination  





13:15 - 14:30 Session 2C-C3 (abstracts/résumés - page 176) **Thornbrough Building, room/local 1307**

Spatial Statistics

Statistique spatiale

Contributed Paper Session / Séance de communications libres
Chair/Président: Douglas Woolford (Wilfrid Laurier University)

- 13:15 - 13:30 Olaf Berke (University of Guelph)
The 350th Anniversary of Statistical Epidemiology: A Spatial Re-analysis of John Graunt's London Plague Data / 350e anniversaire de l'épidémiologie statistique : analyse spatiale des données sur la peste de Londres de John Graunt  
- 13:30 - 13:45 Masud Rana, Shahedul A. Khan & Longhai Li (University of Saskatchewan), Joel A. Dubin (University of Waterloo)
Monitoring and Understanding Atmospheric CFC Decline with the Spatial-Longitudinal Bent-Cable Model / Surveillance et compréhension du déclin des CFC atmosphériques par le modèle "bent-cable" longitudinal  
- 13:45 - 14:00 Jonathan Lee, Reg Kulperger & Hao Yu (University of Western Ontario)
Large-scale Spatial Simulation and Computation Using Parallel Computing / Simulation spatiale à grande échelle et calcul parallèle  

- 14:00 - 14:15 Alisha Albert-Green (Princess Margaret Hospital), Charmaine Dean (University of Western Ontario), Douglas Woolford (Wilfrid Laurier University), David Martell (University of Toronto)
A Nonparametric Framework for Quantifying Temporal Trends in the Seasonality of Forest Fire Risk / Cadre non-paramétrique de quantification des tendances temporelles de la saisonnalité du risque d'incendies forestiers  
- 14:15 - 14:30 Devan Becker & Douglas Woolford (Wilfrid Laurier University), David Martell (University of Toronto)
Exploring the Lifetime of Forest Fires / Exploration de la durée de vie des incendies forestiers  

13:15 - 14:45 Session 2C-C4 (abstracts/résumés - page 179) **MacNaughton Building, room/local 113**

Studying Gene Expression/Association

Étude de l'expression et de l'association des gènes

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Shelley Bull (Samuel Lunenfeld Research Institute)

- 13:15 - 13:30 Farnoosh Abbas Aghababazadeh, Mayer Alvo & David R. Bickel (University of Ottawa)
Statistical Methods to Analyze Genome-Wide Association Data / Méthodes statistiques pour analyser des données d'association sur tout le génome  
- 13:30 - 13:45 Gun Ho Jang & Rui Feng (University of Pennsylvania)
A Statistical Model for Copy Number Polymorphism Calling Using Haplotypes / Un modèle statistique pour l'identification du nombre de copies de polymorphismes à l'aide d'haplotypes  
- 13:45 - 14:00 Ahmed Hossain (McMaster University), Andrew Willan (University of Toronto), Joseph Beyene (McMaster University)
A Flexible Nonparametric Approach for Identifying Correlated Differentially Expressed Genes in Microarray Experiments / Une approche non-paramétrique flexible pour identifier des gènes corrélés et exprimés différemment dans des expériences de micropuces  
- 14:00 - 14:15 Thierry Chekouo Tekougang & Alejandro Murua (Université de Montréal)
The Gibbs-plaid Biclustering Model / Le modèle Gibbs-plaid de biclustering  
- 14:15 - 14:30 Maggie Chu & Wenqing He (University of Western Ontario)
Model-Based Clustering of Time-Course RNA-seq Data / Regroupement modélisé de données ARN-seq sur les temps d'absorption  
- 14:30 - 14:45 Martin Leclerc & Lajmi Lakhil-Chaieb (Université Laval)
Estimation of Dependence Between Lifetimes in the Presence of Selection Bias: the Case of Breast Cancer / Estimation de la dépendance entre des durées de vie en présence de biais de sélection : le cas du cancer du sein  

13:15 - 15:15 Session 2C-P (abstracts/résumés - page 182)

Science Complex Atrium

Poster Session II

Séance d'affichage II

Contributed Poster Session / Séance d'affichage libre

- 13:15 - 15:15 Justin Angevaere & Daniel Gillis (University of Guelph)
The Utility of Catch Per Unit Effort Variance / Utilité de la variance des captures par unité d'effort
- 13:15 - 15:15 Shaheena Bashir (University Health Network, Toronto), Celia Greenwood (Lady Davis Institute for Medical Research and McGill University / Lady Davis Institute for Medical Research et McGill University)
Joint Tests of Effect Size in Multivariate Meta-Analysis / Tests simultanés de l'ampleur de l'effet dans une méta-analyse multivariée

- 13:15 - 15:15 Victoria Borg Debono, Shiyuan (William) Zhang, Chenglin Ye & James Paul (McMaster University), Lindsay Hurlburt (University of Toronto), Yamini Murthy & Lehana Thabane (McMaster University)
The Quality of Reporting of Randomized Controlled Trials (RCTs) in Postoperative Pain Management / Qualité des rapports sur les essais contrôlés randomisés (ECR) dans la gestion de la douleur postopératoire
- 13:15 - 15:15 Zhengfei Chen (University of Toronto)
Riemann Manifold HMCMC for Log-Gaussian Cox Processes / MCMC hamiltonien sur des variétés de Riemann pour des processus de Cox log-gaussiens
- 13:15 - 15:15 Jing Ding (Acadia University)
Population Assessments by Using Age-at-harvest Data and Catch Effort / Évaluations des populations au moyen des données sur l'âge des prises et l'effort de pêche
- 13:15 - 15:15 Jayde Eustace & J. Concepción Loredó-Osti (Memorial University of Newfoundland)
A Gumbel State Space Model for Extreme-value Data / Modèle à espace d'états de Gumbel pour des valeurs extrêmes
- 13:15 - 15:15 Remya M. John & Angelo Canty (McMaster University)
Statistical Analysis of RNA-Seq Data: A Comparison of Three R Packages / Analyse statistique de données d'ARN-Seq : comparaison de trois bibliothèques R
- 13:15 - 15:15 Xiaoqing Niu (University of Alberta), Pengfei Li (University of Waterloo)
Testing Homogeneity in Normal Scale Mixture Models / Test d'homogénéité pour des modèles gaussiens à dispersion aléatoire
- 13:15 - 15:15 Matthew Rueffer, Julie Horrocks & Daniel Gillis (University of Guelph)
Spatial and Temporal Modelling of Change in Marine Harvests / Modélisation spatiale et temporelle de la variation des pêches
- 13:15 - 15:15 Kathleen Ryan, Daniel Gillis & Stephen Crawford (University of Guelph)
Depth Distribution of Larval Lake Whitefish (*Coregonus clupeaformis*) in a Great Lakes Embayment / Distribution en profondeur des larves de grand corégone (*Coregonus clupeaformis*) dans une baie des Grands Lacs.
- 13:15 - 15:15 Olga Shestakovska, Janice Pogue, John Eikelboom & Stuart J. Connolly (Population Health Research Institute, McMaster University and Hamilton Health Sciences)
Large Hazard Ratios Do Not Always Mean High Predictive Accuracy: A Simulation Study / De grands ratios de risque ne signifient pas toujours une grande exactitude prédictive : étude de simulation
- 13:15 - 15:15 Monica Sirski (University of Manitoba)
On the Statistical Analysis of Functional Data Arising from Designed Experiments / À propos de l'analyse statistique de données fonctionnelles découlant de plans d'expérience
- 13:15 - 15:15 Gary Sneddon (Mount Saint Vincent University), Matthew Kerby (Memorial University of Newfoundland)
Panel Data Models and Electoral Success of Ex-Cabinet Ministers / Modèles pour données longitudinales et succès électoral d'anciens ministres
- 13:15 - 15:15 Mary Thiess (Fisheries and Oceans Canada / Pêches et Océans Canada), Skip McKinnell (PICES - North Pacific Marine Organization), Marc Trudel (Fisheries and Oceans Canada / Pêches et Océans Canada)
Time Series Analysis: Assessing the "pink salmon" effect / Analyse de séries chronologiques : évaluation de "l'effet saumon rose"
- 13:15 - 15:15 Monica H.T. Wong, Paul McNicholas & David M. Mutch (University of Guelph)
Cluster Analyses to Assess Weight Loss Maintenance: An Application of Clustering in Nutrigenomics / Analyse de regroupement pour évaluer le maintien d'une perte de poids : application des analyses de regroupements à la nutriginomique
- 13:15 - 15:15 Haiyan Yang & J. Concepción Loredó-Osti (Memorial University of Newfoundland)
Joint Modelling of Genetic Linkage and Association / Modélisation conjointe de la liaison et de l'association génétique

- 13:15 - 15:15 Li Hua Yue & Wenqing He (University of Western Ontario)
Logistic Prediction on Prostate Cancer Status Using Imaging Data / Prévisions logistiques de l'état du cancer de la prostate à l'aide de données d'imagerie
- 13:15 - 15:15 Wei Yuhong & Paul McNicholas (University of Guelph)
Mixture-model Averaging / Moyennage de modèles de mélange







14:45 - 15:15
Coffee Break
Pause-café

Science Complex Atrium

15:15 - 16:45 Session 2D-I1 (abstracts/résumés - page 190) **Rozanski Hall, room/local 103**
Nonparametric Bayes Methods
Méthodes bayésiennes non paramétriques

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Saman Muthukumarana (University of Manitoba)

- 15:15 - 15:45 Alexandre Bouchard-Côté, Ardavan Saeedi & John Petkau (University of British Columbia), Arnaud Doucet (University of Oxford, U.K.)
The Gamma-Exponential Process / Le processus gamma exponentiel  
- 15:45 - 16:15 Simon Guillotte (Université du Québec à Montréal), François Perron (Université de Montréal)
A Sieves Model for Copulas / Un modèle de tamis pour copules  
- 16:15 - 16:45 Tim Swartz (Simon Fraser University), Paramjit Gill (University of British Columbia, Okanagan)
Stylometric Analyses using Dirichlet Process Mixture Models / Analyses stylométriques par l'utilisation de modèles de mélanges de processus de Dirichlet  









15:15 - 16:35 Session 2D-I2 (abstracts/résumés - page 192) **Rozanski Hall, room/local 101**
Recent Advances in Analysis of Survival Data from Cross Sectional Sampling
Progrès récents en analyse de survie à partir de données transversales

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Masoud Asgharian (McGill University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 15:15 - 15:35 Pierre-Jérôme Bergeron (University of Ottawa), Masoud Asgharian (McGill University), Ewa Makvandi-Nejad (University of Ottawa)
Nonparametric Estimation of the Survival Function for Prevalent and Incident Cases under Stationary Incidence / Estimation nonparamétrique de la fonction de survie pour des cas prévalents et incidents lorsque l'incidence est stationnaire  
- 15:35 - 15:55 Marco Carone (University of California, Berkeley), Daniel Scharfstein (Johns Hopkins Bloomberg School of Public Health), Mark van der Laan (University of California, Berkeley), Masoud Asgharian (McGill University)
Quantifying the Impact of Disease on Lifetime Under Cross-sectional Sampling / Quantifier l'impact d'une maladie sur la durée de vie à l'aide d'un échantillon transversal  
- 15:55 - 16:15 Ashkan Ertefaei (University of Michigan), Masoud Asgharian & David Stephens (McGill University)
Double Bias / Double biais  
- 16:15 - 16:35 Taki R. Shinohara & Mei-Cheng Wang (Johns Hopkins Bloomberg School of Public Health)
Alternating Event Processes during Lifetimes: Population Dynamics and Statistical Inference / Processus d'événements alternant au cours de la durée de vie : dynamique des populations et inférence statistique  

15:15 - 16:45 Session 2D-I3 (abstracts/résumés - page 195) Alexander Hall, room/local 200

Teaching and Learning Statistics with Technology

Enseignement et apprentissage de la statistique à l'aide des nouvelles technologies







Invited Paper Session / Présentations sur invitation

Chair/Président: Saumen Mandal (University of Manitoba)

Organizers/Responsables: Saumen Mandal and/et Zeneida Mateo (University of Manitoba)

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

- 15:15 - 15:45 Bruce Dunham & Eugenia Yu (University of British Columbia), Joanne Quan (University of Toronto)
Using On-Line Simulation Tools to Promote Student Understanding of Sampling Distributions.
/ L'utilisation d'outils de simulation en ligne pour favoriser la compréhension par les étudiants
des distributions d'échantillonnage  
- 15:45 - 16:15 Bethany White (University of Western Ontario)
Harnessing the Power of Technology to Get Students Talking about Statistics / Exploiter la
puissance de la technologie pour inciter les étudiants à parler de statistique  
- 16:15 - 16:45 Roger Woodard (North Carolina State University)
A Large Course Redesign of Introductory Statistics: You Don't Need 65 Other People with
You to Learn a Definition / Un grand remaniement d'un cours d'introduction à la statistique :
vous n'avez pas besoin de 65 personnes à vos côtés pour apprendre une définition  

15:15 - 16:45 Session 2D-I4 (abstracts/résumés - page 197) Rozanski Hall, room/local 104







The Census of Population in Canada: Looking Back at 2011 and Into the Future

Le Recensement de la population au Canada : Retour sur 2011 et regard vers l'avenir

Invited Paper Session / Présentations sur invitation

Chair/Président: Don McLeish (University of Waterloo)

Organizer/Responsable: Julie Trépanier (Statistics Canada / Statistique Canada)

- 15:15 - 15:45 Patrice Mathieu (Statistics Canada / Statistique Canada)
Data Collection for the 2011 Census and National Household Survey / La collecte du Recen-
sement de la population et de l'Enquête nationale auprès des ménages de 2011  
- 15:45 - 16:15 Normand Laniel (Statistics Canada / Statistique Canada)
The Estimation Methodology for the 2011 Census of Population and the National Household
Survey / La méthodologie d'estimation du Recensement de la population et de l'Enquête
nationale auprès des ménages de 2011  
- 16:15 - 16:45 Julie Trépanier (Statistics Canada / Statistique Canada)
The 2016 Census Strategy Project / Le projet de la stratégie du Recensement de 2016  

15:15 - 16:45 Session 2D-I5 (abstracts/résumés - page 199) Alexander Hall, room/local 100

Theory and Applications of Self-Normalized Processes

Théorie et application des processus auto-normalisés







Invited Paper Session / Présentations sur invitation

Chair/Président: Hao Yu (University of Western Ontario)

Organizer/Responsable: Yuliya Martsynyuk (University of Manitoba)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

- 15:15 - 15:45 Yuliya Martsynyuk (University of Manitoba)
Invariance Principles for a Multivariate Student Process in the Generalized Domain of Attraction of the Multivariate Normal Law / Principes d'invariance d'un processus multivarié de Student dans le domaine d'attraction généralisé d'une loi normale multivariée  
- 15:45 - 16:15 Masoud Nasari & Miklós Csörgő (Carleton University)
Another Look at Bootstrapping the Student t-Statistic / Un autre regard sur le bootstrap de la statistique t de Student  
- 16:15 - 16:45 Miklós Csörgő (Carleton University)
Weak Invariance Principles and Strong Limit Laws for Self-Normalized Processes / Principes d'invariance faible et lois limites fortes pour des processus autonormalisés  

15:15 - 16:30 Session 2D-C1 (abstracts/résumés - page 201) **Rozanski Hall, room/local 105**

Business and Industrial Statistics and Quality Control

Statistique industrielle et de gestion et contrôle de la qualité

Contributed Paper Session / Séance de communications libres

Chair/Président: Roman Viveros-Aguilera (McMaster University)

- 15:15 - 15:30 Davtalab Olyaie Mostafa & Gholamreza Jahanshahloo (Tarbiat Moallem University, Iran), Masoud Asgharian (McGill University), Vahid Partovi Nia (École Polytechnique de Montréal)
Efficiency in Stochastic Data Envelopment Analysis / Efficacité dans l'analyse d'enveloppement de données stochastiques  
- 15:30 - 15:45 Gemai Chen (University of Calgary)
Discovery Models — A Least Squares Approach With Applications / Modèles de découverte — méthode des moindres carrés et applications  
- 15:45 - 16:00 Amir Hosein Homaie Shandizi & Vahid Partovi Nia (École Polytechnique de Montréal)
A Test for Attribute Gauge R&R / Un test pour le gage R&R  
- 16:00 - 16:15 Jayasankar Vatathoor & Asokan Mulayath Variyath (Memorial University of Newfoundland)
Robust Control Charts for Multivariate Individual Observations / Cartes de contrôle robustes pour les observations individuelles multivariées  
- 16:15 - 16:30 Ick Huh, Roman Viveros-Aguilera & Narayanaswamy Balakrishnan (McMaster University)
The MEWMA with Two Unequal Smoothing Parameters / Le MEWMA avec deux paramètres de lissage inégaux  







15:15 - 16:45 Session 2D-C2 (abstracts/résumés - page 204) **Rozanski Hall, room/local 102**







Financial Applications

Applications en finance

Contributed Paper Session / Séance de communications libres

Chair/Président: Michael Evans (University of Toronto)

- 15:15 - 15:30 René Ferland & François Watier (Université du Québec à Montréal)
Mean-Variance Efficiency in a Market with Regime-Switching Zero-Coupons / Efficacité moyenne-variance dans un marché à changement de régime  
- 15:30 - 15:45 François Watier (Université du Québec à Montréal), Chantal Labbé (HEC Montréal)
Goal Achieving Probabilities of Optimal Cone-Constrained Mean-variance Portfolios / Probabilités d'atteinte des objectifs de portefeuilles moyenne-variance optimaux à contrainte conique  
- 15:45 - 16:00 Taehan Bae (University of Regina), Ian Iscoe (Algorithmics Inc.)
Sum of Bernoulli Mixtures: Beyond Conditional Independence / Somme de mélanges de Bernoulli : au-delà de l'indépendance conditionnelle  

- 16:00 - 16:15 Zhenyu Cui, Carole Bernard & Don McLeish (University of Waterloo)
Nearly Exact Option Price Simulation using Characteristic Functions / Simulation presque exacte de prix d'options à partir de fonctions caractéristiques  
- 16:15 - 16:30 Brian M. Hartman (University of Connecticut), Chris Groendyke (Robert Morris University)
Model Selection and Averaging in Regime-Switching Models of Various Types / Sélection et moyennage de modèles à changements de régimes de divers types  
- 16:30 - 16:45 Chengguo Weng (University of Waterloo)
Constant Proportion Portfolio Insurance under Regime Switching Lévy Exponential Process / Assurance portefeuille à proportions constantes dans le cadre d'un processus exponentiel de Lévy avec changements de régime  



15:15 - 16:45 Session 2D-C3 (abstracts/résumés - page 207) **MacNaughton Building, room/local 105**

Mixture Models, Clustering and Classification

Modèles de mélanges, partitionnement et classification

Contributed Paper Session / Séance de communications libres

Chair/Président: Abbas Khalili (McGill University)

- 15:15 - 15:30 Liam Callaghan & Ayesha Ali (University of Guelph)
Modelling Plant-Pollinator Interactions with Mixtures of Linkage Rules / Modélisation des interactions plante-pollinisateur avec diverses règles de liaison  
- 15:30 - 15:45 Armin Hatefi & Mohammad Jafari Jozani (University of Manitoba)
Fisher Information in Ranked Set Samples from Finite Mixture Models / Information de Fisher et échantillonnage d'ensembles ordonnés appliqués aux modèles de mélanges finis  
- 15:45 - 16:00 Pooyan Khajehpour Tadavani & Ali Ghodsi (University of Waterloo)
Partitioning High Dimensional Data to Low Dimensional and Localized Subspaces / Partitionnement de données de grande dimension en sous-espaces de petite dimension localisés  
- 16:00 - 16:15 Irene Urbik & Paul McNicholas (University of Guelph)
Clustering and Classification with Parsimonious Skew-t Mixture Models / Regroupement et classification à l'aide de modèles de mélange parcimonieux de distributions t asymétriques  
- 16:15 - 16:30 Brian Franczak, Paul McNicholas & Ryan Browne (University of Guelph)
The ParSAL Family of Mixture Models / Modèles de mélanges de distributions d'une famille parcimonieuse de lois Laplace asymétriques décentrées  
- 16:30 - 16:45 Rachel O'Reilly & Paul McNicholas (University of Guelph)
Cross-Validation for Mixture Model Selection / Validation croisée pour sélection de modèles de mélange  



15:15 - 16:45 Session 2D-C4 (abstracts/résumés - page 210) **MacNaughton Building, room/local 113**

Modelling the Spread of Infectious Diseases

Modélisation de la propagation des maladies infectieuses

Contributed Paper Session / Séance de communications libres

Chair/Président: Patrick Brown (Cancer Care Ontario)

- 15:15 - 15:30 Lin Zhang & Rob Deardon (University of Guelph)
Comparing Bayesian Statistical Models of Infectious Disease Outbreaks via Probability Scoring / Comparaison de modèles bayésiens d'épidémies de maladies infectieuses au moyen d'une pondération probabiliste  

- 15:30 - 15:45 Lilia Leticia Ramirez Ramirez (Instituto Tecnológico Autónomo de México (ITAM)), Mary Thompson & Yulia Gel (University of Waterloo), Eileen De Villa (Peel Public Health), Matt McPherson (Infonaut)
Simulation of Infectious Disease Spread as a Tool to Evaluate Control Measures During an Epidemic Outbreak / Simulation de la propagation d'une maladie infectieuse comme outil pour évaluer les mesures de lutte en cas de flambée épidémique  
- 15:45 - 16:00 Nadia Bifulchi, Rob Deardon & Zeny Feng (University of Guelph)
Spatial Approximations of Network-based Individual Level Infectious Disease Models / Approximations spatiales des modèles de maladie infectieuse individualisés en réseau  
- 16:00 - 16:15 Lorna Deeth, Rob Deardon & Daniel Gillis (University of Guelph)
Variations of the Deviance Information Criterion for Infectious Disease Models / Variations sur le critère d'information de déviance pour les modèles de maladies infectieuses  
- 16:15 - 16:30 Chen Zhang & Alwell Oyet (Memorial University of Newfoundland)
Estimation and Forecasting of a Lag 2 Dynamic Model for Infectious Diseases / Estimation et prévision d'un modèle dynamique à décalage de deux périodes pour les maladies infectieuses  
- 16:30 - 16:45 Rajat Malik, Rob Deardon & Grace P.S. Kwong (University of Guelph), Ben Cowling (University of Hong Kong)
Modeling the Spread of Influenza within Households in Hong Kong / Modéliser la propagation de l'influenza dans les ménages de Hong Kong  

16:50 - 17:50

Rozanski Hall, room 102

Information on 2012 NSERC Discovery Grant Competition

Information sur le concours 2012 des subventions à la découverte du CRSNG



Wednesday June 6**mercredi 6 juin**

8:45 - 9:45 Session 3A-A1 (abstract/résumé - page 213) Rozanski Hall, room/local 101
CJS Award Address**Allocution du récipiendaire du Prix de la RCS**

Special Session / Conférence spéciale

Chair/Président: Paul McNicholas (University of Guelph)



Organizer/Responsable: Denis Larocque (HEC Montréal)

8:45 - 9:45 Chong Gu and Ping Ma (Purdue University; University of Illinois, Urbana-Champaign)
 Nonparametric regression with cross-classified responses / Régression non-paramétrique
 avec variables réponses provenant de classifications croisées  

8:45 - 9:45 Session 3A-A2 (abstract/résumé - page 214) Rozanski Hall, room/local 104
CRM-SSC Prize in Statistics Address**Allocution du récipiendaire du Prix CRM-SSC**

Special Session / Conférence spéciale



Organizer and Chair / Responsable et président: David Stephens (McGill University)

8:45 - 9:45 Changbao Wu (University of Waterloo), Jiahua Chen (University of British Columbia), Jae-Kwang
 Kim (Iowa State University)
 Semiparametric Fractional Imputation and Sparse and Efficient Replication Variance Estima-
 tion for Complex Surveys / Imputation fractionnaire semi-paramétrique et estimation parcimo-
 nieuse et efficace de la variance par réplication pour des enquêtes complexes  

8:45 - 9:45 Session 3A-A3 (abstract/résumé - page 215) Rozanski Hall, room/local 103
Pierre Robillard Award Address**Allocution du récipiendaire du Prix Pierre-Robillard**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et présidente: Keumhee Carriere Chough (University of Alberta)

8:45 - 9:45 Bei Chen (McMaster University), Bovas Abraham & Yulia Gel (University of Waterloo)
 Linearization Methods in Time Series Analysis / Méthodes de linéarisation pour l'analyse de
 séries chronologiques  

9:45 - 10:15**Coffee Break****Pause-café****Science Complex Atrium**

10:15 - 11:45 Session 3B-I1 (abstracts/résumés - page 216) Alexander Hall, room/local 100
Confidentiality**Confidentialité**



Invited Paper Session / Présentations sur invitation





Chair/Président: Mike Hidioglou (Statistics Canada / Statistique Canada)

Organizer/Responsable: Jean-Louis Tambay (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

10:15 - 10:45 Anne-Sophie Charest (Université Laval)
 Differential Privacy for Synthetic Datasets / Confidentialité différentielle pour jeux de données
 synthétiques  

- 10:45 - 11:15 Éric Gagnon & Valérie Roy (Institut de la statistique du Québec)
Statistical Disclosure Control of Remote Access Microdata Files / Contrôle de divulgation statistique pour fichiers de microdonnées accessibles à distance  
- 11:15 - 11:45 Julia Hoeninger (Federal Statistical Office, Germany), Jörg Höhne (State Statistical Institute Berlin-Brandenburg, Germany)
Morpheus – Remote Access to Micro Data / Morpheus – Accès à distance aux microdonnées  

10:15 - 11:45 Session 3B-I2 (abstract/résumé - page 218) **Rozanski Hall, room/local 104**

Current Issues in the Design and Analysis of Cluster Randomization Trials



Problèmes actuels dans la conception et l'analyse d'essais randomisés par grappes

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Yutaka Yasui (University of Alberta)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 10:15 - 11:15 Allan Donner (University of Western Ontario)
Current Issues in the Design and Analysis of Cluster Randomization Trials / Problèmes actuels dans la conception et l'analyse d'essais randomisés par grappes  
- 11:15 - 11:45 Robert Platt, Discussant (McGill University)







10:15 - 11:45 Session 3B-I3 (abstracts/résumés - page 219) **Rozanski Hall, room/local 103**

Spatial and Environmental Statistics

Statistique spatiale et environnementale

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Theodoro Koulis (University of Manitoba)

- 10:15 - 10:45 Rob Deardon (University of Guelph)
Computationally Efficient Forms of Individual-Level Models for Large-Scale Spatial Infectious Disease Systems / Formes de modèles à temps de calcul efficace pour systèmes spatiaux de maladies infectieuses à grande échelle au niveau des individus  
- 10:45 - 11:15 Subash Lele, Peter Solymos & Erin Bayne (University of Alberta)
Zero Inflated Models, Detection Error and Conditional Likelihood Inference with Applications in Ecology / Modèles à surabondance de zéros, erreur de détection et inférence fondée sur la vraisemblance conditionnelle avec applications en écologie  
- 11:15 - 11:45 Jeffrey Picka (University of New Brunswick)
Assessing the Fit of Models for Disordered Spatial Patterns / Évaluation de l'adéquation de modèles pour les structures spatiales désordonnées  

10:15 - 11:45 Session 3B-I4 (abstracts/résumés - page 221) **Alexander Hall, room/local 200**

Stochastic Models and Applications



Modèles stochastiques et applications





Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Shui Feng (MacMaster University)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

- 10:15 - 10:45 Deli Li (Lakehead University)
On Jiang's Asymptotic Distribution of the Largest Entry of a Sample Correlation Matrix / À propos de la distribution asymptotique de Jiang du plus grand élément d'une matrice de corrélation échantillonnale  

- 10:45 - 11:15 Mahmoud Zarepour & Luai Al Labadi (University of Ottawa)
Some Asymptotic Theory in Nonparametric Bayesian Inference / Résultats asymptotiques pour l'inférence bayésienne non-paramétrique  
- 11:15 - 11:45 Fang Xu & Shui Feng (McMaster University)
The Gamma-Dirichlet Structure and the Quasi Invariance Property / La structure de Gamma-Dirichlet et la propriété de quasi-invariance  

10:15 - 11:45 Session 3B-I5 (abstracts/résumés - page 223) **Rozanski Hall, room/local 101**

Teaching Statistical Consulting with a Mixture of Classroom and Practicum







Enseignement de la consultation statistique : un mélange d'exposés magistraux et d'expériences pratiques

Invited Paper Session / Présentations sur invitation

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

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

- 10:15 - 10:45 Tony Panzarella (University of Toronto)
Lab Course in Biostatistics – Providing Practical Experience for Future Consulting Biostatisticians / Cours de laboratoire en biostatistique - Fournir une expérience pratique de consultation aux futurs biostatisticiens  
- 10:45 - 11:15 Tamara Arenovich (University of Toronto)
Biostatistics Practicum – A Supervisor's Perspective / Stage en biostatistique - La perspective d'une superviseure  
- 11:15 - 11:45 David Soave (University of Toronto)
Statistical Consulting with a Mixture of Classroom and Practicum – A Student Perspective / Consultation statistique avec un mélange de cours et de pratique – Une perspective étudiante  









10:15 - 11:30 Session 3B-C1 (abstracts/résumés - page 225) **MacNaughton Building, room/local 105**



Clustering and Classification

Méthodes de partitionnement et de classification

Contributed Paper Session / Séance de communications libres

Chair/Président: Matías Salibián-Barrera (University of British Columbia)

- 10:15 - 10:30 Malika Charrad, Nadia Ghazzali, Azam Niknafs, Véronique Boiteau & Marc-Olivier Robert Lambert (Université Laval)
An Examination of Indices for Determining the Number of Clusters : the Nb.Clusters Package / Analyse des indices de détermination du nombre de classes : la librairie Nb.Clusters  
- 10:30 - 10:45 Thuvaraha Vanniyasingam & Lehana Thabane (McMaster University), Reitze N. Rodseth (University of KwaZulu-Natal, South Africa), Giovana A. Lurati Buse & Daniel Bolliger (University Hospital Basel, Switzerland), Bruce M. Biccard (University of KwaZulu-Natal, South Africa), Christoph S. Burkhardt (University Hospital Basel, Switzerland)
Using the Minimum P-Value Method to Predict the Occurrence of Cardiovascular Complications Within 30 Days of Vascular Surgery / Utilisation de la méthode de la valeur p minimale pour prévoir l'apparition de complications cardiovasculaires dans les 30 jours suivant une chirurgie vasculaire  
- 10:45 - 11:00 Tolulope Sajobi, Lisa Lix & Bolanle Dansu (University of Saskatchewan)
Robust Measures of Variable Importance for Two Independent Groups / Mesures robustes de l'importance des variables pour deux groupes indépendants  
- 11:00 - 11:15 Javed Hossain Tomal, William Welch & Ruben H. Zamar (University of British Columbia)
Ensembling Descriptor Sets to Rank Activity of Compounds in QSAR Studies / Assemblage d'ensembles de descripteurs pour classer l'activité des substances dans les études RQSA  

11:15 - 11:30 Daniel Gillis (University of Guelph)
Signal Processing for Species Identification / Traitement des signaux pour l'identification d'espèces  

10:15 - 11:45 Session 3B-C2 (abstracts/résumés - page 228) **MacNaughton Building, room/local 113**

Mixed Effects Models

Modèles à effets mixtes

Contributed Paper Session / Séance de communications libres

Chair/Président: Georges Monette (York University)

10:15 - 10:30 Vahid Partovi Nia (École Polytechnique de Montréal), Reza Drikvandi (Shahid Beheshti University, Iran), Geert Verbeke (Katholieke Universiteit Leuven, Belgium), Ahmad Khodadadi (Shahid Beheshti University, Iran)
Testing Variance Components in Linear Mixed Effects Models / Le test de composantes de la variance dans les modèles linéaires avec effets mixtes  

10:30 - 10:45 Yan Yan Wu (Samuel Lunenfeld Research Institute), Georges Monette & Augustine Wong (York University), Laurent Briollais (Samuel Lunenfeld Research Institute)
Third-Order Method For the Test of Variance Components in Linear Mixed Model / Méthode de troisième ordre pour le test des composantes de la variance dans le modèle linéaire mixte  

10:45 - 11:00 Eleanor M. Pullenayegum & Feng Xie (McMaster University)
Methodology for Scoring the EQ-5D / Méthodologie de notation du EQ-5D  

11:00 - 11:15 Weiqiang Wang & Zeny Feng (University of Guelph)
Modeling the Uncertainty of Susceptible and Non-susceptible Multiple Outcomes / Modélisation de l'incertitude de résultats multiples sensibles et non sensibles  

11:15 - 11:30 Haocheng Li & Grace Y. Yi (University of Waterloo)
A Variable Selection Method for Random Effects Models / Une méthode de sélection de variables pour les modèles à effets aléatoires  

11:30 - 11:45 Jesse D. Raffa & Joel A. Dubin (University of Waterloo)
Multivariate Longitudinal Data Analysis Using Hidden Markov Models / Analyse de données longitudinales multivariées par modèles de Markov cachés  



10:15 - 11:45 Session 3B-C3 (abstracts/résumés - page 231) **Rozanski Hall, room/local 102**



Models with Binary Responses









Modèles à variables indépendantes binaires

Contributed Paper Session / Séance de communications libres

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

10:15 - 10:30 Alan Ker (University of Guelph), Tolga Ergun (Suffolk University), Abdoul Sam (Ohio State University)
Semiparametric Estimation of the Link Function in Binary-Choice Single-Index Models / Estimation semi-paramétrique de la fonction lien dans les modèles de choix binaire à indice simple  

10:30 - 10:45 François Larochelle & Gou Tingting (Research and Development Canada, Department of National Defence)
A Complementary Perspective on the Use of ROC Curve to Assess Logistic Regression Models / Une perspective complémentaire sur l'utilisation de la courbe ROC pour évaluer un modèle de régression logistique  

- 10:45 - 11:00 Jinhui Ma, Parminder Raina, Joseph Beyene & Lehana Thabane (McMaster University)
Comparison of Marginal and Cluster-Specific Models in Analysis of Cluster Randomized Trials with Missing Binary Outcomes / Comparaison de modèles marginaux et spécifiques aux grappes dans l'analyse d'essais randomisés avec des réponses binaires manquantes  
- 11:00 - 11:15 Bingrui Sun & Brajendra Sutradhar (Memorial University of Newfoundland)
Correlations Versus Odds Ratios Based Inferences in Bivariate Binary Models / Inférences fondées sur les corrélations ou les rapports de cotes dans les modèles binaires bivariés  
- 11:15 - 11:30 Rong Chu, Stephen D. Walter, Gordon Guyatt, P.J. Devereaux, Michael Walsh, Kristian Thorlund & Lehana Thabane (McMaster University)
Methodological Issues on Prognostic Imbalance in Randomized Controlled Trials (RCTs) Evaluating A Binary Outcome / Questions méthodologiques concernant le déséquilibre pronostique dans les essais contrôlés randomisés (ECR) évaluant un résultat binaire  
- 11:30 - 11:45 Rafiqul Chowdhury & Yun-Hee Choi (University of Western Ontario)
A Prediction Model to Identify Elderly Who Get Hospitalized – An Application Using Longitudinal Data / Modèle de prévision pour identifier les aînés à hospitaliser – une application fondée sur des données longitudinales  









10:15 - 11:15 Session 3B-C4 (abstracts/résumés - page 234) Rozanski Hall, room/local 105

Spectral Methods in Time series

Analyse spectrale de séries chronologiques

Contributed Paper Session / Séance de communications libres

Chair/Président: Rafal Kulik (University of Ottawa)

- 10:15 - 10:30 Charlotte L. Haley & David J. Thomson (Queen's University)
Jackknifed Error for Multitaper Estimates of Autocovariance in Time Series / Estimation de l'erreur par la méthode du jackknife pour les estimateurs à fenêtres de pondération multiples de l'autocovariance dans des séries chronologiques  
- 10:30 - 10:45 Aaron Springford (Queen's University)
Identification of Time Series Outliers Applied to Schumann Resonance Data / Identification des valeurs aberrantes dans des séries chronologiques avec application aux données sur la résonance de Schumann  
- 10:45 - 11:00 David J. Thomson (Queen's University)
Incautious Use of Robust Spectrum Estimation Procedures / Utilisation imprudente des procédures robustes d'estimation du spectre  
- 11:00 - 11:15 David Riegert, Aaron Springford, Wesley Burr & David J. Thomson (Queen's University)
Is There Structure in High-Frequency Variation of Great Lakes Water Levels? / Y a-t-il une structure dans la variation des hautes fréquences du niveau de l'eau dans les Grands Lacs?  

11:45 - 13:15

Lunch / Dîner

Creelman Hall

13:15 - 14:45 Session 3C-A (abstract/résumé - page 236) **Rozanski Hall, room/local 104**



Survey Methods Section Presidential Invited Address: Empirical Likelihood in Survey Sampling
Allocution de l'invité du Président du Groupe des méthodes d'enquêtes : Vraisemblance empirique et échantillonnage

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: David Haziza (Université de Montréal)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

13:15 - 14:45 Changbao Wu (University of Waterloo)
 Empirical Likelihood in Survey Sampling / La vraisemblance empirique dans les enquêtes par sondage  

13:15 - 14:45 Session 3C-I1 (abstracts/résumés - page 237) **Rozanski Hall, room/local 101**

Causality and Graphical Markov Models in High Dimensions
Causalité et modèles graphiques markoviens en plusieurs dimensions



Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Ayesha Ali (University of Guelph)

13:15 - 13:45 Frederick Eberhardt (Washington University, St. Louis), Antti Hyttinen & Patrik Hoyer (University of Helsinki, Finland)

Causal Discovery from Experiments / Découverte causale à partir d'expériences  

13:45 - 14:15 Helene Massam (York University), Gerard Letac (Université Paul Sabatier, France)
 Bayes Factors and the Geometry of Discrete Loglinear Models / Facteurs de Bayes et géométrie du modèle discret log-linéaire  

14:15 - 14:45 Erica Moodie (McGill University), Bibhas Chakraborty (Columbia University)
 Q-learning for Estimating Optimal Dynamic Treatment Rules from Observational Data / Q-learning pour l'estimation des règles optimales de traitement dynamiques de données d'observation  



13:15 - 14:45 Session 3C-I2 (abstracts/résumés - page 239) **Alexander Hall, room/local 200**



Copula Models
Modèles de copules



Invited Paper Session / Présentations sur invitation

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

Organizer/Responsable: Johanna Nešlehová (McGill University)

13:15 - 13:45 Louis-Paul Rivest (Université Laval)
 Applications of Exchangeable Copula Models / Quelques applications des modèles de copule échangeables  

13:45 - 14:15 Johanna Nešlehová & Christian Genest (McGill University), Bruno Rémillard (HEC Montréal)
 On the Multi-linear Empirical Copula Process / À propos du processus de copule empirique multilinéaire  

14:15 - 14:45 Bruno Rémillard (HEC Montréal)
 Testing Hypotheses for the Copula of Dynamic Models / Tests d'hypothèses pour la copule de modèles dynamiques  





13:15 - 14:45 Session 3C-I3 (abstracts/résumés - page 241) **Rozanski Hall, room/local 103**
Design and Analysis of Experiments in Modern Agriculture
Conception et analyse d'expériences en agriculture moderne

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Tom Loughin (Simon Fraser University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion





- 13:15 - 13:45 Chris Brien (University of South Australia), Bronwyn D. Harch & Ray L. Correll (CSIRO Mathematical and Information Sciences, Australia), Rosemary A. Bailey (Queen Mary University of London, U.K.)
 Principles in the Design of Multiphase Experiments with a Later Laboratory Phase: Orthogonal Designs / Principes généraux dans la conception de plans d'expérience à plusieurs phases comprenant une phase ultérieure en laboratoire : plans orthogonaux  
- 13:45 - 14:15 George Milliken (Kansas State University)
 Basic Designs of Experiments for Comparing Precision Agriculture Treatments on Research and Commercial Fields / Plans d'expérience simples pour comparer des traitements agricoles de précision dans des champs de recherche et commerciaux  
- 14:15 - 14:45 Ken McRae, Discussant (Agriculture and Agrifood Canada / Agriculture et Agroalimentaire Canada)

13:15 - 14:45 Session 3C-I4 (abstracts/résumés - page 243) **Alexander Hall, room/local 100**
Robustness for Complex Data
Robustesse et données complexes

Invited Paper Session / Présentations sur invitation

Chair/Président: Sanjoy Sinha (Carleton University)







Organizer/Responsable: Joanna Mills Flemming (Dalhousie University)





- 13:15 - 13:45 Chris Field (Dalhousie University)
 Robustness Issues in Molecular Evolution / Questions de robustesse en évolution moléculaire  
- 13:45 - 14:15 Matías Salibián-Barrera (University of British Columbia), Graciela Boente (Universidad de Buenos Aires, Argentina)
 On Robust Functional Principal Components / Composantes principales robustes pour données fonctionnelles  
- 14:15 - 14:45 Sanjoy Sinha, Discussant (Carleton University)

13:15 - 14:30 Session 3C-C1 (abstracts/résumés - page 245) **Rozanski Hall, room/local 102**
Financial Applications and Heavy-Tailed Distributions
Applications en finance et distributions à ailes relevées

Contributed Paper Session / Séance de communications libres

Chair/Président: François Watier (Université du Québec à Montréal)

- 13:15 - 13:30 Vincenzo Coia & Mei Ling Huang (Brock University)
 On Estimation of Heavy Tailed Distributions / De l'estimation des distributions à ailes relevées  
- 13:30 - 13:45 David Edward Wilson (University of Waterloo)
 Estimating the Meixner Process / Estimation du processus de Meixner  
- 13:45 - 14:00 Zhaoxia (Jeana) Ren & Adam Kolkiewicz (University of Waterloo)
 Estimation of the Jump Frequency in Jump-Diffusion Models / Estimation de la fréquence de chocs dans les modèles de distribution de chocs  

- 14:00 - 14:15 Nader Nematollahi (Allameh Tabataba'i University, Iran and University of Manitoba / Allameh Tabataba'i University, Iran et University of Manitoba), Azadeh Kiapour (Islamic Azad University, Iran)
Estimation of Premium and Prediction of Claim Size under Scale Invariant Squared Error Loss Function / Estimation des primes et prédiction de la taille des réclamations sous une fonction de perte quadratique invariante sous un changement d'échelle  
- 14:15 - 14:30 Franck Adekambi (University of the Witwatersrand, South Africa)
Ruin Probability in Health Care Insurance Pricing with Constant Force of Interest / Probabilité de ruine dans la fixation des prix d'assurance avec intensité d'intérêt constante  









13:15 - 14:15 Session 3C-C2 (abstracts/résumés - page 248) **MacNaughton Building, room/local 113**

Functional Data, Principal Components and Clustering

Données fonctionnelles, composantes principales et méthodes de partitionnement

Contributed Paper Session / Séance de communications libres

Chair/Président: Giles Hooker (Cornell University)

- 13:15 - 13:30 James Ramsay (McGill University)
Multivariate and Functional Principal Components without Eigenanalysis / Composantes principales multivariées et fonctionnelles sans analyse de valeurs propres  
- 13:30 - 13:45 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.  
- 13:45 - 14:00 Yunqi Ji & Antonio Ciampi (McGill University)
Model Based Clustering of Longitudinal Data: Application to Modeling Disease Course / Classification basée sur un modèle pour données longitudinales : application à la modélisation de l'évolution d'une maladie  
- 14:00 - 14:15 Sanjeena Dang & Paul McNicholas (University of Guelph)
Variational Approximations for Flexible Mixture Model-Based Clustering / Approximations variationnelles pour classification par mélange de modèles  







13:15 - 14:45 Session 3C-C3 (abstracts/résumés - page 250) **MacNaughton Building, room/local 105**







Methods for Genetic Association

Méthodes d'association génétique

Contributed Paper Session / Séance de communications libres

Chair/Président: Angelo Canty (McMaster University)

- 13:15 - 13:30 Antonio Ciampi (McGill University), Kathleen Oros Klein, J. Brent Richards & Celia Greenwood (Lady Davis Institute for Medical Research, Jewish General Hospital)
The Role of Statistical Learning Methods in Genetic Association Studies / Rôle des méthodes d'apprentissage statistique dans les études d'association génétique  
- 13:30 - 13:45 Binod Neupane, Jemila Hamid & Joseph Beyene (McMaster University)
Multivariate Meta-Analysis in Genetic Association Studies / Méta-analyse multivariée dans les études d'association génétique  
- 13:45 - 14:00 Julia Taleban, Laura L. Faye, Lei Sun, Apostolos Dimitromanolakis & Shelley Bull (Samuel Lunenfeld Research Institute)
Bootstrap Bias-reduction in Genetic Association Analysis of Time-to-event Outcomes / Réduction du biais par bootstrap dans une analyse d'association génétique du temps avant événement  

- 14:00 - 14:15 Andriy Derkach & Lei Sun (University of Toronto), Jerry Lawless (University of Waterloo)
Combining Linear and Quadratic Tests for Rare Variants Provides a Robust Test Across Genetic Models / Combiner des tests linéaires et quadratiques pour des variantes rares produit un test robuste pour différents modèles génétiques  
- 14:15 - 14:30 Zhijian Chen (Samuel Lunenfeld Research Institute), Radu V. Craiu (University of Toronto), Shelley Bull (Samuel Lunenfeld Research Institute)
Sequential Two-Phase Stratified Designs for Regional Sequencing Following Genome-Wide Association Study: A Bayesian Approach / Plans stratifiés séquentiels à deux phases pour séquençage régional suite à une étude d'association pangénomique : une approche bayésienne  
- 14:30 - 14:45 Zeynep Baskurt (University of Toronto), Lisa J. Strug (The Hospital For Sick Children)
A Composite Likelihood Approach Using the Evidential Paradigm to Analyze Genetic Association in Pedigrees / Approche de vraisemblance composite par le paradigme de preuves pour analyser l'association génétique dans les lignées  

Abstracts • Résumés

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

Organizer and Chair / Responsable et président: John Brewster (University of Manitoba)

Rozanski Hall, room/local 104

Abstract/Résumé

JAMES O. BERGER (Duke University)

[Monday June 4 / lundi 4 juin, 8:30]

Reproducibility and Multiplicity

Reproductibilité et multiplicité

An increasing fraction of scientific findings seem not to be reproducible. For instance, Bayer Healthcare examined 67 published studies but was able to replicate less than 1/4 of the findings, with over 2/3 exhibiting major inconsistencies. After discussing the evidence for a growing lack of reproducibility, suspected causes of the problem will be reviewed – at least those causes that relate to statistics. This increase in irreproducible science coincides with the increasing data deluge, suggesting that a major cause is incorrect handling of multiplicity. Some promising Bayesian directions towards resolution of the multiplicity problem are discussed.

Une proportion croissante des résultats scientifiques ne semblent pas reproductibles. Par exemple, après avoir examiné 67 études publiées, Bayer Healthcare n'a pu reproduire seulement que le quart des résultats, parmi lesquels plus des deux tiers comportaient des incohérences majeures. Nous discutons d'abord de ce phénomène grandissant de non-reproductibilité, puis nous examinons les causes présumées du problème qui sont liées à la statistique. Cette hausse de la non-reproductibilité des résultats scientifiques coïncide avec la multiplication des données, suggérant que le traitement incorrect de la multiplicité serait une cause majeure du problème. Nous présentons quelques orientations bayésiennes prometteuses visant à résoudre cette question.

1B-V: Nonparametric Bayes Methods: an Overview Statistique bayésienne nonparamétrique: un survol

Organizer and Chair / Responsable et président: Alexandre Leblanc (University of Manitoba)

Rozanski Hall, room/local 104

Abstract/Résumé

STEVE MACEACHERN (Ohio State University)

[Monday June 4 / lundi 4 juin, 10:15]

Nonparametric Bayes Methods: An Overview

Méthodes bayésiennes non-paramétriques : une vue d'ensemble

Nonparametric Bayesian methods provide a powerful set of tools that allow us to build realistic, complex models for many phenomena. The methods allow us to inject information into an analysis without rigidly enforcing unrealistic constraints, thereby alleviating one of the main objections to the use of Bayesian methods. The methods have also proven effective as a means of flexibly fitting data under low-information specification of the prior distribution. This talk, aimed at the non-expert, will survey some of the major nonparametric Bayesian models, providing motivation, a glimpse into details, modelling techniques, and illustration of performance in a number of examples.

Les méthodes bayésiennes non-paramétriques fournissent un ensemble d'outils efficaces permettant de construire des modèles réalistes et complexes pour divers phénomènes. Ces techniques permettent d'intégrer de l'information à une analyse sans y imposer de contraintes irréalistes de façon stricte, atténuant ainsi l'une des principales critiques quant à l'utilisation des méthodes bayésiennes. Celles-ci se sont également montrées efficaces comme moyen flexible pour ajuster des données lorsque peu d'information est disponible pour spécifier la distribution a priori. Cet exposé, qui s'adresse aux non-experts, étudie quelques-uns des principaux modèles bayésiens non-paramétriques, en offrant une justification de leur utilisation, un aperçu des détails, des techniques de modélisation, et une illustration de leur performance au moyen d'une série d'exemples.

1B-I1: Extending Mixtures of Factor Analyzers Généralisations des mélanges avec covariances d'analyse factorielle

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

Alexander Hall, room/local 200

Abstracts/Résumés

JEFFREY ANDREWS & PAUL MCNICHOLAS (University of Guelph)

[Monday June 4 / lundi 4 juin, 10:15]

Mixtures of Modified t -Factor Analyzers

Mélanges de lois de Student modifiées avec covariance d'analyse factorielle

A family of 24 modified t -factor analyzer models, known as the MOD t family, is introduced. This family is discussed in the context of model-based clustering and classification, and applied to several real data sets. Comparisons with its Gaussian counterpart, as well as other popular mixture model families, are made. Finally, potential directions of related future research are discussed.

Nous présentons une famille de 24 lois de Student modifiées avec une covariance d'analyse factorielle, connue sous le nom de famille modifiée MOD t . Nous discutons de cette famille dans le cadre d'analyses de regroupement et de classification, et l'appliquons à divers jeux de données réelles. Nous la comparons avec sa contrepartie Gaussienne, ainsi qu'avec d'autres modèles de mélanges populaires. Finalement, nous discutons d'orientations potentielles pour la recherche future.

JANGSUN BAEK (Chonnam National University, South Korea)

[Monday June 4 / lundi 4 juin, 10:45]

Applications of Various Mixtures of Common Factor Analyzers to Clustering of High-Dimensional Data

Application de divers mélanges à covariance d'analyse factorielle pour l'analyse de regroupement de données de grande dimension

Mixtures of common factor analyzers are a useful technique for clustering high-dimensional data where the number of observations is small relative to the number of variables. It implements mixtures of factor analyzers with common component-factor loadings depending on various types of distribution of component-error and factor: multivariate normal, multivariate t , and multivariate lognormal distributions. It can handle data with not only normal tails, but also tails longer than that of the normal distribution or skewed tails which are commonly observed in microarray experiments. Furthermore, it produces robust clustering results against outliers, and displays the data in a low-dimensional factor space.

Les mélanges avec covariance d'analyse factorielle sont une bonne technique d'analyse de regroupement pour les données de grande dimension où le nombre d'observations est petit par rapport au nombre de variables. Cette technique met en place des mélanges à covariance d'analyse factorielle avec des saturations communes selon le type de distribution pour le facteur et l'erreur de chaque composante : distribution normale multivariée, distribution de Student multivariée, distribution log-normale multivariée. Elle peut traiter des données avec des queues normales, mais aussi avec des queues asymétriques, ou plus relevées que la normale, ce qui est généralement observé dans les expériences sur micro-réseaux. Elle produit également des analyses de regroupement robustes aux valeurs aberrantes et présente les données dans un espace factoriel de petite dimension.

1B-I1: Extending Mixtures of Factor Analyzers Généralisations des mélanges avec covariances d'analyse factorielle

SALVATORE INGRASSIA (Università di Catania, Italy), **PAUL MCNICHOLAS** (University of Guelph), **ANTONIO PUNZO** (Università di Catania, Italy), **SANJEENA SUBEDI** (University of Guelph)

[Monday June 4 / lundi 4 juin, 11:15]

Cluster Weighted Factor Analyzers

Lois à covariance d'analyse factorielle pour modèles de grappes pondérées

Cluster-Weighted Modelling (CWM) is a flexible statistical framework for modelling local relationships in heterogeneous populations on the basis of weighted combinations of local models. In particular, CWM models the joint density $p(\mathbf{x}, y)$ of a random vector \mathbf{X} and a random variable Y which can be considered as the input and the output variables, respectively. It can be proven that, under suitable assumptions, CWM includes finite mixtures of regressions as special cases. We extend Cluster Weighted Models in order to include an underlying latent factor structure for the input variables. This allows modelling dataset with large input variables.

La modélisation par pondération de grappes (MPG) est une structure statistique flexible qui établit des relations locales entre des populations hétérogènes par la combinaison pondérée de modèles locaux. Plus précisément, le MPG modélise la densité conjointe $p(\mathbf{x}, y)$ d'un vecteur aléatoire de variables explicatives \mathbf{X} et d'une variable réponse Y . Nous montrons que sous des hypothèses convenables, les MPG incluent les mélanges finis de régressions. Nous étendons par ailleurs les MPG de façon à inclure une structure latente factorielle sous-jacente pour les variables explicatives. Cela nous permet, en particulier, de modéliser des données lorsque les variables explicatives sont nombreuses.

1B-I2: Responsive Designs Conception adaptée de plans d'échantillonnage

Chair/Président: David Haziza (Université de Montréal)

Organizer/Responsable: Wesley Yung (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Rozanski Hall, room/local 103

Abstracts/Résumés

PETE BRODIE (Office for National Statistics, U.K.)

[Monday June 4 / lundi 4 juin, 10:15]

Developing a Response Chasing Strategy for Business Surveys within the Office for National Statistics

Développement d'une stratégie de rappel pour les enquêtes auprès des entreprises à l'Office for National Statistics

The Office for National Statistics like other National Statistical Offices employs very simple processes for ensuring that data is returned for our many business surveys. These traditional processes rely on prescribed targets for numbers of returned questionnaires often coupled with targets based on auxiliary information available from the sampling frame (employment size, etc.) so as to ensure that "important" responders provide their data. Here we explore some alternative methods for maximizing the quality of the resultant outputs. These methods use the basic maxim that the responders that will best improve quality should be prioritised within the response chasing strategy.

Comme d'autres organismes statistiques nationaux, l'Office for National Statistics utilise des processus très simples pour s'assurer que les données de ses nombreuses enquêtes auprès des entreprises sont récupérées. Ces processus traditionnels s'appuient sur des cibles définies pour le nombre de questionnaires retournés souvent jumelées à des cibles basées sur la disponibilité de l'information auxiliaire provenant de la base de sondage (nombre d'employés, etc.), de façon à s'assurer que les répondants « importants » fournissent leurs renseignements. Nous explorons quelques méthodes alternatives pour maximiser la qualité des résultats. Ces méthodes recourent à la maxime de base affirmant qu'une priorité devrait être accordée aux répondants les plus susceptibles d'améliorer la qualité dans le cadre de la stratégie de rappel.

DANIEL TOTH & POLLY PHIPPS (U.S. Bureau of Labor Statistics), **NICK HORTON** (Smith College)

[Monday June 4 / lundi 4 juin, 10:45]

Analyzing Establishment Nonresponse Using an Interpretable Regression Tree Model with Linked Administrative Data

Analyse de non-réponse d'établissements à l'aide d'un arbre de régression interprétable avec des données administratives liées

To gain insight into how characteristics of an establishment are associated with nonresponse, we applied a recursive partitioning algorithm to the Occupational Employment Statistics May 2006 survey data. The resulting regression tree models an establishment's propensity to respond to the survey given certain establishment characteristics. We

Pour comprendre la relation entre les caractéristiques d'un établissement et la non-réponse, nous avons appliqué un algorithme de partitionnement récursif aux données de l'enquête Occupational Employment Statistics de mai 2006. Les arbres de régression résultants modélisent la probabilité de réponse à l'enquête d'un établissement selon certaines de ses caractéristiques. Nous testons le modèle avec les don-

1B-I2: Responsive Designs Conception adaptée de plans d'échantillonnage

test the model using data from the November 2006 Occupational Employment Statistics survey. This representation is then used along with frame-level administrative wage data linked to sample data to illustrate the potential for non-ignorable nonresponse bias. We then propose a method to adjust for nonresponse using the administrative data to correct for this bias.

nées de novembre 2006 pour la même enquête. Cette représentation est ensuite utilisée avec des données administratives sur les salaires au niveau de la base de sondage liées aux données de l'échantillon pour illustrer le potentiel de biais de non-réponse non-ignorable. Nous proposons ensuite une méthode pour corriger ce biais en ajustant pour la non-réponse à l'aide des données administratives.

1B-I3: Statistical Inference for Differential Equation Models
Inférence statistique pour les modèles d'équations différentielles

Organizer and Chair / Responsable et président: Jiguo Cao (Simon Fraser University)

Alexander Hall, room/local 100

Abstracts/Résumés

MICHAEL DOWD (Dalhousie University)

[Monday June 4 / lundi 4 juin, 10:15]

Data Assimilation for Large-Scale Spatio-Temporal Dynamic Models

Assimilation de données pour modèles dynamiques spatio-temporels à grande échelle

Data assimilation estimates the system state and parameters for dynamic models. These numerical models solve nonlinear time-space dependent partial differential equations. Major application areas are in oceanography and meteorology. I present recent progress in developing sequential Monte Carlo approaches for these large scale systems using two studies: (i) Sequential importance resampling to determine the biogeochemical state and parameters using a 3D ocean circulation model and satellite imagery; and (ii) A new approach that uses particle filtering/smoothing in space, as well as in time, to allow for sequential Monte Carlo methods to be effectively applied to large scale stochastic dynamic systems.

L'assimilation de données est une méthode estimant l'état et les paramètres du système pour les modèles dynamiques. Les modèles numériques employés permettent de résoudre des équations aux dérivées partielles non linéaires dépendant de l'espace et du temps. Les principaux champs d'application sont l'océanographie et la météorologie. Je présente les récents progrès dans l'élaboration de méthodes de Monte-Carlo séquentielles pour ces systèmes à grande échelle au moyen de deux études : (i) le rééchantillonnage d'importance séquentiel qui permet de déterminer l'état et les paramètres biogéochimiques à l'aide d'imagerie satellite et d'un modèle de la circulation océanique en 3D ; et (ii) une nouvelle approche qui utilise la filtration et le lissage de particules dans l'espace et dans le temps afin de pouvoir appliquer efficacement les méthodes de Monte Carlo séquentielles à des systèmes dynamiques stochastiques.

GILES HOOKER & LIEFUR THORBERGSSON (Cornell University), **KEVIN LIN** (Arizona State University), **BRUCE ROGERS** (SAMSU)

[Monday June 4 / lundi 4 juin, 10:45]

Control Theory and the Design of Dynamic Experiments

Théorie de la régulation et conception d'expériences dynamiques

We consider experimental designs in which a dynamical system modelled with time-varying inputs can be externally controlled. Experimental protocols typically fix system inputs at constant values or use predefined functional forms for input values. In this talk, we examine the problem of designing experimental protocols that can adapt to observations in order to optimize the precision of parameter estimates. This can be particularly important when certain parameters affect system behaviour only in particular regions of the state space. This prob-

Nous considérons des plans d'expérience dans lesquels un système dynamique modélisé au moyen d'entrées variant avec le temps peut être régulés à l'externe. Les protocoles d'expérimentation fixent habituellement les entrées à des valeurs constantes, ou utilisent des formes fonctionnelles prédéfinies pour l'entrée de valeurs. Dans cet exposé, nous examinons le problème de conception de protocoles expérimentaux qui peuvent s'adapter aux observations afin d'optimiser la précision des estimateurs des paramètres. Ceci peut être particulièrement important lorsque certains paramètres ont des répercussions sur le comportement du

1B-I3: Statistical Inference for Differential Equation Models Inférence statistique pour les modèles d'équations différentielles

lem can be expressed within the context of control theory and we explore the properties of maximally informative designs.

système seulement dans des régions précises de l'espace d'états. Ce problème peut être exprimé dans le cadre de la théorie de la régulation. Nous explorons ensuite les propriétés des plans les plus informatifs.

AARON A. KING (University of Michigan)

[Monday June 4 / lundi 4 juin, 11:15]

POMP: Statistical Inference for Stochastic Dynamical Systems

Processus markovien partiellement observable : l'inférence statistique pour les systèmes stochastiques dynamiques

Partially-observed Markov models can be used to confront scientific hypotheses with time series data. The computational complexity of the inference problem for such models and the need for modeling flexibility impose restrictions on the class of inference algorithms of practical use. *Plug-and-play* methods are attractive insofar as multiple hypotheses are to be entertained and computing power is plentiful. The R package *pomp* is a modeling and inference tool and a platform for plug-and-play inference algorithms. This talk will show off the package's modeling and inference facilities and extend an invitation for new contributions to the project.

Les modèles markoviens partiellement observés peuvent être utilisés pour confronter des hypothèses scientifiques avec les données d'une série chronologique. La complexité de calcul du problème d'inférence pour ces modèles et la nécessité de flexibilité dans la modélisation imposent des restrictions sur la classe des algorithmes d'inférence utilisables en pratique. Les méthodes *prêtes à l'utilisation* sont attrayantes étant donné que de nombreuses hypothèses sont à vérifier et que la puissance de calcul est grande. Le progiciel *pomp* pour R est un outil de modélisation et d'inférence et une plate-forme pour des algorithmes prêts à l'utilisation. Cet exposé présentera les capacités de modélisation et d'inférence du progiciel et invitera à apporter de nouvelles contributions à ce projet.

1B-I4: Uncertainty Quantification in Analysis of Computer Models

Quantifier l'incertitude dans l'analyse des modèles informatiques

Organizer and Chair / Responsable et président: William Welch (University of British Columbia)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Rozanski Hall, room/local 101

Abstracts/Résumés

DORIN DRIGNEI (Oakland University)

[Monday June 4 / lundi 4 juin, 10:15]

Sensitivity Analysis of a Dynamical Computer Model for Vehicle Suspension Systems

Analyse de sensibilité d'un modèle informatique dynamique pour les systèmes de suspension des véhicules

The study of natural phenomena is aided by computer (or simulation) models. In order to understand the influence of inputs on outputs in computer models, formal sensitivity analyses are usually carried out. In this talk we present methods for sensitivity analysis in computer models with time series output. In addition, the sensitivity indices discussed will be used to screen out inactive parameters in computer model calibration. The methods will be illustrated with a computer model for the simulation of vehicle suspension systems.

Les modèles informatiques (ou par simulations) aident à étudier les phénomènes naturels. Pour comprendre l'influence des entrées sur les sorties des modèles informatiques, des analyses formelles de sensibilité sont habituellement effectuées. Dans cet exposé, nous présentons des méthodes d'analyse de sensibilité dans des modèles informatiques dont les sorties sont des séries chronologiques. De plus, les indices de sensibilité traités seront utilisés pour filtrer les paramètres inactifs lors de la calibration de modèle informatique. Les méthodes sont illustrées au moyen d'un modèle informatique par la simulation de systèmes de suspension de véhicules.

JOSLIN GOH & DEREK BINGHAM (Simon Fraser University), **JAMES PAUL HOLLOWAY, CAROLYN C. KURANZ, MICHAEL J. GROSSKOPF, ERICA RUTTER & FORREST W. DOSS** (University of Michigan)

[Monday June 4 / lundi 4 juin, 10:45]

Tuning with Multiple Emulators of Different Fidelities

Ajustement au moyen de plusieurs émulateurs de niveau de fidélité différent

Computer codes are used widely to describe physical processes in lieu of physical observations. In some cases, more than one emulator, each with different degrees of fidelity, can be used to explore the physical system. In this work, we combine field observations and model runs from deterministic multi-fidelity emulators to build a predictive model for the real process. The resulting model can be used to perform sensitivity analysis, solve inverse problems and make predictions. The approach is Bayesian and will be illustrated through a simple example, followed by a real application at

Les programmes informatiques sont fréquemment utilisés à la place d'observations physiques pour décrire des processus physiques. Dans certains cas, plusieurs émulateurs possédant chacun un niveau de fidélité différent peuvent être utilisés pour explorer le système physique. Ce travail combine des observations sur le terrain et des simulations d'émulateurs déterministes à niveaux de fidélité multiples dans le but de construire un modèle de prévision pour le véritable processus. Le modèle qui en résulte permet d'effectuer une analyse de sensibilité, de résoudre les problèmes inverses et de faire des prévisions. Il s'agit d'une approche bayésienne que nous illustrons à l'aide d'un exemple simple,

1B-I4: Uncertainty Quantification in Analysis of Computer Models Quantifier l'incertitude dans l'analyse des modèles informatiques

the Center for Radiative Shock Hydrodynamics.

suivi d'une application réelle au Center for Radiative Shock Hydrodynamics.

JASON LOEPPKY (University of British Columbia, Okanagan), **WILLIAM WELCH** (University of British Columbia),
DEREK BINGHAM (Simon Fraser University)

[Monday June 4 / lundi 4 juin, 11:15]

Quantification of Computer Model Bias

Quantification du biais d'un modèle informatique

Computer models to simulate physical phenomena are widely available in engineering and science. Before relying on a computer model, a natural first step is to compare its output with physical data to assess whether the model reliably represents reality. Physical data can also be used to calibrate or tune unknown parameters in the computer model. Calibration is particularly problematic in the presence of systematic discrepancies between the computer model and physical observations. In this talk we investigate and illustrate some of the many trade-offs that can occur between the discrepancies function, the estimated computer model and the error process.

La simulation de phénomènes physiques au moyen de modèles informatiques est très répandue en science et génie. Avant de se fier à un modèle informatique, la première étape habituelle consiste à comparer ses résultats à des données physiques pour juger de la fiabilité du modèle. Les données physiques peuvent également servir à calibrer ou ajuster les paramètres inconnus d'un modèle informatique. La calibration est particulièrement problématique en présence de divergences systématiques entre le modèle informatique et les observations physiques. Dans cet exposé, nous étudions et illustrons quelques-uns des compromis possibles entre la fonction de divergence, le modèle informatique estimé et le processus d'erreur.

1B-C1: Biostatistics: Modelling and Study Design
Biostatistique: modélisation et planification d'étude

Chair/Présidente: Yanqing Yi (Memorial University of Newfoundland)

MacNaughton Building, room/local 105

Abstracts/Résumés

LEHANA THABANE (McMaster University), **JANUSZ KACZOROWSKI** (Université de Montréal), **LISA DOLOVICH** (McMaster University), **LARRY CHAMBERS** (University of Ottawa), **ON BEHALF OF CHAP INVESTIGATORS**

[Monday June 4 / lundi 4 juin, 10:15]

Dealing with Criticisms and Controversies of Pragmatic Trials

Traitement des critiques et des controverses entourant les essais pragmatiques

Pragmatic trials are designed to answer the practical question of whether offering an intervention compared with some alternative (eg. usual care) in routine health care does more good than harm. In this presentation, I will review the similarities and differences between pragmatic and explanatory trials; and common criticisms and controversies of pragmatic trials. I will use the Cardiovascular Health Awareness Program (CHAP) trial (see www.CHAPprogram.ca) which was a community randomization pragmatic trial designed to assess whether offering a highly organized, community-based CHAP intervention compared to usual care can reduce cardiovascular related outcomes to illustrate how we addressed some of the criticisms.

Les essais pragmatiques sont conçus pour répondre à la question pratique de savoir si la réalisation d'une intervention, par opposition à une méthode alternative (comme les soins habituels), est plus utile que dommageable dans le cadre des soins de santé courants. Dans cet exposé, nous examinerons les similitudes et les différences entre les essais pragmatiques et explicatifs, ainsi que les critiques et controverses communes en matière d'essais pragmatiques. Pour illustrer notre traitement de certaines critiques, nous utiliserons l'essai du Programme de sensibilisation à la santé cardiovasculaire (PSSC, voir www.CHAPprogram.ca), qui consiste en un essai pragmatique de randomisation du milieu conçu pour évaluer si la pratique d'une intervention PSSC hautement structurée et fondée sur le milieu peut réduire les risques de problèmes liés au système cardiovasculaire, comparativement aux soins habituels.

SHIYUAN (WILLIAM) ZHANG & LEHANA THABANE (McMaster University)

[Monday June 4 / lundi 4 juin, 10:30]

Critical Appraisal and Application of Different Statistical Methods of Analyzing Continuous Outcomes in RCTs

Appréciation critique et application de différentes méthodes d'analyse de données continues d'ECR

In the analysis of continuous outcomes in Randomized Controlled Trials (RCTs), standard statistical methods are commonly employed to evaluate the treatment effect by the comparison of groups using: i) the post-treatment scores, ii) change in scores, iii) percentage change from baseline and iv) post-treatment scores adjusting for baseline as a covariate. In this talk, we present a critical appraisal of methods and application to MOBILE trial designed to assess whether the addition of gabapentin to a multimodal-perioperative analge-

Dans l'analyse de données continues provenant d'essais contrôlés randomisés (ECR), des méthodes statistiques standards sont couramment employées pour évaluer l'effet du traitement par la comparaison des groupes au moyen : i) des scores post-traitement, ii) de la variation dans les scores, iii) de la variation en pourcentage par rapport à la référence et iv) des scores post-traitement utilisant la référence comme covariable. Dans cet exposé, nous présentons une appréciation critique des méthodes et de l'application de l'essai MOBILE conçu pour évaluer si l'ajout de gabapentine à un traitement analgésique périopératoire multimo-

1B-C1: Biostatistics: Modelling and Study Design Biostatistique: modélisation et planification d'étude

sia regimen can reduce morphine consumption or improve analgesia following total joint arthroplasty. We will use range of motion and pain scores to illustrate the applications.

dal peut réduire la consommation de morphine ou améliorer l'analgésie à la suite d'une chirurgie d'arthroplastie totale. Nous illustrons les applications à l'aide des mesures d'amplitude de mouvement et de douleur.

MEAGHAN S. CUERDEN, CECILIA COTTON & RICHARD J. COOK (University of Waterloo)

[Monday June 4 / lundi 4 juin, 10:45]

Directed Donor Transfusions in Infants Admitted to the Neonatal Intensive Care Unit

Transfusions par don dirigé pour nourrissons admis à l'unité de soins intensifs néonataux

Parents of infants in the neonatal intensive care unit often donate blood through a directed donor program designed to minimize risk of exposure for their infants. Controversy persists regarding the effect of directed donation on adverse health outcomes. Challenges in assessing the impact of directed donation include data collection issues, physician and parent preferences, parent eligibility, scheduling of transfusions, etc. We discuss challenges in causal inferences and describe a simulation study of the need for transfusions, directed donor status, and infant outcomes. Data collected prospectively from 592 infants admitted to Hamilton Health Sciences acute care hospitals are analyzed for illustration.

Les parents de nourrissons à l'unité de soins intensifs néonataux donnent souvent du sang via un programme de dons dirigé dans le but de minimiser le risque de contamination pour leurs enfants. L'effet de dons dirigés sur la santé est toutefois encore controversé. Les facteurs qui compliquent l'évaluation de l'impact des dons dirigés incluent des problèmes de collecte de données, les préférences du médecin et des parents, l'admissibilité des parents, l'horaire des transfusions, etc. Nous discutons des difficultés pour l'inférence causale et décrivons une étude de simulation concernant le besoin de transfusions, la situation du donneur direct et les résultats sur la santé de l'enfant. Les données d'une étude prospective de 592 enfants admis aux hôpitaux de soins aigus du Hamilton Health Sciences sont analysées pour illustrer nos résultats.

OLLI SAARELA, ERICA MOODIE & DAVID STEPHENS (McGill University)

[Monday June 4 / lundi 4 juin, 11:00]

On Bayesian Estimation of Marginal Structural Models

De L'estimation bayésienne des modèles structurels marginaux

Bayesian estimation of marginal causal contrasts requires a full probability model specification and integration over intermediate variables and confounders. This is avoided in marginal structural models, estimated using inverse probability of treatment (IPT) weighting, which motivates studying Bayesian counterparts of IPT weighted methods. We propose an approach based on posterior predictive distribution of the weighted estimator. We also outline how this approach can be utilized in testing modeling assumptions should one wish to model the outcome process. The methods are illustrated with simulations, with specific interest in variance estimation, and with data from the Canadian Co-infection Cohort study.

L'estimation bayésienne des contrastes causaux marginaux exige la spécification et l'intégration d'un modèle de probabilité sur des variables et facteurs confusionnels intermédiaires. Ces hypothèses ne sont pas nécessaires dans les modèles structurels marginaux, estimés par pondération par l'inverse de probabilité de traitement (IPT). Ceci conduit à étudier les homologues bayésiens des modèles pondérés IPT. Nous proposons une approche fondée sur la distribution prédictive a posteriori de l'estimateur pondéré, puis montrons comment l'utiliser pour tester les hypothèses de modélisation et pour modéliser le processus résultant. Nous illustrons ces méthodes par des simulations spécifiquement pour estimer la variance, et sur des données de la Cohorte canadienne de co-infection.

1B-C1: Biostatistics: Modelling and Study Design Biostatistique: modélisation et planification d'étude

RAMON CLÈRIES & JOSEPA RIBES (Cancer Registry of Catalonia and University of Barcelona, Spain / Cancer Registry of Catalonia et University of Barcelona, Spain), **JOSE MIGUEL MARTINEZ** (CISAL / GREDS-EMCONET - University Pompeu Fabra, Spain), **YUTAKA YASUI** (University of Alberta)

[Monday June 4 / lundi 4 juin, 11:15]

Decomposing Changes in Disease Incidence and Mortality Counts into Overall and Age-Specific Changes in Risk and Population Size

Décomposition des changements de l'incidence d'une maladie et de la mortalité en changements globaux et spécifiques à l'âge du risque et de la taille de la population

Understanding differences in disease incidence and mortality counts between time points or geographical areas is useful in the assessment and development of public health policies. We consider a decomposition of these differences into overall and age-specific differences in “risk” and “population size” using an Age-Period model. The method allows describing an overall period effect on “risk” and “population size”, and differential period effects across age groups. The utility of the method will be illustrated using cancer incidence data from two Spanish regions and mortality data from the World Health Organisation database. Graphical presentations for stakeholders are also discussed.

Une compréhension des différences temporelles et géographiques de l'incidence de la maladie et de la mortalité est utile pour l'évaluation et le développement de politiques de santé publiques. Nous décomposons ces différences en différences de « risque » et de « taille de population » globales et spécifiques à l'âge par un modèle Âge - Période permettant de décrire un effet de période global sur le risque et la taille de population et des effets de périodes différenciés par groupes d'âge. Nous en illustrons l'efficacité par des données sur l'incidence du cancer provenant de deux régions espagnoles et des données de mortalité de la base de données de l'Organisation Mondiale de la Santé. Nous discutons de la présentation graphique faite aux intervenants.

QING GUO, LEHANA THABANE, GEOFFREY HALL, RON GOEREE, MARGARET MCKINNON & ELEANOR M. PULLENAYEGUM (McMaster University)

[Monday June 4 / lundi 4 juin, 11:30]

The Reporting of Sample Size Calculations in Functional Magnetic Resonance Imaging Studies: A Systematic Review

Publication des calculs de la taille d'échantillon dans les études d'imagerie par résonance magnétique fonctionnelle : un examen systématique

Anecdotal evidence suggests that functional MRI (fMRI) studies rarely consider power when setting sample size. This raises concerns, since under-sized studies may fail to detect effects of interest, and encourage data dredging. Although sample size methodology in fMRI studies exists, implementation may be hampered by uncertainty of input parameters in the calculations. We systematically evaluated how often sample size calculations were reported in observational fMRI studies. Moreover, we assessed the proportion of articles that reported estimates of effect size and variance components in their results sections, as these are needed to facilitate sample size determination for future studies.

Selon des preuves anecdotiques, les études par imagerie par résonance magnétique fonctionnelles (IRMf) considèrent rarement la puissance lorsqu'elles déterminent la taille d'un échantillon. Ceci est préoccupant dans la mesure où des études de taille insuffisante risquent de ne pas détecter les effets escomptés, favorisant la torture des données. Bien que la méthodologie de calcul de la taille d'échantillon existe pour les études par imagerie par résonance magnétique fonctionnelle, la mise en oeuvre de celle-ci peut être restreinte par l'incertitude liée aux paramètres d'entrée dans les calculs. Nous évaluons systématiquement la fréquence à laquelle les calculs de tailles d'échantillon ont été fournis dans les études d'observation par IRMf. De plus, nous avons déterminé la proportion des articles qui ont communiqué des estimés de l'ampleur de l'effet et des composantes de la variance dans leurs sections de résultats, étant donné

1B-C1: Biostatistics: Modelling and Study Design
Biostatistique: modélisation et planification d'étude

qu'elles sont nécessaires pour déterminer plus facilement la taille des échantillons dans les études futures.

1B-C2: Frailty Models and the Analysis of Recurrent Events Modèles de fragilité et analyse d'événements récurrents

Chair/Présidente: Katherine Davies (University of Manitoba)

MacNaughton Building, room/local 113

Abstracts/Résumés

AVIK HALDER & GLEN TAKAHARA (Queen's University)

[Monday June 4 / lundi 4 juin, 10:15]

Bayesian Semiparametric Inference of Frailty Model Using Lévy Process Priors

Inférence bayésienne semi-paramétrique d'un modèle de la fragilité utilisant un processus de Lévy comme loi a priori

Lee and Kim (2003) first proposed the use of a Levy process prior for Bayesian inference of the cumulative hazard function in a proportional hazard model. We extend this model by incorporating a frailty component for heterogeneity among the cumulative intensity processes for different subjects. The Lévy process prior generalizes that used by Sinha (1993) and Sinha et al.(1998), who do consider frailty models. It is assumed that random effects are Gamma with mean one and variance θ which makes our model semi-parametric. We characterize the joint posterior and use blocked Gibbs sampling for sampling from the joint posterior.

Lee et Kim (2003) ont été les premiers à proposer l'utilisation du processus de Lévy comme loi a priori pour l'inférence bayésienne de la fonction de risque cumulé dans un modèle de régression à risque proportionnel. Nous généralisons ce modèle en incorporant une composante de fragilité pour l'hétérogénéité des processus d'intensité cumulative des différents sujets. Utiliser le processus de Lévy comme loi a priori généralise les travaux de Sinha (1993) et Sinha et coll. (1998), qui considèrent les modèles de fragilité. On suppose que les effets aléatoires sont de loi Gamma avec moyenne un et une variance θ , ce qui rend notre modèle semi-paramétrique. Nous décrivons la loi a posteriori conjointe et utilisons l'échantillonnage de Gibbs par bloc pour générer des pseudo-observations de la loi conjointe a posteriori.

CHRISTIAN BOUDREAU (University of Waterloo), **HUGO VARET & SANDRINE KATSAHIAN** (Hôpital Henri Mondor)

[Monday June 4 / lundi 4 juin, 10:30]

Multivariate Frailty Models for Competing Risks

Modèles de fragilité multivariés pour risques concurrents

In this talk we extend Fine and Gray's model for the subdistribution of a competing risk to include multivariate frailties that are assumed to have a joint log-normal distribution. This extension allows for the modelling of situations where a more complex frailty structure is needed than the usual shared frailty. For example, when multiple frailty terms occur within each cluster, or when frailties are nested. Estimation is done using a penalized partial likelihood approach in combination with a Laplace approximation of the partial likelihood. The proposed method is then illustrated using bone-marrow transplantation data for acute myeloid leukemia.

Dans cet exposé, nous généralisons le modèle de Fine et Gray pour la sous-distribution de risques concurrents pour inclure des modèles de fragilité multivariée que nous supposons avoir une distribution conjointe log-normale. Cette extension permet de modéliser des situations pour lesquels une structure de fragilité plus complexe que celle des modèles usuels à fragilité partagée est requise. Par exemple, lorsqu'il y a plusieurs termes de fragilité dans un même groupe, ou lorsque les fragilités sont emboîtées. L'estimation est effectuée à l'aide d'une approche de vraisemblance partielle pénalisée en conjonction avec une approximation de Laplace de la vraisemblance partielle. La méthode proposée est illustrée à l'aide de données sur la transplantation de moelle osseuse pour la leucémie myéloïde aiguë.

1B-C2: Frailty Models and the Analysis of Recurrent Events Modèles de fragilité et analyse d'événements récurrents

BALAKUMAR SWAMINATHAN & YUN-HEE CHOI (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 10:45]

Ascertainment Corrected Penetrance Estimate for Bivariate Event Time Outcomes using Frailty Model
Estimation de la pénétrance corrigée par constatation à l'aide d'un modèle de fragilité

In the analysis of complex diseases, affecteds may experience successive events as a consequence of a primary event. We aim to quantify the dependence between sequential event times and estimate the age dependent penetrance. Shared gamma frailty model with Weibull baseline is used and adjusted for family based study design. Using 12 Lynch Syndrome families from Newfoundland, penetrance of first colorectal cancer by age 70 was 99% for male mutated gene carriers and 86% for female carriers. Penetrance to develop a second colorectal cancer in 5 years following first event was 20% for male carriers and 11% for female carriers.

Dans les maladies complexes, les patients peuvent considérer des événements successifs comme conséquence d'un événement initial. Nous tentons de quantifier la dépendance entre les temps séquentiels des événements et d'estimer la pénétrance dépendante de l'âge. Nous utilisons un modèle de fragilité gamma partagée avec ligne de base Weibull et l'ajustons au plan d'étude familial. Pour 12 familles de Terre-Neuve souffrant du syndrome de Lynch, la pénétrance d'un premier cancer colorectal à l'âge de 70 ans était de 99% pour les porteurs de gène modifié mâles et de 86% pour les porteuses. La pénétrance d'un second cancer colorectal dans les 5 ans suivant le premier événement était de 20% pour les porteurs et de 11% pour les porteuses.

LIQUN DIAO & RICHARD J. COOK (University of Waterloo)

[Monday June 4 / lundi 4 juin, 11:00]

Analysis of Recurrent Exacerbations in Chronic Disease via Vine Copula Models

Analyse d'exacerbations récurrentes dans les maladies chroniques via des modèles de copules avec des vignes

In some chronic disease settings recurrent events have a non-ignorable duration and are characterized as episodes. We formulate models in which the onset times are generated according to a Markov time scale and the durations of the episodes are governed by a semi-Markov process. To reflect multiple dependencies, we consider a construction according to vine copula models in which the multivariate joint density is decomposed into a cascade of pairwise conditional density functions. We propose candidate models, study properties of them, assess treatment effects, and examine the biases arising from simple analyses which ignore some dependencies.

Dans certains contextes de maladies chroniques, les événements récurrents ont une durée non-ignorable et sont définis comme des épisodes. Nous formulons des modèles dans lesquels les temps de début des épisodes sont générés par une échelle de temps de Markov et les durées des épisodes sont gouvernées par un processus semi-markovien. Pour refléter les dépendances multiples, nous considérons une construction de modèles de copules avec des vignes, dans lesquels la densité multivariée jointe est décomposée en une cascade de fonctions de densités conditionnelles par paires. Nous proposons des modèles possibles, étudions leurs propriétés, évaluons les effets du traitement et examinons les biais des analyses simples qui ignorent certaines dépendances.

HUA SHEN & RICHARD J. COOK (University of Waterloo)

[Monday June 4 / lundi 4 juin, 11:15]

A Sequential Mover-Stayer Model for Recurrent Events with "Burn-out"

Modèle mobile-stable séquentiel pour les événements récurrents avec épuisement

Patients with affective disorder experience repeated exacerbations but some researchers in

Les patients souffrant de troubles affectifs vivent des exacerbations répétées, mais certains chercheurs en psychia-

1B-C2: Frailty Models and the Analysis of Recurrent Events Modèles de fragilité et analyse d'événements récurrents

psychiatry think this process may “burn-out” after which patients will not experience any new exacerbations. We describe a sequential mover-stayer process to model such recurrent events. “Movers” follow a continuous-time multistate Markov model with non-zero transition intensity. When an individual becomes a “stayer” they enter a subject-specific absorbing state and hence acquire a null transition intensity. An EM algorithm is developed to estimate the parameters for the dynamic mover-stayer process as well as the transition intensities for the hidden Markov model.

trie pensent que ce processus peut s'épuiser, après quoi les patients ne souffriront plus d'exacerbations. Nous décrivons un processus mobile-stable séquentiel pour modéliser de tels événements récurrents. Les individus “mobiles” suivent un modèle de Markov multi-état en temps continu avec intensité de transition non nulle. Lorsqu'un individu devient “stable”, il atteint un état sans retour spécifique au sujet et une intensité de transition nulle. Nous développons un algorithme EM pour estimer les paramètres du processus mobile-stable dynamique et les intensités de transition pour le modèle de Markov caché.

1B-C3: Predictive Models and Survey Methods Modèles de prédiction et méthodes d'enquête

Chair/Président: Christian Léger (Université de Montréal)

Rozanski Hall, room/local 105

Abstracts/Résumés

CHRISTOPHER VAN BOMMEL, ANGELA THOMPSON, LAURIE WADSWORTH & XU WANG (St. Francis Xavier University)

[Monday June 4 / lundi 4 juin, 10:15]

Keeping Pace: Physical Activity and Dietary Intake of Nova Scotia Students

Keeping Pace : Activité physique et apport alimentaire des étudiants de Nouvelle-Écosse

Keeping Pace is a study of physical activity and dietary intake of students in Grades 3, 7, and 11 in Nova Scotia conducted in 2009. The goals of this project are to determine the physical activity levels and usual food intakes of these boys and girls, and the factors that influence their physical activity and dietary intake. In this presentation, I discuss the method used for sample selection which involved a stratification and two-stage cluster sampling, the methods used to analyze the data collected, and some key results.

Keeping Pace est une étude réalisée en 2009 sur l'activité physique et l'apport alimentaire des étudiants de 3e, 7e et 11e année de Nouvelle-Écosse. Les objectifs de ce projet consistent à déterminer le niveau d'activité physique et l'apport alimentaire habituel de ces jeunes garçons et filles, ainsi que les facteurs qui influencent ces variables. Dans cet exposé, je discute la méthode de sélection des échantillons utilisée, qui comporte une stratification et un échantillonnage en grappes stratifié à deux degrés, les méthodes utilisées pour l'analyse des données recueillies ainsi que certains résultats clés.

MOSHE FEDER & DANNY PFEFFERMANN (University of Southampton, U.K.)

[Monday June 4 / lundi 4 juin, 10:30]

Empirical Likelihood Approach to Non-Response

Approche par la vraisemblance empirique pour la non-réponse

Surveys are often affected by informative nonresponse (INR), which is a challenging problem when fitting models to survey data. Methods involving weighting-class adjustments and calibration were developed years ago. Recently, methods that assume parametric models for the response probabilities and calibration equations were proposed (Chang & Kott, 2008). Qin et al. (2002) consider an empirical likelihood (EL) method for INR assuming a parametric model for the response probabilities and known population means of the covariates. In this presentation we study EL methods for model estimation that account for informative sampling and INR, without assuming a PM for the response probabilities.

La non-réponse informative a souvent une incidence sur les enquêtes, ce qui constitue un problème de taille lors de l'ajustement de modèles aux données d'enquête. Les méthodes qui comportent des ajustements par pondération de classes et de la calibration ont été conçues il y a plusieurs années. Récemment, Chang & Kott (2008) ont proposé des méthodes basées sur des modèles paramétriques pour la probabilité de réponse et les équations de calibrage. Qin et al. (2002) considèrent une méthode par la vraisemblance empirique pour la non-réponse informative en supposant un modèle pour les probabilités de réponse et la connaissance de la moyenne des covariables dans la population. Dans cette présentation, nous étudions les méthodes par la vraisemblance empirique pour estimer un modèle en tenant compte de la non-réponse et des sondages informatifs, sans faire l'hypothèse d'un modèle paramétrique pour les probabilités de réponse.

1B-C3: Predictive Models and Survey Methods Modèles de prédiction et méthodes d'enquête

ISABEL MOLINA & NIRIAN MARTÍN (Universidad Carlos III de Madrid, Spain)

[Monday June 4 / lundi 4 juin, 10:45]

Empirical Best Estimation under a Nested Error Linear Regression Model with Log Transformation

Meilleure estimation empirique pour un modèle de régression linéaire avec erreurs emboîtées et transformation logarithmique

Log transformation of the response is often needed when fitting models to socio-economic variables. The reason is that monetary welfare variables such as income are markedly skewed and this transformation might make their distribution approximately normal. However, typical parameters of interest are the area means of the untransformed welfare variables, which will be means of the exponentials. For these particular non-linear parameters an exact analytical expression for best predictor is obtained. An analytical expression for the mean squared error of the new empirical best predictor that is second order correct for large number of areas is also derived.

Une transformation logarithmique de la réponse est souvent nécessaire pour ajuster des modèles pour des variables socio-économiques. Ceci s'explique par le fait que les variables monétaires comme le revenu sont fortement asymétriques et que cette transformation peut rendre leur distribution approximativement normale. Toutefois, les paramètres d'intérêt typiques sont les moyennes dans les domaines pour les variables non-transformées, qui seront les moyennes de valeurs dont on aura pris l'exponentielle. Pour ces paramètres non-linéaires spécifiques, on obtient une expression analytique exacte pour le meilleur estimateur. Une expression analytique pour l'erreur moyenne au carré du nouveau meilleur prédicteur empirique est aussi obtenue. Celle-ci est correcte au deuxième ordre s'il y a un grand nombre d'aires.

FATEMEH DORRI & ALI GHODSI (University of Waterloo)

[Monday June 4 / lundi 4 juin, 11:00]

Covariate Shift Adaptation by Metric Learning

Adaptation au décalage d'une covariable par apprentissage métrique

A common assumption in most predictive models is that training data and test data are drawn from the same underlying distribution. In many applications, however, this assumption may not be correct. We propose a novel algorithm to address this problem. The proposed method computes a transformation that projects the data to a new space with two properties. First, the distribution of training and test data are as close as possible in the transformed space, and second, dependency between predictors and response variable is maximized. This method can also reduce the dimensionality of the data while it preserves the aforementioned properties.

Une hypothèse commune à la plupart des modèles de prévision est que les données d'entraînement et les données de test proviennent de la même distribution. Pour plusieurs applications, cette hypothèse peut toutefois être incorrecte. Nous proposons un nouvel algorithme pour faire face à ce problème. La méthode suggérée calcule une transformation qui projette les jeux de données dans un nouvel espace ayant deux propriétés spécifiques. Premièrement, la distribution des jeux de données d'entraînement et de test sont aussi semblables que possible dans ce nouvel espace et, deuxièmement, la dépendance entre les prédicteurs et la variable dépendante est maximisée. Cette méthode peut également réduire la dimensionnalité des données tout en conservant ces deux propriétés.

YAN YUAN (University of Alberta)

[Monday June 4 / lundi 4 juin, 11:15]

Prediction Error Estimation with a Changing Covariates Distribution

Estimation de l'erreur de prévision lorsque la distribution des covariables est changeante

1B-C3: Predictive Models and Survey Methods Modèles de prédiction et méthodes d'enquête

When statistical models are used to predict the values of unobserved random variables, the prediction accuracy is often quantified via loss functions. The expected loss over the covariates distribution is prediction error. Current estimators of prediction error assume that the distribution of covariates is unchanged in the new data as in the “training” data used to specify the predictive model. We relax this condition on covariate distribution and propose an estimator that weighs the prediction losses on the training data to estimate the average loss over the covariate distribution in the new data. Different methods of choosing weights are compared.

Quand des modèles statistiques sont utilisés pour prédire les valeurs de variables aléatoires non observées, l'exactitude de la prévision est souvent quantifiée par des fonctions de perte. L'espérance de la perte sur la distribution de covariables constitue l'erreur de prévision. Les estimateurs actuels de l'erreur de prévision supposent que la distribution des covariables reste inchangée dans les nouvelles données par rapport aux données “d'entraînement” utilisées pour la spécification du modèle prédictif. Nous relâchons cette condition pour la distribution des covariables et nous proposons un estimateur qui pondère les pertes relatives aux prévisions sur les données d'entraînement pour estimer la perte moyenne sur la distribution des covariables dans les nouvelles données. Nous comparons différentes méthodes pour choisir les poids.

1B-C4: Times Series Models Modèles de séries chronologiques

Chair/Présidente: Yulia Gel (University of Waterloo)

Rozanski Hall, room/local 102

Abstracts/Résumés

MACIEJ AUGUSTYNIAK (Université de Montréal)

[Monday June 4 / lundi 4 juin, 10:15]

Estimation of the Markov-switching GARCH Model by a Monte Carlo EM Algorithm

Estimation pour le modèle GARCH à changement de régime markovien par un algorithme EM Monte Carlo

Markov-switching generalized autoregressive conditional heteroscedasticity (MS-GARCH) models offer rich dynamics to model financial data. Estimating these path dependent models is a challenging task because the exact computation of the likelihood is infeasible in practice. This led some authors to propose estimation methods that do not depend on the likelihood while others suggested estimating by maximum likelihood modified versions of the model. In this presentation, I propose a novel approach based on the Monte Carlo expectation-maximization algorithm to calculate the maximum likelihood estimator of the MS-GARCH model. Moreover, I discuss its practical implementation and demonstrate its effectiveness.

Les modèles GARCH (autorégressifs conditionnellement hétéroscédastiques généralisés) à changement de régime markovien permettent une modélisation dynamique et riche des données financières. Cependant l'estimation de ces modèles est difficile vu l'impossibilité pratique d'un calcul précis de la vraisemblance. Certains auteurs proposent donc des méthodes d'estimation indépendantes de la vraisemblance, d'autres une estimation par vraisemblance maximale à partir de versions modifiées du modèle. Je propose une approche originale, fondée sur l'algorithme d'espérance-maximisation de Monte Carlo pour calculer l'estimateur du maximum de vraisemblance du modèle MS-GARCH. J'en discute l'application pratique et en prouve l'efficacité.

WEIWEI LIU (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 10:30]

General Markov Tools with Application to GARCH Models

Outils de Markov généralisés avec application aux modèles GARCH

Generalized autoregressive conditional heteroscedastic (GARCH) processes are important tools in financial econometrics. To estimate GARCH models, quasi-maximum likelihood (QML) estimation is often utilized as it produces a consistent estimator with limiting Gaussian distribution when certain parametric constraints are imposed. Thus finding weak conditions that ensure such asymptotic properties has attracted considerable attention during the past decade. This presentation studies the pure GARCH processes under a general Markov model framework and develops the asymptotic theory of the QML estimator from a set of Markov

Les processus autorégressifs conditionnels hétéroscédastiques généralisés (GARCH) sont des outils importants pour l'économétrie financière. Pour estimer ces modèles, la méthode du quasi-maximum de vraisemblance est souvent utilisée car elle produit un estimateur convergent de loi limite gaussienne sous certaines contraintes paramétriques. En conséquence, une attention particulière a été apportée dans la dernière décennie à identifier des conditions faibles qui garantissent ces propriétés asymptotiques. Cet exposé étudie le processus GARCH dans le cadre d'un modèle de Markov général. Le résultat obtenu avec les méthodes Markov est comparé avec les résultats présents dans la littérature.

1B-C4: Times Series Models Modèles de séries chronologiques

tools. The Markov methods result obtained is compared with established results from literature.

VYACHESLAV LYUBCHICH & YULIA GEL (University of Waterloo)

[Monday June 4 / lundi 4 juin, 10:45]

ARCH/GARCH Models and Trends: From Finance to Medicine, Weather and Further

Modèles ARCH/GARCH et tendance : de la finance à la médecine, météorologie et plus encore

Although conditional heteroscedasticity in time series was originally attributed only to financial processes, similar ARCH/GARCH effects are now detected in a variety of other applications, from wind speed prediction and air/water quality analysis to heart rate variability and patent volatility. In this talk we focus on practical aspects of detecting trends of a general form in time series with an ARCH/GARCH structure and discuss how it could change our interpretation of temporal dynamics in social, medical and environmental studies.

Bien que l'hétéroscédasticité conditionnelle des données en séries ait d'abord été attribuée uniquement aux processus financiers, des effets ARCH/GARCH similaires sont maintenant détectés dans une variété d'autres applications, de la prédiction des vitesses du vent et l'analyse de la qualité de l'eau et de l'air à l'analyse de la variabilité du rythme cardiaque et la volatilité des brevets. Dans cet exposé, nous nous concentrons sur les aspects pratiques de la détection de tendance générale dans des données chronologiques avec une structure ARCH/GARCH et examinons la façon dont cela pourrait changer notre interprétation des dynamiques temporelles dans des études sociales, médicales et environnementales.

ZI ZHEN LIU, REG KULPERGER & HAO YU (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 11:00]

Variable Selection and Estimation in AR Modelling via Adaptive LASSO Procedure

Sélection de variables et estimation pour modèles AR via une procédure LASSO adaptative

The LASSO (Tibshirani, 1996) and the adaptive LASSO (Zou, 2006) are widely used in regression analysis for simultaneous variable selection and parameter estimation. We apply the adaptive LASSO to autoregressive time series modelling by formulating adaptive weights as functions of the Yule-Walker estimators and the partial autocorrelations of an AR process. The earlier papers ignore the autocorrelation information inherent in an AR process. The proposed adaptive LASSO shows promising results for modelling stationary AR processes.

Le LASSO (Tibshirani, 1996) et le LASSO adaptatif (Zou, 2006) sont utilisés couramment en régression pour simultanément choisir les variables et estimer les paramètres. Nous appliquons le LASSO adaptatif à la modélisation de données chronologiques autorégressives (AR) en formulant des poids adaptatifs qui sont fonction des estimateurs de Yule-Walker et des autocorrélations partielles d'un processus AR. Les articles publiés à ce jour ignorent l'information sur l'autocorrélation inhérente à un processus AR. Le LASSO adaptatif proposé donne des résultats prometteurs pour la modélisation des processus AR stationnaires.

WESLEY BURR & GLEN TAKAHARA (Queen's University), **HWASHIN H. SHIN** (Health Canada)

[Monday June 4 / lundi 4 juin, 11:15]

Towards Developing an Air Health Index using Spectral Decomposition of GAMs

Élaboration d'un indice de la qualité de l'air au moyen de la décomposition spectrale des MAG

The Air Health Index seeks to model the national risk due to short-term effects of air pollution on

L'indice de la qualité de l'air vise à modéliser le risque, à l'échelle nationale, relié aux effets à court terme de la pollu-

1B-C4: Times Series Models Modèles de séries chronologiques

mortality. The standard model uses natural-spline-based smoothers to account for long-term variations in the mortality. From a spectral point of view, it is natural to think of smooth, long-term variations as low-frequency components. This points to spectral filtering as a possible alternative technique, where the smoothers are replaced by high-pass filters. We apply high-pass filters to the predictors and response of the GAM. The results are comparable to previous work, with several valuable improvements, including interpretation, computational efficiency, and reduced degrees-of-freedom used.

tion de l'air sur la mortalité. Le modèle standard tient compte des variations à long terme de la mortalité par un lissage au moyen de fonctions splines naturelles. D'un point de vue spectral, il est naturel de considérer les variations lisses à long terme comme des composantes de basse fréquence. Ainsi, le filtrage spectral constitue une autre technique possible, où les outils de lissage sont remplacés par des filtres passe-haut. Nous appliquons les filtres passe-haut aux prédicteurs et aux réponses des MAG. Les résultats sont comparables à ceux des travaux antérieurs, mais comportent plusieurs améliorations importantes, y compris en matière d'interprétation, d'efficacité de calculs et d'utilisation réduite de degrés de liberté.

1C-A: SSC Impact Award Address
Allocution du récipiendaire du prix pour impact de la SSC

Organizer and Chair / Responsable et président: James Stafford (University of Toronto)

Rozanski Hall, room/local 104

Abstract/Résumé

FRANCIS ZWIERS (Pacific Climate Impacts Consortium, University of Victoria)

[Monday June 4 / lundi 4 juin, 13:15]

On Bridging Two Solitudes – Examples of a Slow Diffusive Process

Sur la réunion de deux solitudes – exemples d'un processus de diffusion lent

While the interaction between the climatic and statistical sciences has improved steadily, and is very active in some areas, the interaction process is still best characterized as a slow diffusion. This comes about for a number of reasons – including the use of different scientific languages, dissemination practices, and ingrained community practices. But the slowness of the process also comes about because many climate research problems are simply hard from a statistical perspective, or because technological limitations slow the implementation of advanced statistical methods. I will describe three such examples, all of which would benefit from additional input from statisticians.

Bien que l'interaction entre la science du climat et la statistique s'améliore constamment, le processus de communication entre les deux domaines demeure encore bien caractérisé comme étant une diffusion lente. Cela s'explique par plusieurs raisons – y compris l'utilisation de différents langages scientifiques, les différences dans les méthodes de transmission de connaissances et les pratiques communautaires enracinées. Mais la lenteur du processus vient aussi simplement du fait que de nombreux problèmes de recherche sur le climat sont complexes d'un point de vue statistique, ou encore parce que les limites technologiques ralentissent la mise en œuvre de méthodes statistiques complexes. Je décrirai trois exemples qui appartiennent à cette catégorie et pour lesquels la contribution de statisticiens serait bénéfique.

1C-11: Asymptotic Methods in Stochastics: a Session in Honour of Miklós Csörgő's 80th Birthday

Séance en l'honneur du 80e anniversaire de Miklós Csörgő

Chair/Président: Don McLeish (University of Waterloo)

Organizer/Responsable: Barbara Szyszkowicz (Carleton University)

Rozanski Hall, room/local 103

Abstracts/Résumés

DON DAWSON (Carleton University)

[Monday June 4 / lundi 4 juin, 13:15]

Path Properties of Fifty Years of Research in Probability and Statistics: a Tribute to Miklós Csörgő

Les propriétés d'une trajectoire de cinquante ans de recherche en probabilités et en statistique : un hommage à Miklós Csörgő

Over the past fifty years Miklós Csörgő has produced a remarkable series of contributions of great importance to probability and statistics that have had a profound impact on the development of these subjects. At the same time he has provided inspiration for the host of graduate students and post-doctoral fellows he has supervised and who have gone on to have an important impact on the development of probability and statistics across Canada and abroad. In my talk I will give some personal observations on his exceptional contributions to our discipline and our years together as colleagues.

Au cours des cinquante dernières années, Miklós Csörgő a produit une série remarquable de contributions scientifiques très importantes qui ont contribué à façonner les fondements des sujets qu'il a abordés. Ce faisant, il est demeuré une source d'inspiration pour de nombreux jeunes chercheurs qu'il a formés, et qui contribuent à leur tour à l'avancement de la science. Dans mon exposé, je vais partager quelques observations personnelles à propos de ses contributions exceptionnelles à notre discipline, mais aussi du temps que nous avons passé ensemble en tant que collègues.

DON MCLEISH (University of Waterloo)

[Monday June 4 / lundi 4 juin, 13:45]

Simulation of some Continuous-time Models in Finance

Simulation de modèles en temps continu en finance

Currently popular continuous-time financial models include diffusions, Lévy processes, stochastic volatility models and others for which simple forms for the transition densities are often not available. We explore methods for unbiased simulations of such models, including pathwise importance sampling, simulating based on characteristic functions. Examples include unbiased estimates of the prices of forward starting options in the Heston stochastic volatility model and continuously monitored Parisian options. The author owes a long-standing debt to Miklós Csörgő, not only for his related research contributions to pathwise properties of stochastic processes, but for his infectious en-

Parmi les modèles financiers en temps continu actuellement populaires figurent les diffusions, les processus de Lévy, les modèles de volatilité stochastique et d'autres modèles pour lesquels des formes simples de la densité de transition sont souvent indisponibles. Nous explorons des méthodes de simulations non biaisées de tels modèles, y compris l'échantillonnage préférentiel par trajectoire, basées sur des fonctions caractéristiques. Des exemples comprennent des estimations sans biais des prix d'options démarrant à terme ("forward starting options") dans le modèle de volatilité stochastique de Heston, et des prix d'options parisiennes en temps continu. L'auteur est profondément reconnaissant à l'endroit de Miklós Csörgő, non seulement pour ses contributions en recherche sur les propriétés des trajectoire de

1C-I1: Asymptotic Methods in Stochastics: a Session in Honour of Miklós Csörgő's 80th Birthday

Séance en l'honneur du 80e anniversaire de Miklós Csörgő

thusiasm for research in probability and its applications.

processus stochastiques, mais également pour son enthousiasme contagieux dans la recherche en probabilités et ses applications.

JON N.K. RAO (Carleton University)

[Monday June 4 / lundi 4 juin, 14:15]

Large-sample Theory for Sampling from Finite Populations: Selected Developments

Théorie asymptotique pour l'échantillonnage à partir de populations finies : quelques développements

Extensive developments in asymptotic theory exist for the IID case. In the case of survey sampling, IID assumption is not satisfied due to survey design features such as stratification, unequal selection probabilities and clustering. In this talk, I will trace some key developments in large-sample theory for surveys, based on extensions and modifications of the IID theory. I will present some central limit theorems and results on design-consistency of variance estimators that can be used to construct asymptotically valid confidence intervals. Some recent work on goodness-of-fit tests will also be mentioned.

La théorie asymptotique pour le cas IID est largement développée. Dans le cas d'enquêtes par sondage, l'hypothèse IID ne tient plus à cause des caractéristiques du plan de sondage, comme la stratification, les probabilités inégales de sélection et les grappes. Dans cet exposé, je retrace quelques développements importants de la théorie asymptotique dans les enquêtes. Ces développements se fondent sur des extensions et des modifications de la théorie dans le cas IID. Je présente quelques théorèmes centraux limites ainsi que des résultats de convergence sous le plan d'estimateurs de la variance, qui peuvent être utilisés pour construire des intervalles de confiance asymptotiquement valides. J'aborde également quelques travaux récents sur les tests d'adéquation.

1C-I2: Inference and Applications using Estimating Functions Inférence à l'aide de fonctions d'estimation et applications

Chair/Président: Aerambamoorthy Thavaneswaran (University of Manitoba)

Organizer/Responsable: Melody Ghahramani (University of Winnipeg)

Alexander Hall, room/local 100

Abstracts/Résumés

PETER X.-K. SONG, FEI WANG & LU WANG (University of Michigan)

[Monday June 4 / lundi 4 juin, 13:15]

Estimating Equations Approach to Merging Longitudinal Data

Approche par équations d'estimation pour la fusion de données longitudinales

Merging data from multiple studies has been widely adopted in biomedical studies to reach more powerful statistical analyses. This talk concerns two major issues related to merging longitudinal datasets collected from similar studies. First, we develop a rigorous hypothesis testing procedure to assess the validity of data merging. Second, we propose a flexible joint estimation approach for analyzing merged data and accounting for different within subject correlations and follow-up schedules in different studies. We establish large sample properties for the proposed methods. Using simulation studies, we compare our method with several popular methods. We illustrate our methods through data examples.

La fusion de données à partir de plusieurs études a été largement adoptée dans les études biométriques pour obtenir des analyses statistiques plus puissantes. Cet article concerne deux défis importants pour la fusion de jeux de données longitudinales recueillis à partir d'études similaires. D'abord, nous élaborons une procédure de test d'hypothèse rigoureuse pour analyser la validité de la fusion des données. Ensuite, nous proposons une approche d'estimation flexible pour analyser les données fusionnées en tenant compte des différences dans les corrélations propres à un sujet, et faisant le suivi des différents échéanciers dans les diverses études. Nous établissons les caractéristiques asymptotiques des méthodes proposées. À l'aide d'études de simulation, nous comparons notre méthode avec d'autres méthodes populaires. Nous illustrons notre méthode par des exemples de données.

GRACE Y. YI (University of Waterloo), **NANCY REID** (University of Toronto)

[Monday June 4 / lundi 4 juin, 13:45]

Biased Estimating Functions and Applications

Fonctions d'estimation biaisées et applications

I will discuss the use of estimating functions that are not unbiased. Typically, to result in consistent estimators, unbiasedness of estimating functions is a pre-requisite. However, it is sometimes easier to find a useful estimating function that is biased, especially in the presence of missing data or misclassified observations. I will show that the root of the estimating function can be modified to give a consistent and asymptotically normal estimator, and illustrate this on several examples with binary data. I will compare this to the alternative approach

Je vais discuter de l'utilisation de fonctions d'estimation qui sont biaisées. Généralement, l'absence de biais est requise pour obtenir des estimateurs convergents. Cependant, il est parfois plus facile de trouver une fonction d'estimation utile qui est biaisée, particulièrement en présence de données manquantes ou d'observations classées incorrectement. Je vais montrer que la solution aux fonctions d'estimation peut être modifiée pour donner un estimateur convergent et asymptotiquement normal, ce que j'illustrerai par différents exemples sur des données binaires. Je vais comparer cette approche d'ajustement à l'alternative habi-

1C-I2: Inference and Applications using Estimating Functions Inférence à l'aide de fonctions d'estimation et applications

of adjusting the estimating function, and show that it can be more efficient.

tuelle et montrerai qu'elle peut être plus efficace.

MICHELLE QIAN ZHOU (Harvard University), **PETER X.-K. SONG** (University of Michigan), **MARY THOMPSON** (University of Waterloo)

[Monday June 4 / lundi 4 juin, 14:15]

Information Ratio Test for Model Misspecification in Quasi-Likelihood Inference

Test du rapport d'information pour l'erreur de spécification de modèle dont l'inférence est basé sur la quasi-vraisemblance

In general, model misspecification can lead to invalid inference for parameter estimation. In the context of quasi-likelihood inference, most of the existing statistical methods focus on assessing the validity of the mean structure. However, limited work addresses the adequacy of the variance/covariance (Var/Cov) structure, and more specifically, there lacks a powerful systematic statistical test for model misspecification in Var/Cov. In this talk, I will introduce a novel and unified framework for testing such misspecification, through a comparison between two forms of information matrix. Our method shows substantial improvement and is more robust in comparison to several popular existing statistical methods.

En général, l'erreur de spécification du modèle peut engendrer une inférence invalide pour l'estimation de paramètres. Dans le cadre d'inférence par la quasi-vraisemblance, la plupart des méthodes statistiques existantes mettent l'accent sur la validation de la structure de la moyenne. Cependant, peu de travaux mettent en question l'adéquation de la structure de covariance. Plus particulièrement, il manque un test statistique systématique puissant pour l'erreur de spécification de la covariance du modèle. Dans cet exposé, je vais présenter un nouveau cadre unifié pour tester cette erreur de spécification par une comparaison entre deux formes de matrices d'information. Notre méthode montre une amélioration substantielle et est plus robuste comparée aux autres méthodes statistiques populaires existantes.

1C-13: Modern Multiple Imputation Approaches to Missing Data Problems Approches modernes à l'imputation multiple pour les problèmes de données manquantes

Organizer and Chair / Responsable et président: Russell Steele (McGill University)

Session sponsored by the Biostatistics Section
Séance parrainée par le Groupe de biostatistique

Rozanski Hall, room/local 101

Abstracts/Résumés

BEN GOODRICH (Columbia University)

[Monday June 4 / lundi 4 juin, 13:15]

Evaluating the Missing-At-Random Assumption after Multiply Imputing Missing Data

Évaluation de l'hypothèse de « données manquantes au hasard » après l'imputation multiple de données manquantes

Most algorithms for multiply imputing missing values make a “Missing-At-Random” (MAR) assumption, which loosely states that the probability that a data point is missing is conditionally independent from the values of the missing data. Thus, many researchers believe that the resulting completed data cannot falsify the MAR assumption. However, the presence of observed data among the completed data can provide some leverage. This study investigates ways to falsify the MAR assumption by applying selection models and covariance structure analysis to the completed data. Simulations are conducted to determine how effective these techniques are.

La plupart des algorithmes d'imputation multiple de valeurs manquantes reposent sur l'hypothèse de « données manquantes au hasard » (MAH), qui affirme grosso modo que la probabilité qu'une donnée soit manquante est conditionnellement indépendante des valeurs de celle-ci. Par conséquent, plusieurs chercheurs croient que les données complétées résultantes ne peuvent pas falsifier l'hypothèse MAH. Cependant, la présence de données observées parmi les données complétées peut constituer un point d'appui. Nous étudions des manières de falsifier l'hypothèse MAH en appliquant des modèles de sélection et une analyse de la structure de covariance aux données complétées. L'efficacité de ces techniques est évaluée à l'aide de simulations.

NATHANIEL SCHENKER (National Center for Health Statistics and Centers for Disease Control and Prevention)

[Monday June 4 / lundi 4 juin, 13:45]

Multiple Uses for Multiple Imputation at the U.S. National Center for Health Statistics

Diverses utilisations de l'imputation multiple au National Center for Health Statistics des États-Unis

This talk describes three recent or potential applications of multiple imputation at the National Center for Health Statistics, to illustrate various types of problems for which the technique can be used. One application involves missing data on body scans in the National Health and Nutrition Examination Survey (NHANES). The second seeks to improve on analyses of self-reported data from the National Health Interview Survey by imputing clinical values using models fitted to the smaller NHANES. The third uses observed birthweights

Cet exposé décrit trois applications récentes ou potentielles de l'imputation multiple au National Center for Health Statistics des États-Unis visant à illustrer différents types de problèmes pour lesquels cette technique peut être employée. L'une des applications concerne des données manquantes sur des images de scanner corporel du National Health and Nutrition Examination Survey (NHANES). La seconde cherche à améliorer les analyses des données autodéclarées du National Health Interview Survey par l'imputation de valeurs cliniques en utilisant des modèles ajustés au plus petit NHANES. La troisième utilise les poids observés à la

1C-13: Modern Multiple Imputation Approaches to Missing Data Problems Approches modernes à l'imputation multiple pour les problèmes de données manquantes

and mixture models to identify questionable gestational ages in U.S. natality data, and seeks to impute more plausible gestational ages.

naissance et des modèles de mélange pour identifier les âges gestationnels incertains dans les données de natalité aux États-Unis et cherche à imputer des âges gestationnels plus plausibles.

YAJUAN SI & JEROME P. REITER (Duke University)

[Monday June 4 / lundi 4 juin, 14:15]

Multiple Imputation in Panel Studies with Attrition and Refreshment Samples

Imputation multiple dans des études de panel avec érosion et échantillons ajoutés

Panel data provides advanced understanding of mass behavior along time with multiple follow-up waves. However, panel studies suffer from attrition, which reduces the effective sample size and can lead to biased estimates if the tendency to drop out is systematically related to the substantive outcome of interest. A refreshment sample as an external data source includes new, randomly sampled respondents who are provided with the same questionnaire at the same time as a subsequent wave of the panel. We utilize refreshment samples to assess the effects of panel attrition and to correct for biases via multiple imputation.

Les données de panel permettent une meilleure compréhension du comportement de masse au fil du temps à l'aide de vagues de suivi multiples. Cependant, les études de panel souffrent d'érosion, ce qui diminue la taille effective de l'échantillon et peut entraîner un biais dans les estimations si le désistement est systématiquement lié au résultat d'intérêt. L'échantillon ajouté en tant que source externe de données comprend de nouveaux répondants sélectionnés aléatoirement à qui le même questionnaire est administré conjointement avec une vague subséquente du panel. Ces échantillons ajoutés sont employés pour évaluer les effets de l'érosion du panel et pour corriger les biais au moyen de l'imputation multiple.

1C-I4: Reflections on 40 Years of Teaching Statistics Réflexions sur 40 ans d'enseignement de la statistique

Organizer and Chair / Responsable et présidente: Alison Gibbs (University of Toronto)

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

Alexander Hall, room/local 200

Description

This panel session convenes five experienced Canadian statistics educators to comment on the past, present, and future of statistics education. What has changed over the last 40 years? What is the same? And what should change or should stay the same as we go forward? The panelists will share their observations, experiences, transformations, and lessons they have learned and opinions they have cultivated during their careers teaching statistics. The session will feature reflections from each of the panelists, followed by discussion led by the moderator, and ending with the opportunity for comments and questions from the audience.

Ce panel rassemble cinq enseignants canadiens de statistique expérimentés qui commenteront le passé, le présent et le futur de l'enseignement de la statistique. Qu'est-ce qui a changé au cours des 40 dernières années? Qu'est-ce qui est demeuré inchangé? Qu'est-ce qui devrait changer ou demeurer inchangé dans l'avenir? Les panélistes partageront leurs observations et expériences, ainsi que les transformations, leçons tirées et opinions cultivées au cours de leur carrière dans l'enseignement de la statistique. Au cours de la séance, chacun des panélistes présentera ses réflexions, qui seront suivies d'une discussion animée par le modérateur. L'auditoire aura ensuite l'occasion de commenter et de poser des questions.

JOHN BREWSTER (University of Manitoba)

PAUL COREY (University of Toronto)

CHRIS FIELD (Dalhousie University)

JERRY LAWLESS (University of Waterloo)

LARRY WELDON (Simon Fraser University)

1C-C1: Analysis of Longitudinal Data Analyse de données longitudinales

Chair/Président: Paramjit Gill (University of British Columbia, Okanagan)

MacNaughton Building, room/local 105

Abstracts/Résumés

HYUN (JUNE) LIM, PROSANTA MONDAL & STUART SKINNER (University of Saskatchewan)

[Monday June 4 / lundi 4 juin, 13:15]

Joint Modelling of Repeated CD4+ Count Measurements and Survival Time in HIV Study

Modélisation conjointe de mesures répétées de dénombrement de CD4+ et de temps de survie dans une étude sur le VIH

In many clinical studies, data have two types of outcome, longitudinal biomarker measurements and survival time. These two outcomes are often separately analyzed using a mixed effects model and survival model. However, separate models are overly simplified by ignoring the association between two components. Joint modelling is sophisticated, complex approach but enables the longitudinal measurements and survival time to be modelled whilst accounting for association between them. Using a HIV dataset with repeated CD4+ count measurements and survival time, we combine a random effects sub-model and Cox or Weibull sub-models through their shared dependence on latent variable.

Dans plusieurs études cliniques, les données sont de deux types : mesures longitudinales de biomarqueurs et temps de survie. Ces deux réponses sont souvent analysées séparément à l'aide d'un modèle à effets mixtes et d'un modèle de survie. Toutefois, les modèles séparés simplifient excessivement le problème en ignorant l'association entre ces composantes. La modélisation conjointe est sophistiquée et compliquée, mais elle permet de modéliser les mesures longitudinales et les temps de survie en prenant en compte leur association. À l'aide d'un jeu de données sur le VIH contenant des mesures répétées du CD4+ et des temps de survie, nous combinons, par l'entremise de leur dépendance à une variable latente, un sous-modèle avec effets aléatoires et un sous-modèle de Cox ou Weibull.

TASNEEM ZAIHRA (Concordia University), **SARA AHMED & PIERRE ERNST** (McGill University)

[Monday June 4 / lundi 4 juin, 13:30]

Evaluating Predictive Validity of Patient Reported Asthma Control, Belief about Medications and Self-Efficacy

Évaluation de la validité prédictive du contrôle de l'asthme, des croyances à l'égard des médicaments et de l'autoefficacité déclarés par le patient

The purpose of this study is to develop statistical models to evaluate the predictors of long-term asthma control. A cohort of 1437 patients with asthma were followed over a one-year period and completed the Asthma Control Test (ACT), mini-asthma quality of life questionnaire (AQLQ), asthma self-efficacy (opinion, knowledge about asthma) scale and beliefs about medication. Long-term asthma control was evaluated by assessing overuse of rescue medication and emergency department (ED) visits. Even a 15% reduction in

Cette étude vise à développer des modèles statistiques pour évaluer les variables explicatives du contrôle à long terme de l'asthme. Une cohorte de 1437 patients atteints d'asthme a été suivie pour une période d'un an. Ils ont tous répondu au questionnaire sur le contrôle de l'asthme (ACT) et au mini questionnaire sur la qualité de vie (AQLQ), ainsi qu'à l'échelle d'autoefficacité (opinions, connaissances sur l'asthme) et à des questions sur leurs croyances à l'égard des médicaments. Le contrôle à long terme de l'asthme a été évalué en fonction de l'utilisation excessive des médicaments de secours et du nombre de visites au service d'ur-

1C-C1: Analysis of Longitudinal Data Analyse de données longitudinales

asthma-related ED visits and hospitalization based on efficient predictive models could result in an important reduction in health-care costs.

gence. D'après des modèles prédictifs efficaces, une réduction de 15% du nombre de visites au service d'urgence et d'hospitalisation en raison de l'asthme pourrait entraîner une importante réduction des coûts des soins de santé.

ABDERAZZAK MOUIHA (Centre de Recherche de l'Institut Universitaire de Santé Mentale de Québec), **SIMON DUCHESNE** (Université Laval)

[Monday June 4 / lundi 4 juin, 13:45]

The Dynamic of Alzheimer's Disease: a Longitudinal Study of Patients Progressing From MCI to AD

La dynamique de la maladie d'Alzheimer : une étude longitudinale des patients ayant progressé de MCI vers AD

Numerous studies suggest that medial temporal lobe atrophy, tracked via hippocampal (HC) volumes and longitudinal rates of change, predicts Alzheimer's disease (AD) in patients with amnesic mild cognitive impairment (aMCI). It should relate to neuropsychological assessment. To determine the shape of this relationship and take into account within-individual changes in HC over time, we used a general mixed model. The control group is used to transform scores. From the ADNI (Alzheimer's Disease Neuroimaging Initiative) study, 48 females and 86 males, who progressed from MCI to AD over time, were included. Results confirm the hypothesis proposed by Jack et al. 2010.

Plusieurs études suggèrent qu'une atrophie du lobe temporal médian, mesurée grâce au volume hippocampique (HC), prédit le déclin de la maladie d'Alzheimer (MA) chez les patients ayant un trouble cognitif léger (TCL). Cette atrophie doit être reliée aux résultats neuropsychologiques. Pour déterminer la forme de cette relation, en tenant compte de la variabilité intra-individuelle du HC, nous utilisons un modèle mixte général. Le groupe contrôle est utilisé pour normaliser les scores. À partir de l'étude ADNI (Alzheimer's Disease Neuroimaging Initiative), 48 femmes et 86 hommes dont la MA a progressé ont été sélectionnés. Nos résultats infirment l'hypothèse proposée par Jack et coll. 2010.

NARGES NAZERI RAD & JERRY LAWLESS (University of Waterloo)

[Monday June 4 / lundi 4 juin, 14:00]

Multi-state Models for Viral Load Patterns

Modèles multi-états pour la charge virale

In many cohorts of individuals infected with the human immune deficiency virus (HIV), viral load is measured at intermittent visit times. Multi-state models with states defined by viral load levels provide a simplistic but useful approach for examining viral load dynamics, and phenomena such as viral blips or viral rebound. In this talk we consider the effect of the gap time between viral load measurements on the estimation of process parameters. Insights gained also apply to other longitudinal studies where biomarker measurements are taken at intermittent visits.

Dans de nombreuses cohortes d'individus séropositifs, la charge virale est mesurée de façon intermittente. Les modèles multi-états dont les états sont définis par la charge virale offrent une approche simpliste mais utile pour étudier la dynamique de la charge et des phénomènes tels les sursauts ou rebonds viraux. Nous étudions l'effet du délai entre les mesures de la charge virale sur l'estimation des paramètres de processus. Nos résultats s'appliquent également à d'autres études longitudinales où des biomarqueurs sont mesurés lors de visites intermittentes.

1C-C1: Analysis of Longitudinal Data Analyse de données longitudinales

DEPENG JIANG, ANDREA E. BOMBAK & QIER TAN (University of Manitoba)

[Monday June 4 / lundi 4 juin, 14:15]

Latent Class Growth Model for National Longitudinal Data

Modèle de croissance à classes latentes pour données longitudinales nationales

This study will discuss statistical assumptions, strengths, limitations, and areas of interpretational confusion for using latent class growth model (LCGM) to analyze longitudinal data. We use data from the National Longitudinal Survey of children and Youth (Statistics Canada and Human Resources Development Canada, 1994-2011) to illustrate how latent class growth model are commonly used in longitudinal research, and what one should take into consideration. We explain how the integration of the latent class growth model and conventional growth model can lead to a more complete understanding of the processes and patterns of human development.

Cette étude examine les hypothèses statistiques, les points forts, les limites et les difficultés d'interprétation de l'utilisation d'un modèle de croissance à classes latentes (LCGM) pour l'analyse de données longitudinales. Nous utilisons des données de l'Enquête longitudinale nationale sur les enfants et les jeunes (Statistique Canada et Ressources humaines et Développement des compétences Canada, 1994-2011) pour illustrer comment les modèles de croissance à classes latentes sont utilisés dans la recherche impliquant des données longitudinales, ainsi que ce qu'on devrait prendre en considération. On explique aussi comment l'intégration du modèle de croissance à classes latentes et du modèle de croissance conventionnel peut mener à une meilleure compréhension des processus et des structures de développement chez l'humain.

CHENGLIN YE, GARY FOSTER, LISA DOLOVICH & LEHANA THABANE (McMaster University)

[Monday June 4 / lundi 4 juin, 14:30]

Comparison of Methods to Model Diastolic and Systolic Blood Pressure Measurements over Time

Comparaison de méthodes de modélisation des mesures de la tension artérielle diastolique et systolique au fil du temps

The pathology behind high blood pressure is still unclear. Most analyses of data from interventional studies treat systolic blood pressure (SBP) and diastolic blood pressure (DBP) separately, ignoring the plausible correlation between them. Such analyses assumed that the two measures were independent of each other. We use data from the Cardiovascular Health Awareness Program (CHAP, see www.CHAPprogram.ca) study to compare the results of the longitudinal analyses of DBP and SBP using univariate (i.e. separately) and multivariate (i.e. jointly) approaches as functions of several baseline risk factors.

La pathologie liée à une tension artérielle élevée n'est pas encore clairement définie. La plupart des analyses de données provenant d'études interventionnelles traitent la tension artérielle systolique (TAS) et la tension artérielle diastolique (TAD) séparément, ignorant la corrélation plausible entre les deux. Ces analyses assument que les deux mesures sont indépendantes l'une de l'autre. À partir de données de l'étude du CHAP (Programme de sensibilisation à la santé cardiovasculaire, voir www.CHAPprogram.ca), nous comparons les résultats des analyses longitudinales des TAS et TAD en utilisant des approches univariées (ou séparées) et multivariées (ou combinées) comme des fonctions de plusieurs facteurs de risque de base.

1C-C2: Estimating Population Size and Growth Estimation de la taille et de la croissance des populations

Chair/Présidente: Llwellyn Armstrong (Ducks Unlimited Canada)

Rozanski Hall, room/local 105

Abstracts/Résumés

NOEL CADIGAN (Marine Institute of Memorial University of Newfoundland)

[Monday June 4 / lundi 4 juin, 13:15]

A Nonparametric Approach to Fitting a Stock-Recruitment Relationship, and Implications for Fisheries Management Reference Points

Approche non-paramétrique d'ajustement d'une relation stock-recrutement et implications pour les points de référence de gestion des pêcheries

Understanding the relationship between parental stock size and subsequent reproduction and recruitment of juveniles to a fishery is a fundamental component of fisheries management. I illustrate a nonparametric approach to estimate stock-recruitment relationships. The approach preserves compensatory density dependence in which the recruitment rate monotonically decreases as stock size increases, which is a basic assumption of commonly used parametric stock-recruitment models. I also illustrate the implications of the nonparametric estimates on maximum sustainable yield reference points, which are fundamental in many fisheries management frameworks around the world. I use the approach to provide nonparametric bootstrapped confidence intervals for reference points.

Comprendre la relation entre la taille des stocks, la reproduction et le recrutement des jeunes pour la pêche est fondamental en gestion des pêcheries. J'illustrerai une approche non-paramétrique de l'estimation des relations entre la taille et le recrutement qui conserve une dépendance face à la densité compensatoire : le taux de recrutement diminue de façon monotone lorsque la taille des stocks augmente, une hypothèse fondamentale des modèles de taille-recrutement paramétriques communément utilisés. J'illustrerai les implications des estimations non-paramétriques pour les points de référence du rendement constant maximal. J'offrirai des intervalles de confiance bootstrap non-paramétriques pour ces points de référence.

MARYLÈNE TROUPÉ & JEAN VAILLANT (Université des Antilles et de la Guyane), **LÉONIDE CELINI** (Université de Paris-Est-Créteil, France)

[Monday June 4 / lundi 4 juin, 13:30]

Sampling of Thinned Point Processes: applications to Aphis Gossypii mortality rates

Échantillonnage de processus ponctuels amincis : application aux taux de mortalité d'Aphis Gossypii

Population abundance is often evaluated from sampling unit count data. Statistical inference on these population's parameters can be carried out with different procedures. We are interested in situations for which there are two sampling dates, one before and one after a particular event affecting the abundance. By considering distributions of individuals as point process realizations, we infer

L'abondance de populations est souvent évaluée à partir de données de comptage dans des unités d'échantillonnage. L'inférence statistique sur certains paramètres de ces populations peut alors être conduite par différents procédés statistiques. Nous nous intéressons au cas où cet échantillonnage est effectué à deux dates d'observation situées avant et après un événement ayant un effet supposé sur cette abondance. En assimilant les répartitions d'individus aux

1C-C2: Estimating Population Size and Growth Estimation de la taille et de la croissance des populations

parameters of the stochastic removal of individuals. The method is applied to count data of *Aphis gossypii* (Glover) collected before and after thunderstorms.

réalisations d'un processus ponctuel, nous estimons et testons les paramètres du modèle stochastique de disparition des individus. Nous appliquons la méthode aux données de comptage d'*Aphis gossypii* (Glover) recueillies avant et après plusieurs orages.

LAWRENCE MCCANDLESS, MICHELLE PATTERSON, LAUREN CURRIE, AKM MONIRUZZAMAN & JULIAN SOMERS (Simon Fraser University)

[Monday June 4 / lundi 4 juin, 13:45]

Counting the Unsheltered Homeless in Edmonton Using Plant-Capture Studies from Toronto

Dénombrement des sans-abris sans refuge à Edmonton à l'aide d'une étude « placer-capturer » de Toronto

Every two years, the city of Edmonton counts the unsheltered homeless during a single day in October. Volunteer enumerators are instructed to approach individuals on predetermined walking routes throughout the city where the homeless are known to congregate. The total count is then reported as an estimate of the size of the homeless population. However, interval estimates are lacking and the sampling frame is poorly understood. In this talk I will use Bayesian analysis to combine data from Edmonton with plant-captures studies of homelessness from Toronto.

Tous les deux ans, la ville d'Edmonton compte les sans-abris sans refuge pendant une journée spécifique en octobre. Des énumérateurs bénévoles sont chargés d'approcher des individus sur les routes piétonnes à travers la ville où les sans-abris se regroupent habituellement. Le total est utilisé comme estimateur de la taille de la population des sans-abris. Toutefois, il n'y a pas d'intervalle de confiance, et le plan d'échantillonnage n'est pas très bien compris. Dans cet exposé, je vais utiliser l'analyse bayésienne pour combiner les données d'Edmonton avec des études «placer-capturer» (plant-capture) des sans-abris de Toronto.

B. JOHANNES EDINGER (First Nations Statistical Institute)

[Monday June 4 / lundi 4 juin, 14:00]

A Stochastic Population Projection Model for Canada's First Nations

Un modèle stochastique de projection de la population pour les Premières Nations du Canada

A stochastic population projection model for First Nations in Canada is produced and demonstrated. The projection model uses time series data and expert knowledge from traditional cohort-component based projections to specify parametric distributions for the stochastic processes governing population growth or decline. Probabilistic intervals for the future course of Canada's First Nations' population trajectory are calculated using Monte Carlo simulations. The advantages of the stochastic approach to population projections over the traditional approach are shown, not only in the context of First Nations, but also for municipalities, cities and economic regions.

Un modèle stochastique de projection de la population pour les Premières Nations au Canada est produit et illustré. Le modèle de projection utilise des données longitudinales et la connaissance d'expert des projections traditionnelles basées sur la méthode des composantes de cohorte pour spécifier les distributions paramétriques pour le processus stochastique gouvernant la croissance ou le déclin des populations. Des intervalles probabilistes pour l'évolution future des Premières Nations au Canada sont calculés par simulations Monte Carlo. Les avantages de l'approche stochastique de projections de population par rapport aux approches traditionnelles sont illustrés, non seulement dans le contexte des Premières Nations, mais aussi pour des municipalités, villes et régions économiques.

1C-C3: Generalized Linear Models, Overdispersion and Zero-Inflation Modèles linéaires généralisés, inflation de la variance et surabondance de zéros

Chair/Président: Thierry Duchesne (Université Laval)

MacNaughton Building, room/local 113

Abstracts/Résumés

MICHAEL EVANS & ZEYNEP BASKURT (University of Toronto)

[Monday June 4 / lundi 4 juin, 13:15]

Bayesian Goodness of Fit and g -Priors

Adéquation bayésienne et lois a priori de type g

Zellner (1986) introduced g -priors for the regression coefficients in a normal linear model as reference informative priors. While the priors are quite natural, the argument on which they are based seems somewhat artificial as it involves the concept of imaginary observations. We provide a direct argument for a generalization of g -priors in the context of generalized linear models. An easily elicited prior is used to derive a Bayesian goodness of fit for the model and the generalized g -prior then arises via conditioning this prior when the model passes. An application to regression with correlated binary data is discussed.

Zellner (1986) a développé les lois a priori de type g pour les coefficients d'un modèle de régression linéaire normal, en tant que lois a priori de référence informatives. Alors que les lois a priori sont assez intuitives, l'argument sur lequel elles sont basées semble quelque peu artificiel puisqu'il utilise le concept d'observations imaginaires. Nous apportons un argument direct pour une généralisation des lois a priori g dans le cadre de modèles linéaires généralisés. Une loi a priori facilement obtenue est utilisée pour évaluer l'adéquation bayésienne du modèle, et la loi a priori g généralisée découle du conditionnement de cette loi a priori lorsque le modèle n'est pas rejeté. Nous discutons d'une application en régression avec des données binaires corrélées.

NIROSHAN WITHANAGE & ALEXANDER R. DE LEON (University of Calgary)

[Monday June 4 / lundi 4 juin, 13:30]

Joint Estimation of Specificities and Sensitivities in Reader-based Diagnostic Studies

Estimation conjointe des spécificités et des sensibilités des études fondées sur la lecture du diagnostic

Binocular data typically arise in ophthalmology where pairs of eyes are evaluated, through some diagnostic procedure, for the presence of certain diseases or pathologies. Treating eyes as independent and adopting the usual approach in estimating the sensitivity and specificity of a diagnostic test may yield incorrect inferences. This study is concerned with several pathologies in eyes which are read by several readers. A likelihood-based method of estimating sensitivities and specificities via hierarchical generalized linear models is proposed to meaningfully delineate the various correlations in the data. Data from a study on diabetic retinopathy are analyzed to illustrate the methodology.

Les données binoculaires sont typiques du domaine de l'ophtalmologie, où des paires d'yeux sont examinées selon certaines procédures de diagnostic pour détecter la présence de maladies ou de pathologies. Le traitement indépendant des yeux et l'adoption de l'approche courante d'estimation de la sensibilité et de la spécificité d'un test diagnostic peuvent donner lieu à des conclusions erronées. Cette étude porte sur plusieurs pathologies des yeux dont le diagnostic est posé par plusieurs lecteurs. Nous proposons une méthode d'estimation des sensibilités et des spécificités basée sur la vraisemblance par le biais de modèles linéaires hiérarchiques généralisés visant à délimiter correctement les diverses corrélations dans les données. La méthodologie est illustrée par l'analyse des données d'une étude sur la rétinopathie diabétique.

1C-C3: Generalized Linear Models, Overdispersion and Zero-Inflation Modèles linéaires généralisés, inflation de la variance et surabondance de zéros

CATHERINE CREA & AYESHA ALI (University of Guelph)

[Monday June 4 / lundi 4 juin, 13:45]

On the Robustness of Dirichlet-Multinomial Regression in the Context of Modeling Pollination Networks

De la robustesse de la régression Dirichlet-multinomiale dans le contexte de la modélisation des réseaux de pollinisation

Recent studies suggest that the structure of plant-pollinator networks is driven by two opposing theories: neutrality and linkage rules. We propose Dirichlet-Multinomial (DM) regression to model plant-pollinator interactions as a function of plant-pollinator characteristics for evaluating the contribution of each process to network structure. This research evaluates the robustness of DM regression to misspecification of dispersion structure, and compares the performance of DM regression to grouped conditional logit (GCL) regression through simulation studies. Results of these studies suggest that DM regression is a robust statistical method for modeling quantitative plant-pollinator networks and outperforms the GCL regression when data are over-dispersed.

Les études récentes suggèrent que la structure des réseaux plante-pollinisateur tient à deux théories contradictoires : la neutralité et les règles de liaison. Nous proposons la régression Dirichlet-multinomiale (DM) pour modéliser les interactions plante-pollinisateur comme fonction des caractéristiques plante-pollinisateur et évaluer la contribution de chaque processus à la structure du réseau. Nous évaluons la robustesse de la régression DM aux erreurs de spécification de la structure de dispersion et en comparons la performance par rapport à la régression logit conditionnelle groupée (LGC) à l'aide de simulations. Les résultats suggèrent que la régression DM est une méthode statistique robuste pour modéliser les réseaux plante-pollinisateur quantitatifs, avec des résultats supérieurs à la régression LGC quand les données sont sur-dispersées.

XUEMAO ZHANG (Lakehead University), **SUDHIR PAUL** (University of Windsor), **DELI LI** (Lakehead University)

[Monday June 4 / lundi 4 juin, 14:00]

Modeling Longitudinal Binomial Responses by Generalized Estimating Equations

Modélisation de réponses binomiales longitudinales par des équations d'estimation généralisées

Extra-binomial variation or overdispersion of longitudinal binomial data is frequently observed in biomedical and observational studies. Current generalized estimating equations (GEE) packages treat the overdispersion parameter as a constant across all subjects. In this talk, a two-parameter variance function modeling the extraneous variance is proposed to account for overdispersion due to heterogeneity between subjects. The new approach allows modeling the extra-binomial variation as a function of the mean.

On observe fréquemment une variation extra-binomiale ou une surdispersion des données binomiales longitudinales dans des études biomédicales et observationnelles. Les librairies actuelles pour les équations d'estimation généralisées (EEG) considèrent le paramètre de la surdispersion comme étant constant chez tous les sujets. Cet exposé propose une fonction de la variance à deux paramètres permettant de modéliser la variance superflue pour tenir compte de la surdispersion attribuable à l'hétérogénéité des sujets. La nouvelle approche permet de modéliser la variation extra-binomiale en fonction de la moyenne.

S. M. KHURSHID ALAM & SUDHIR PAUL (University of Windsor), **XUEMAO ZHANG** (Lakehead University)

[Monday June 4 / lundi 4 juin, 14:15]

Negative Binomial Behrens-Fisher Problem

Problème de Behrens-Fisher pour la binominale négative

Traditional Behrens-Fisher problem is to test the equality of the means μ_1 and μ_2 of two normal

Le problème traditionnel de type Behrens-Fisher consiste à tester l'égalité des moyennes μ_1 et μ_2 de deux populations

1C-C3: Generalized Linear Models, Overdispersion and Zero-Inflation Modèles linéaires généralisés, inflation de la variance et surabondance de zéros

populations where the variances σ_1^2 and σ_2^2 are unknown. The purpose of this paper is to deal with the corresponding problem for over-dispersed count data. We develop six tests and compare these in terms of size and power.

normales où les variances σ_1^2 et σ_2^2 sont inconnues. L'objectif de cet article est de traiter le problème des données de dénombrement surdispersées. Nous avons mis au point six tests et les avons comparés en termes de seuil et de puissance.

CONNIE STEWART (University of New Brunswick), **CHRIS FIELD** (Dalhousie University)

[Monday June 4 / lundi 4 juin, 14:30]

Statistical Methods in Diet Estimation Using Quantitative Fatty Acid Signature Analysis

Méthodes statistiques pour estimer le régime alimentaire via l'analyse quantitative de la signature des acides gras

Quantitative Fatty Signature Analysis produces diet estimates containing the proportion of each species of prey in a predator's diet. Since the estimates are compositional, often contain an abundance of zeros, and samples sizes are generally small, inference problems require the use of non-standard statistical methodology. In this talk, we discuss three recently proposed zero-inflated distributions that may be used to model the diet estimates. Parametric confidence intervals based on these models are compared for real-life seal data. We also consider a nonparametric approach for handling a second inference problem, namely that of testing for a difference in diet.

L'analyse quantitative de la signature des acides gras produit des estimations indiquant la proportion de chaque espèce de proie dans le régime alimentaire d'un prédateur. Comme les estimations sont compositionnelles, elles contiennent souvent une abondance de zéros. Aussi, puisque les tailles d'échantillon sont généralement petites, une méthodologie statistique non standard est requise pour les problèmes d'inférence. Nous discutons de trois distributions à inflation zéro récemment proposées pour modéliser les estimations de régime. Nous comparons les intervalles de confiance paramétriques déduits de ces modèles sur des données réelles concernant les phoques. Nous explorons aussi une approche non paramétrique pour traiter un autre problème d'inférence, à savoir comment établir une différence significative de régime alimentaire.

1C-C4: Methods for High-Dimensional Data/Data Reduction Analyse de données en plusieurs dimensions

Chair/Président: Renjun Ma (University of New Brunswick)

Thornbrough Building, room/local 1307

Abstracts/Résumés

XIAOGANG (STEVEN) WANG (York University)

[Monday June 4 / lundi 4 juin, 13:15]

Dimensionality Reduction for Categorical Data

Réduction de dimensionnalité pour des données catégorielles

We introduce a general class of global association measures for categorical variables. We also propose an efficient method to reduce the dimensionality for high dimensional categorical data. Our method is capable of detecting redundant variables and assessing the collective effect of a group of categorical variables. The method is demonstrated using synthetic and real data sets.

Nous présentons une classe générale de mesures d'association globales pour des variables catégorielles. Nous proposons également une méthode efficace pour réduire la dimensionnalité des données catégorielles de grande dimension. Notre méthode permet de détecter les variables redondantes et d'évaluer l'effet collectif d'un groupe de variables catégorielles. Notre approche est démontrée au moyen de jeux de données synthétiques et réelles.

BABAK ALIPANAHI RAMANDI (University of Waterloo), **NATHAN KRISLOCK** (INRIA Grenoble Rhône-Alpes, France), **ALI GHODSI** (University of Waterloo)

[Monday June 4 / lundi 4 juin, 13:30]

Large-Scale Manifold Learning by Semidefinite Facial Reduction

Apprentissage de variétés à grande échelle par réduction de faces semidéfinie

The problem of nonlinear dimensionality reduction can be formulated as a semidefinite programming (SDP) problem. However, only SDP problems of limited size can be solved directly using current SDP solvers. To overcome this difficulty, we propose a novel SDP formulation for dimensionality reduction based on semidefinite facial reduction that significantly reduces the number of variables and constraints of the SDP problem, allowing us to solve very large dimensionality reduction problems. We also provide a supervised version of our algorithm that incorporates the response variable into the embedding.

Le problème de la réduction non linéaire de la dimension peut se formuler en terme d'un problème de programmation semidéfinie (PSD). Toutefois, seuls les problèmes PSD de taille limitée peuvent se résoudre directement avec les solveurs PSD existants. Pour surmonter cette difficulté, nous proposons une formulation PSD originale de réduction de la dimension fondée sur la réduction de face semidéfinie qui réduit considérablement le nombre de variables et de contraintes du problème PSD, nous permettant de résoudre des problèmes de réduction de très grande dimension. Nous proposons une version supervisée de notre algorithme qui inclut la variable réponse.

SUN MAKOSSO-KALLYTH (Centre de Recherche du Centre Hospitalier Universitaire de Québec)

[Monday June 4 / lundi 4 juin, 13:45]

Symbolic Data Analysis of m Sets of Intervals

Analyse de données symboliques pour m ensembles d'intervalles

1C-C4: Methods for High-Dimensional Data/Data Reduction Analyse de données en plusieurs dimensions

This work presents a new approach to analyze series of m tables of symbolic interval variables. We firstly define a space of intervals with laws of composition. Then, we define a compromise matrix of type intervals, a measure of covariance between interval variables, a new measure of correlation between interval. This way, we achieve a symbolic PCA of compromise. To express the variability of m tables, they are projected on the principal axes of the PCA of intervals of the compromise. For the interpretation, a new measure of correlation will be used.

Ce travail présente une approche nouvelle pour analyser une série de m tableaux contenant des variables de type intervalle. Ces tableaux décrivent les mêmes individus et les mêmes variables. On définit premièrement un espace d'intervalles que l'on munit de lois de composition. On définit ensuite une matrice de compromis de type intervalle et une mesure de covariance et de corrélation entre des intervalles. On obtient ainsi une analyse en composante principale du compromis. Pour analyser simultanément les multiples tableaux d'intervalles, on les projette sur les axes principaux de l'ACP du compromis. Une nouvelle mesure de corrélation est utilisée pour l'interprétation.

ASHLEY BONNER (McMaster University)

[Monday June 4 / lundi 4 juin, 14:00]

Sparse Principal Component Analysis for High-Dimensional Data: A Comparative Study with Applications in Genetics.
Analyse parcimonieuse en composantes principales pour données de grande dimension : une étude comparative appliquée à la génétique

Through unprecedented advances in technology, high-dimensional datasets have exploded into many fields of observational research. Determining important features proves statistically challenging, as multivariate analysis techniques become flooded and mathematically insufficient when $p > n$. Principal Component Analysis (PCA) is commonly used for dimension reduction and data visualization but suffers from these issues. A collection of Sparse PCA methods have been proposed to counter these flaws but should be compared in detail. Through a wide array of simulation scenarios, this study uncovers performance differences amongst three of the latest methods and then uses them to explore gene expression data.

Avec le progrès technologique, les ensembles de données de grande dimension sont en vogue dans divers domaines de recherche par observation. Mais il est statistiquement difficile de déterminer les facteurs importants car les techniques d'analyse multivariée sont mathématiquement insuffisantes lorsque $p > n$. L'analyse en composantes principales (ACP) permet de réduire la dimension et de visualiser les données, mais souffre de ces mêmes problèmes. Diverses méthodes d'ACP parcimonieuses pourraient y remédier. À partir d'une variété de scénarios de simulation, nous détaillons les écarts de performance entre trois méthodes récentes, puis les utilisons pour explorer des données d'expression génétique.

SATHISH CHANDRA PICHKA & JOSEPH BEYENE (McMaster University)

[Monday June 4 / lundi 4 juin, 14:15]

Sparse Canonical Correlation Analysis (SCCA): A Comparative Study
Analyse de corrélation canonique étendue : étude comparative

Canonical Correlation Analysis (CCA) is a multivariate statistical method that can be used to find linear relationship between two datasets. In high-dimensional data where the number of variables in each dataset is very large and sample size relatively small, findings will lack robustness and biological interpretation. In such situations, sparse

L'analyse de corrélation canonique (ACC) est une méthode statistique multivariée permettant de trouver une relation linéaire entre deux ensembles de données. Dans les données de grande dimension où le nombre de variables de chaque ensemble est très grand et la taille d'échantillon relativement réduite, les résultats manquent typiquement de robustesse et sont difficiles à interpréter. Dans ces situa-

1C-C4: Methods for High-Dimensional Data/Data Reduction

Analyse de données en plusieurs dimensions

CCA (SCCA) approaches are preferred. We compared three different SCCA methods using extensive simulations. We also illustrated the methods using real genomic datasets. SCCA methods consistently selected important variables compared to CCA, but the performances of the sparse methods also varied depending on other factors.

tions, des approches ACC étendue sont préférables. Nous comparons trois méthodes d'ACC étendue à l'aide de nombreuses simulations et offrons des illustrations sur de vrais ensembles de données. Les méthodes d'ACC étendue sélectionnent systématiquement les variables importantes par rapport à l'ACC traditionnelle, mais leurs performances varient en fonction d'autres facteurs.

1C-C5: Modern Regression Methods Méthodes modernes de régression

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

Rozanski Hall, room/local 102

Abstracts/Résumés

SONG CAI (University of British Columbia)

[Monday June 4 / lundi 4 juin, 13:15]

Two Fast Algorithms for L^1 -type Estimation in Regression

Deux algorithmes rapides pour l'estimation de type L^1 en régression

Two algorithms are proposed for minimization of L^1 -type objective functions in regression. The first algorithm is based on mixture of coordinate descent and steepest descent steps to partially relax the two problems of coordinate descent for non-smooth functions: convergence to local non-optimal kinks and slow convergence speed. The second algorithm is a relaxation method which finds a “smart” descent direction at each iteration step to avoid possible zigzag convergence phenomenon in coordinate and steepest descent algorithms. Both methods are efficient and work nicely in practice.

Nous proposons deux algorithmes pour minimiser les fonctions objectif de type L^1 en régression. Le premier est fondé sur un mélange de descente de coordonnées et de descente rapide pour relâcher les deux problèmes de la descente de coordonnées pour les fonctions non lisses : convergence vers des nœuds locaux non optimaux et vitesse de convergence lente. Le second est une méthode de relaxation qui trouve une direction de descente “intelligente” à chaque étape d’itération pour éviter le phénomène de convergence en zigzag dans les algorithmes de descente de coordonnées et de descente rapide. Les deux algorithmes sont efficaces et fonctionnent bien.

FUQI CHEN & SÉVÉRIEN NKURUNZIZA (University of Windsor)

[Monday June 4 / lundi 4 juin, 13:30]

Shrinkage Estimation in Some Model Selection by LASSO in Multivariate Regression with Structural Changes

Estimation par rétrécissement dans certains modèles de sélection par la méthode LASSO en régression multivariée avec changements structurels

In this paper, we propose an estimation method in some multivariate regression models with multiple change-points occurring at unknown times. By a LASSO-type method, we estimate the parameter matrix and unknown change points, and we derive the asymptotic properties of the proposed estimators. Also, we consider the case where the parameter is suspected to be restricted to some linear subspaces. Further, we propose a class of shrinkage estimators and establish their asymptotic optimality. Finally, in order to illustrate the efficiency of the proposed method for small sample sizes, we present some simulation results.

Dans cet article, nous proposons une méthode d’estimation dans certains modèles de régressions multivariées avec changements de régimes multiples intervenus à des dates inconnues. Par une méthode de type LASSO, nous estimons le paramètre matriciel ainsi que les temps de changement de régimes et établissons les propriétés asymptotiques des estimateurs proposés. Aussi, nous considérons le scénario où le paramètre pourrait être restreint à un certain sous-espace linéaire. De plus, nous proposons une classe d’estimateurs à rétrécissement et établissons leur optimalité asymptotique. Finalement, afin d’illustrer l’efficacité de la méthode proposée dans les échantillons de petites tailles, nous présentons quelques résultats de simulation.

1C-C5: Modern Regression Methods Méthodes modernes de régression

NICHOLAS MITSAKAKIS (Toronto Health Economics and Technology Assessment, University of Toronto), **GEORGE TOMLINSON** (Department of Health Policy, Management and Evaluation, University of Toronto)

[Monday June 4 / lundi 4 juin, 13:45]

Bayesian Regression Models for Estimation of Disease-Specific Net Costs using Aggregate Data

Modèles de régression bayésiens pour l'estimation des coûts nets de certaines maladies à l'aide de données agrégées

In order to estimate disease specific “net costs” from aggregate level (e.g. sample means and SD per strata) cost data, which are skew and heteroscedastic, we propose and study a Bayesian Gamma regression mixed model that utilizes as stochastic nodes both sample means and inverse coefficients of variation. We investigate its performance and goodness of fit using simulated and real data, and we compare it with two linear models, assuming known and unknown cost variances per stratum. Our results show that, despite its theoretical justification, the benefits of using the Gamma model over the much simpler linear models are questionable.

Pour estimer les coûts nets de différentes maladies à l'aide de données agrégées de coût (par exemple des moyennes échantillonnelles et écarts type par strates) qui sont asymétriques et hétéroscédastiques, nous proposons et étudions un modèle bayésien de régression gamma mixte qui utilise comme noeuds stochastiques les moyennes des échantillons et l'inverse du coefficient de variation. Nous évaluons sa performance et l'adéquation à l'aide de données réelles et simulées et le comparons à deux modèles linéaires, en supposant des variances connues ou non dans chaque strate. Nos résultats montrent que, malgré sa justification théorique, l'avantage du modèle gamma par rapport aux modèles linéaires beaucoup plus simples est discutable.

VANDA LOURENÇO (Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal), **ANA MARIA PIRES** (Instituto Superior Técnico da Universidade Técnica de Lisboa, Portugal)

[Monday June 4 / lundi 4 juin, 14:00]

Robust Outlier Detection in Genetic Association Studies of Quantitative Traits in the Context of M-regression

Détection robuste d'observations aberrantes dans les études d'association génétique avec modèles de régression de type M

Robust multiple linear regression methods are a valuable asset to genetic association studies of quantitative traits, allowing us not to be concerned with the eventuality of outlying observations being present in the data disrupting our analysis results. However, knowledge of these observations may be of the utmost importance to the researcher in order to assess the underlying mechanisms of the data, since they are not always a result of measurement error. To this respect, we propose and discuss a robust outlier test together with an adequate FDR correction measure to be used in the context of a robust multiple-regression model.

Les modèles de régression robuste sont un outil précieux dans les études d'association génétique, nous permettant de ne pas nous soucier de l'effet possiblement dévastateur des observations aberrantes sur les résultats de notre analyse. Cependant, l'identification de ces observations peut être d'une grande valeur pour aider à la compréhension des mécanismes sous-jacents aux données. En effet, les observations aberrantes ne sont pas toujours le résultat d'une erreur de mesure. Donc, nous proposons et discutons un test robuste pour la détection de ce type d'observations avec une mesure de correction qui sera utilisée dans le cadre d'un modèle de régression multiple robuste.

1C-C5: Modern Regression Methods Méthodes modernes de régression

SALIMAH ISMAIL (University of Victoria), **WENQI SUN** (Simon Fraser University), **FAROUK NATHOO & ARIF BABUL** (University of Victoria), **ALEXANDER MOISEEV** (Down Syndrome Research Foundation of British Columbia), **NAZNIN VIRJI-BABUL** (University of British Columbia), **FAISAL BEG** (Simon Fraser University)

[Monday June 4 / lundi 4 juin, 14:15]

A Skew-t Space-varying Regression Model for the Spectral Analysis of Magnetoencephalography Readings

L'analyse spectrale des mesures de magnétoencéphalographie à l'aide d'un modèle de régression à espace variable et un modèle t asymétrique

Characterizing the brain rhythms of individuals with neurological disorders is of fundamental interest in neuroscience. We develop a Bayesian approach for comparing the resting state brain activity of individuals with Down syndrome (DS) with controls. Magnetoencephalography is used to record time-series of neural activity across several brain regions, and the mean-frequency of the power spectral density is computed at each region. We develop a skew-t model for analysis, and use space-varying regression to examine associations across the scalp. Our analysis suggests spectral slowing in the brain rhythms of individuals with DS, and produces smoothed maps illustrating the scalp-topography of differences.

La caractérisation des changements des rythmes neurologiques sont d'un intérêt fondamental en neuroscience. Nous développons une approche bayésienne à l'aide de la régression à espace variable et un modèle t asymétrique pour faire une comparaison de l'activité mentale spontanée entre les individus qui ont le syndrome de Down et les individus du groupe contrôle. Nous utilisons la magnétoencéphalographie (MEG) pour obtenir des séries chronologiques de l'activité cérébrale dans plusieurs régions du cerveau et nous calculons la fréquence moyenne de la densité spectrale de puissance pour chaque région du cerveau. Notre analyse suggère que les individus avec le syndrome de Down ont un ralentissement spectral. Les différences topographiques sont illustrées à l'aide de cartes lisses.

1C-D1: Case Study I: Mark-Recapture on Atlantic Cod Études de cas I: Marquage-recapture de la morue de l'Atlantique

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

Organizers/Responsables: John Bratney (Fisheries and Oceans Canada / Pêches et Océans Canada), Karla Fox (Statistics Canada / Statistique Canada) and/et Lajmi Lakhal-Chaieb (Université Laval)

Science Complex Atrium

Description

Using data from off Eastern and Southern Newfoundland, students can either estimate the cod exploitation rate or attempt to estimate growth curves using capture-recapture data. Capture-recapture surveys were first used in biology and ecology to estimate the size of animal populations. Although this case-study is about cod populations, we note to the students that capture-recapture surveys and their associated models apply in a much wider range of domains, such as: for estimating at risk populations (e.g., homeless people, prostitutes, illegal migrants, drug addicts, etc.); measuring delicate or rare subcategories (as in the U.S. census undercount procedure and the new French census); measuring consumer habits; estimating defaulting companies; measuring fraud detection; and software debugging. As a result, the students working on this case study will gain insight on a useful survey design with a great deal of potential applications.

En utilisant les données sur la morue de l'Atlantique des côtes orientales et méridionales de Terre-Neuve, les étudiants pourront estimer de taux d'exploitation des morues ou essayer d'estimer les courbes de croissance de population au moyen des données de capture-recapture. Les enquêtes par capture-recapture ont d'abord été utilisées en biologie et en écologie afin d'estimer la taille des populations animales. Même si cette étude de cas s'intéresse aux populations de morues, nous informons les étudiants que les enquêtes par capture-recapture et leurs modèles sous-jacents peuvent s'appliquer dans une gamme beaucoup plus vaste de domaines : estimation des populations à risque (p. ex. sans-abris, prostituées, migrants clandestins, toxicomanes, etc.) ; mesurer des sous-catégories rares ou fragiles (comme dans la procédure de sous-dénombrement du recensement aux États-Unis et du nouveau recensement en France) ; mesurer les habitudes des consommateurs ; estimer les entreprises en défaut de paiement ; mesurer la détection des actes frauduleux ; déboguer des logiciels. Ainsi, les étudiants qui font des recherches sur ce type d'étude de cas pourront acquérir davantage de connaissances sur un plan de sondage utile pouvant s'appliquer à de multiples sujets.

AUDREY BÉLIVEAU & ZHENG SUN (Simon Fraser University)

MAGGIE CHEN & YING QI (University of Toronto)

PAUL VAN DAM-BATES, FAN WU & DEAN KOCH (University of Victoria)

TINGTING ZHAO, ANDY LEUNG & YIYANG PAN (University of British Columbia)

1C-D2: Case Study II: Risk Factors for Osteoporotic Fracture Étude de cas II: Facteurs de risque de fracture ostéoporotique

Chair/Président: Lajmi Lakhal Chaieb (Université Laval)

Organizers/Responsables: Lajmi Lakhal Chaieb (Université Laval) and/et Lisa Lix (University of Saskatchewan)

Science Complex Atrium

Description

Osteoporosis is a condition that causes bones to become thin and porous, decreasing bone strength and leading to increased risk of breaking a bone. The most common sites of osteoporotic fracture are the wrist, spine and hip. No single cause for osteoporosis has been identified. The main purpose of this case study is to identify predictors of osteoporotic fracture (i.e., fractures of the hip, wrist, and spine) in men and women after controlling for various risk factors and to determine if there are differences between men and women. This case study uses data from the National Health and Nutrition Examination Survey, 2007-2008 (NHANES 2007-2008), in which 10,149 individuals of all ages were initially contacted for the survey and physical examination. Data were collected between January 2007 and December 2008.

L'ostéoporose est une maladie dans laquelle les os deviennent minces et poreux et perdent de leur résistance, ce qui augmente le risque de fracture. Les sites les plus fréquents de fracture ostéoporotique sont le poignet, les vertèbres et la hanche. Aucune cause unique d'ostéoporose n'a encore été identifiée. L'objectif principal est d'identifier les prédicteurs de fracture ostéoporotique (c.-à-d. de fractures de la hanche, du poignet et des vertèbres) chez les hommes et les femmes après prise en compte de divers facteurs de risque et de déterminer s'il existe des différences entre les hommes et les femmes. La présente étude de cas utilise des données tirées de la National Health and Nutrition Examination Survey, 2007-2008 (NHANES 2007-2008), dans le cadre de laquelle 10 149 individus de tous âges ont été contactés dans un premier temps par questionnaire et examen médical. Les données ont été recueillies entre janvier 2007 et décembre 2008.

JANIE COULOMBE, MARIE-CHRISTINE DOUCAS & KENZA SALLIER (Université de Montréal)

SHU GUO, JIALIN ZOU & LINGLING FAN (University of Toronto)

WEI LIN, XIMING XU, THI LAM UYEN HOANG & VU THIEN HUONG LE (University of Toronto)

VACHAN MISIR, AZADEH TANHA & YANG CHEN (University of Toronto)

YING QI & JINGXIONG XU (University of Toronto)

NAHID SULTANA, MOHAMMAD SHAHIDUL ISLAM, TAGREED ALMALKI & HUDA ALSULAMI (McMaster University)

YANHUA TIAN AND NISHAN MUDALIGE (York University)

DAOXING WANG, XIN LU, XIN CHEN & XUESHAN ZHANG (University of Waterloo)

1C-P: Poster Session I Séance d'affichage I

Science Complex Atrium

Abstracts/Résumés

PRATHAYANA BALENDRAN & HANNA JANKOWSKI (York University)

[Monday June 4 / lundi 4 juin, 13:15]

Estimation of Songbird Flight Paths Using a Bayesian Modelling Framework

Estimation des trajectoires de vol des oiseaux chanteurs au moyen d'un cadre de modélisation bayésien

Geolocators are presently employed to track songbirds by measuring daily sunset and sunrise times, but these produce highly erroneous observations. Using a Bayesian modelling framework we aim to detect the extreme observations produced by such geolocators and estimate the true flight path of the songbirds.

Des géolocalisateurs sont actuellement employés pour effectuer le suivi des oiseaux chanteurs en mesurant quotidiennement les heures de lever et de coucher du soleil, mais ils produisent des observations hautement erronées. À l'aide d'un cadre de modélisation bayésien, nous cherchons à détecter les observations extrêmes produites par de tels géolocalisateurs et à estimer la véritable trajectoire de vol des oiseaux chanteurs.

JESSICA CONNOLLY (University of Guelph)

[Monday June 4 / lundi 4 juin, 13:15]

Assessing Accuracy and Error Trends in an Underwater Acoustic Telemetry System

Évaluation de l'exactitude et des tendances d'erreur en télémétrie acoustique sous-marine

A simulation-based approach is used to quantify position estimate accuracy and trends in error structure in a three-dimensional underwater environment. Time of arrival differences are used in conjunction with multilateration methods to make positional estimates of a signal source. The network studied here is used to examine fine-scale movement pathways of aquatic organisms within a sensor (hydrophone) array, but general findings are easily extended to other sensory networks. The hydrophone geometry, signal degradation related to distance travelled as well as noise within the underwater system are considered as both independent and additive sources of three-dimensional positioning error.

Nous utilisons une approche basée sur la simulation pour quantifier l'exactitude de l'estimation de la position et les tendances dans la structure d'erreur dans un environnement tridimensionnel sous-marin. Les différences entre les temps d'arrivée sont utilisées conjointement avec des méthodes de multilatération pour effectuer des estimations positionnelles d'une source de signal. Le réseau à l'étude est utilisé pour examiner les trajectoires des mouvements à petite échelle d'organismes aquatiques dans une mosaïque de capteurs (hydrophones), mais les résultats généraux sont facilement applicables à d'autres réseaux sensoriels. En matière de géométrie de l'hydrophone, la dégradation du signal liée à la distance parcourue ainsi que le bruit du système sous-marin constituent des sources d'erreur additives et indépendantes pour le positionnement tridimensionnel.

L. SPECTRUM HAN & JOHN BRAUN (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 13:15]

Diurnal Wind Speed Modelling

Modélisation de la vitesse diurne du vent

1C-P: Poster Session I Séance d'affichage I

Wind speed is an important predictor of rate of spread in wildfire. Diurnal forecast models of wind speed currently in use by fire managers are based on models developed for temperature and are fundamentally based on Gaussians and sinusoidal assumptions. Wind speed is often modelled using Weibull or other skewed distributions and the diurnal pattern are often more complicated than a simple sinusoid. This poster presents alternative models, together with diagnostic assessments.

La vitesse du vent est un prédicteur important de la vitesse de propagation des incendies de forêt. Les modèles de prévision diurne de la vitesse du vent actuellement utilisés par les gestionnaires d'incendie sont basés sur des modèles conçus pour la température et s'appuient essentiellement sur des hypothèses gaussiennes et sinusoïdales. La vitesse du vent est souvent modélisée suivant la loi de Weibull ou d'autres distributions asymétriques et la configuration diurne est souvent plus compliquée qu'une simple sinusoïde. Cette affiche présente des modèles alternatifs ainsi que des évaluations diagnostiques.

SHOFIQUIL ISLAM, JOSEPH BEYENE & JEMILA HAMID (McMaster University)

[Monday June 4 / lundi 4 juin, 13:15]

Multivariate Growth Curve Models for the Analysis of Longitudinal Data

Modèles de courbe de croissance multivariés pour l'analyse de données longitudinales

We consider the Growth Curve Model (GCM), a model particularly useful in the analysis of longitudinal data and other response curves. Using extensive simulations, we compared the performance of a new trace test by Hamid et al. (2011) with Potthoff and Roy's (1964), and Khatri's (1966) likelihood ratio tests as well as a test based on the mixed model. Preliminary results indicate that ignoring time trend in multivariate response data, when it exists, leads to loss of power and tests based on the GCM appeared to have better power when responses follow non-linear growth.

Nous considérons le modèle de courbe de croissance (MCC), particulièrement utile pour l'analyse de données longitudinales et d'autres courbes de réponse. Au moyen de simulations approfondies, nous comparons la performance d'un nouveau test de la trace de Hamid et coll. (2011) avec les tests du rapport de vraisemblance de Potthoff et Roy (1964) et de Khatri (1966), ainsi qu'un test basé sur le modèle mixte. Les résultats préliminaires indiquent que le fait d'ignorer la tendance temporelle dans des données de réponse multivariées, lorsqu'elle existe, mène à une perte de puissance et les tests basés sur le MCC semblent présenter une puissance supérieure quand les réponses suivent une croissance non linéaire.

GRACE P.S. KWONG, ZVONIMIR POLJAK, ROB DEARDON & CATE E. DEWEY (University of Guelph)

[Monday June 4 / lundi 4 juin, 13:15]

Bayesian Analysis of PRRS Virus Spread in Ontario Swine Herds

Analyse bayésienne de la propagation du virus SDRP dans les troupeaux porcins de l'Ontario

Individual-level models (ILMs) for infectious diseases, fitted in a Bayesian MCMC framework, have been used to describe both the spatial and temporal spread of diseases. They are an intuitive and flexible class of models that can consider population heterogeneity via various individual-level covariates. The objective of this study is to identify risk factors for the spread of the genotype 1-18-4 (one of the most common PRRS virus genotypes) from southern Ontario swine data using ILMs. Specifically, we explore networks through

Des modèles de niveau individuel (MNI) pour maladies infectieuses sont ajustés dans un cadre bayésien de MCMC afin de décrire la propagation spatiale et temporelle des maladies. Ils constituent une classe de modèles intuitifs et flexibles qui peuvent tenir compte de l'hétérogénéité d'une population par diverses covariables au niveau individuel. L'objectif de cette étude consiste à identifier au moyen de MNI les facteurs de risque de propagation du génotype 1-18-4 (l'un des génotypes les plus courants du virus SDRP) en utilisant des données sur le porc du sud de l'Ontario. Plus précisément, nous examinons les réseaux par lesquels des

1C-P: Poster Session I Séance d'affichage I

which resources are obtained/delivered, the ownership structure of herds, and identify factors that may contribute to high risk of infection.

ressources sont obtenues et offertes, ainsi que la structure de propriété des troupeaux, et nous identifions les facteurs qui peuvent contribuer à un risque élevé d'infection.

WENKAI MA, JOHN BRAUN & GREGORY KOPP (University of Western Ontario), **MIKE WOTTON** (University of Toronto)

[Monday June 4 / lundi 4 juin, 13:15]

Spotfire Distance Distribution Modelling: Designing a Wind Tunnel Experiment

Modélisation de la distribution des distances de propagation des feux disséminés : conception d'un essai en soufflerie

Spotting is a dangerous and unpredictable feature of wildfires where firebrands are sent aloft by the convection column and carried by the wind, often starting new fires. Physical and stochastic spotfire modelling is required to improve fire management practices, in terms of fire crew safety and suppression resource deployment. An important question concerns the effect of the convection column on the flight distance, firebrand landing variation and distribution. Data can be obtained from a wind tunnel experiment where a vertically-oriented fan is used to model the convection column. This poster describes the experience of designing such a wind tunnel experiment.

Propriété dangereuse et imprévisible des incendies de forêt, la dissémination du feu se produit lorsque des tisons montent en altitude dans les colonnes de convection et sont disséminés par le vent, déclenchant souvent de nouveaux foyers d'incendie. La modélisation physique et stochastique des feux disséminés est nécessaire pour améliorer les pratiques de gestion des incendies, notamment la sécurité de l'équipe d'incendie et la mobilisation des ressources pour la lutte contre les incendies. Une question importante concerne l'effet de la colonne de convection sur la distance de dissémination ainsi que la variation et la distribution de la retombée des tisons. Des données sont obtenues par un essai en soufflerie au cours duquel un ventilateur vertical simule une colonne de convection. Cette affiche décrit la conception d'un tel essai en soufflerie.

NAGHAM MOHAMMAD & IAN A. MCLEOD (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 13:15]

Left-Censored Gamma

Gamma censurée à gauche

Analyzing left-censored gamma data has significant applications to environmental sciences and relevant industries where data from research may be incomplete. In general, censoring occurs when the value of observations is only partially known. We are concerned about the left-censored data where a data point is below a certain value but it is unknown by how much. In our study, we intend to find a method recommendation for analyzing left-censored data using simulation. We simulate left-censored data under the assumption that the underlying distribution is gamma. We propose a new model for the left-censored gamma.

L'analyse d'une distribution gamma censurée à gauche a des applications importantes en sciences de l'environnement et dans certaines industries, quand les données de recherche sont incomplètes. En général, la censure apparaît quand la valeur des observations est partiellement inconnue. Nous examinons les données censurées à gauche lorsqu'un point de donnée est inférieur à une certaine valeur par une marge inconnue. Dans cette étude, nous utilisons une simulation pour énoncer des recommandations sur les méthodes d'analyse de données censurées à gauche. Nous simulons des données censurées à gauche selon l'hypothèse que la distribution sous-jacente est de type gamma. Nous proposons un nouveau modèle pour des données gamma censurées à gauche.

1C-P: Poster Session I Séance d'affichage I

MELISSA MORGAN & ASHOK KRISHNAMURTHY (Mount Royal University)

[Monday June 4 / lundi 4 juin, 13:15]

Bayesian Tracking of Epidemic Infectious Diseases using Data Assimilation Methods

Suivi bayésien des épidémies de maladies infectieuses par des méthodes d'assimilation des données

This project explores the use of Bayesian data assimilation methods for the real-time statistical tracking of infectious diseases and to study the spatial dynamics of a disease spreading rapidly from one city to another. To this purpose, we are adapting a family of data assimilation methods known as “Ensemble Kalman Filters” and “Ensemble Optimal Statistical Interpolation”. Bayesian statistical tracking, in real-time, of an infectious disease as it unfolds is illustrated for a simulated epidemic wave originating in Calgary, AB, Canada and spreading towards Edmonton, AB, Canada.

Ce projet explore l'utilisation de méthodes bayésiennes d'assimilation d'information pour faire le suivi des données statistiques en temps réel des maladies infectieuses et pour étudier la dynamique spatiale de la propagation rapide d'une maladie d'une ville à une autre. Dans ce but, nous avons adapté une famille de méthodes d'assimilation de données appelées “filtres de Kalman d'ensemble” et “interpolation statistique optimale d'ensemble”. Le suivi statistique bayésien d'une maladie infectieuse est illustré en temps réel par la simulation d'une épidémie qui a ses origines à Calgary, Alb., Canada et qui se propage vers Edmonton, Alb., Canada.

LORI L. MURRAY & DAVID R. BELLHOUSE (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 13:15]

Early Smoothing Techniques: Halley's 1701 Map of Magnetic Declination

Techniques de lissage précoces : carte de la déclinaison magnétique de 1701 de Halley

Based on data he collected during sea voyages of the HMS Paramore, Edmond Halley published a map in 1701 showing lines of equal magnetic declination. The map has been held up by many as an early, and good, example of statistical graphics. Halley did not reveal the data analytic techniques that he used in his map construction and they remain unknown to this day. Using some mathematical tools of his day, arithmetical averages and Newton's divided difference method to fit a line to data, a plausible method for the map's construction is given.

En se basant sur des données recueillies lors de voyages en mer du HMS Paramore, Edmond Halley a publié une carte en 1701 montrant les lignes de déclinaison magnétique égale. La carte est considérée par plusieurs comme un bon exemple précoce de graphiques statistiques. Halley n'a pas révélé les techniques analytiques qu'il a utilisées pour construire sa carte et elles restent inconnues à ce jour. Nous proposons une méthode plausible de la construction de la carte à l'aide de certains outils mathématiques de son époque : les moyennes arithmétiques et la méthode de différence divisée de Newton pour ajuster une ligne aux données.

PAUL NGUYEN (University of Toronto), **PATRICK BROWN** (Cancer Care Ontario and University of Toronto / Cancer Care Ontario et University of Toronto), **JAMES STAFFORD** (University of Toronto)

[Monday June 4 / lundi 4 juin, 13:15]

Mapping Disease Risk with Changing Census Boundaries

Cartographie du risque de maladie avec des limites de recensement changeantes

Mapping the risk of rare diseases at high spatial resolution often involves working with data aggregated in space and time to obtain meaningful results. We present a local-EM algorithm that per-

La cartographie du risque de maladies rares à haute résolution spatiale nécessite souvent de travailler avec des données agrégées dans l'espace et le temps pour obtenir des résultats significatifs. Nous présentons un algorithme EM lo-

1C-P: Poster Session I Séance d'affichage I

mits multiple maps with different census boundaries to be combined. This algorithm estimates the risk surface on a tessellation where the cumulative intersections of the various maps are distinct regions. Furthermore, we implement parametric bootstrapping procedures to address the uncertainty of these estimates. These methods were demonstrated on mesothelioma lung cancer data collected in the counties of Lambton and Middlesex of southwestern Ontario, Canada between 1985 and 2007.

cal qui permet la combinaison de plusieurs cartes avec des limites de recensement différentes. Cet algorithme estime la surface de risque d'une mosaïque où les intersections cumulatives des diverses cartes sont des régions distinctes. De plus, nous établissons des procédures paramétriques de bootstrapping pour examiner l'incertitude de ces estimateurs. Ces méthodes ont été démontrées au moyen de données sur le mésothéliome, un cancer pulmonaire, recueillies dans les comtés sud-ontariens de Lambton et Middlesex au Canada, entre 1985 et 2007.

1D-I1: Analysis of Interval-censored Data Analyse de données censurées par intervalles

Organizer and Chair / Responsable et président: John Braun (University of Western Ontario)

Alexander Hall, room/local 100

Abstracts/Résumés

JAMES STAFFORD (University of Toronto), **PATRICK BROWN** (Cancer Care Ontario and University of Toronto / Cancer Care Ontario et University of Toronto), **CHUN-PO STEVE FAN & PAUL NGUYEN** (University of Toronto)

[Monday June 4 / lundi 4 juin, 15:15]

Local-EM and Penalized Likelihood

Algorithme EM local et vraisemblance pénalisée

The use of local likelihood methods in the presence of data that is interval or area censored leads naturally to the consideration of local-EM algorithms. Using a piecewise constant density function at the E-step of the algorithm results in it collapsing explicitly into an EMS algorithm (Silverman et al. 1990). This discovery exposes a previously unknown connection between local-EM algorithms and penalized likelihood that is similar to the more familiar pairing of EM and likelihood. We use local-EM to explore the spatial structure of mesothelioma lung cancer risk in Lambton and Middlesex counties.

L'emploi de méthodes de vraisemblance locale en présence de données censurées par intervalle ou par région conduit naturellement à l'algorithme EM local. L'utilisation d'une fonction de densité constante par morceaux à l'étape E de l'algorithme mène à une équivalence explicite avec un algorithme EMS (Silverman et al. 1990). Cette découverte met en lumière un lien auparavant inconnu entre l'algorithme EM local et la vraisemblance pénalisée qui ressemble à la relation plus familière entre l'algorithme EM et la vraisemblance. Nous utilisons l'algorithme EM local pour explorer la structure spatiale du risque de cancer du poumon par mésothéliome dans les comtés de Lambton et de Middlesex.

JOAN HU (Simon Fraser University), **JOHN BRAUN** (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 15:45]

Two Stage Modeling in Event History Analysis, with Application to Forest Fire Control

Modélisation en deux phases de l'historique des événements avec application au contrôle des feux de forêt

Many applications aim at revealing an underlying process of the observed events rather than the event process itself. This motivated us to adapt a two stage modelling: the first stage models the event data conditional on the underlying process, of which the second stage specifies a probability model. This talk presents the two stage modelling via forest fire data. We focus on employing the duration times of observed fires to make inference on the underlying process. Inference procedures are developed to accommodate the event times with interval censoring and to account for the potential spatial-temporal association of the individual fires.

Plusieurs applications cherchent à déceler le processus sous-jacent des données observées plutôt que le processus d'événements. Ceci nous a motivés à adopter une modélisation en deux phases : la première phase modélisant l'historique des événements conditionnelle au processus sous-jacent et la deuxième phase spécifiant un modèle probabiliste. Cet exposé présente la modélisation en deux phases à l'aide de données sur les feux de forêt. Nous nous concentrons sur l'usage de la durée des feux observés pour inférer le processus sous-jacent. Les procédures d'inférence sont élaborées pour répondre aux temps d'événements censurés par intervalles et pour tenir compte de la possible association spatio-temporelle entre les incendies.

1D-I1: Analysis of Interval-censored Data Analyse de données censurées par intervalles

PATRICK BROWN (Cancer Care Ontario and University of Toronto / Cancer Care Ontario et University of Toronto),
PAUL NGUYEN & JAMES STAFFORD (University of Toronto), **LENNON LI** (Public Health Ontario)

[Monday June 4 / lundi 4 juin, 16:15]

Disease Mapping with Spatially Aggregated Data

Cartographie des maladies à l'aide de données spatialement agrégées

Disease incidence data is usually available in spatially aggregated form, such as counts per administrative area. Typically, such data are modelled with spatially discrete processes, where risk in each region is dependant on the risk of neighbours. This approach has the undesirable property that spatial dependence is dictated by the boundaries of the regions, and these methods are not applicable when regions overlap. This talk describes the use of spatially continuous inhomogeneous Poisson processes with aggregated data, using Bayesian inference and the local-EM algorithm. The methods are illustrated by mapping lung cancer incidence in Ontario and syphilis in North Carolina.

Les données sur l'incidence des maladies sont habituellement disponibles en format spatialement agrégées, tel que des comptes par région administrative. Ces données sont généralement modélisées à l'aide de processus spatialement discrets où le risque de chaque région dépend du risque de ses voisins. Pour cette approche, la dépendance spatiale est imposée par les limites des régions, ce qui n'est pas souhaitable, d'autant plus que les méthodes ne s'appliquent pas lorsqu'il y a un chevauchement des régions. Cet exposé décrit l'utilisation des processus de Poisson inhomogènes spatialement continus sur des données agrégées à l'aide de l'inférence bayésienne et de l'algorithme EM local. La cartographie de l'incidence du cancer du poumon en Ontario et de la syphilis en Caroline du Nord sert à illustrer les méthodes.

1D-I2: Celebrating Our 40th Anniversary Célébrons nos 40 ans

Chair/Président: John Brewster (University of Manitoba)

Organizers/Responsables: Christian Genest (McGill University) and/et Alexandre Leblanc
(University of Manitoba)

Rozanski Hall, room/local 104

Abstracts/Résumés

DAVID R. BELLHOUSE (University of Western Ontario), **CHRISTIAN GENEST** (McGill University)

[Monday June 4 / lundi 4 juin, 15:15]

Celebrating our 40th Anniversary

Célébrons nos 40 ans

The Canadian statistical community is holding this year its 40th annual conference. The first meeting took place in Kingston in 1973 in conjunction with the Learned Societies Conference; the defunct Statistical Science Association of Canada sponsored the event. The Statistical Society of Canada (SSC) took up the torch in 1979. At the 25th Annual Meeting, held in Fredericton (New Brunswick), Christian and I described the tumultuous events that led to the creation of the SSC at the end of the 1970s (Bellhouse and Genest 1999, *Statist. Sci.*).

La communauté statistique canadienne tient cette année son 40e congrès annuel. La première rencontre a eu lieu à Kingston en 1973 dans le cadre du congrès des Sociétés savantes; l'événement ayant été parrainé par la défunte Association canadienne de science statistique. La Société statistique du Canada (SSC) a repris le flambeau en 1979. Lors du 25e congrès annuel, tenu à Fredericton (Nouveau-Brunswick), Christian et moi avons décrit les péripéties ayant mené à la création de la SSC à la fin des années 1970 (Bellhouse et Genest 1999, *Statist. Sci.*).

CHRISTIAN GENEST (McGill University), **DAVID R. BELLHOUSE** (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 16:00]

Celebrating our 40th Anniversary

Célébrons nos 40 ans

After offering a short encore performance David and I will recall various people and events that have since helped shape our association, as we know it today. Some facts and figures will also be provided, along with (sometimes amusing) anecdotes. The audience will be quizzed on its knowledge of SSC history and will be invited to contribute its own material.

Après avoir brièvement rappelé les faits marquants de cette époque, David et moi évoquerons divers événements et personnalités qui ont aidé depuis lors à façonner notre association, telle que nous la connaissons aujourd'hui. Quelques faits et chiffres seront aussi présentés, de même que des anecdotes (parfois cocasses). Les connaissances de l'auditoire seront mises à l'épreuve et le public sera invité à enrichir le propos en partageant ses propres observations.

1D-I3: Linear and Nonlinear Time Series Modèles linéaires et non linéaires de séries chronologiques

Organizer and Chair / Responsable et président: Reg Kulperger (University of Western Ontario)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Rozanski Hall, room/local 101

Abstracts/Résumés

DAVID BRILLINGER (University of California, Berkeley)

[Monday June 4 / lundi 4 juin, 15:15]

The Nicholson Blowfly Experiments: Some History and EDA, II

Les expériences avec les mouches de Nicholson : un historique et EDA, II

The entomologist Alexander J. Nicholson carried out experiments to learn more about a sheep pest, the blowfly. The situation involves a system that is nonlinear, has time lags and might be described as nonstationary. The data are aggregate, age effects exist, and measurements are at discrete times while the phenomenon exists in continuous time. Further a structural change may be taking place. The spectrogram and complex demodulation prove to be useful tools. The goals are to explore some of Nicholson's data to illustrate how these tools and subject matter can elicit information.

L'entomologiste Alexander J. Nicholson a mené des expériences visant à en apprendre davantage sur un ravageur des moutons, la mouche à viande. Le contexte comprend un système non-linéaire avec décalages qui peut être qualifié de non stationnaire. Les données sont agrégées, il existe des effets d'âge et les mesures sont prises en temps discret, alors que le phénomène existe en temps continu. De plus, un changement structurel peut survenir. Le spectrogramme et la démodulation complexe s'avèrent des outils utiles. Notre but est d'explorer des données de Nicholson pour illustrer comment ces outils et la connaissance du sujet peuvent fournir de l'information.

WEIBIN JIANG (BMO Financial Group)

[Monday June 4 / lundi 4 juin, 15:45]

Asymptotic Theory of General Multivariate GARCH Models

Théorie asymptotique pour des modèles GARCH multivariés généraux

This work studies the asymptotic theory of the quasi-maximum likelihood estimator (QMLE) for multivariate generalized autoregressive conditional heteroscedasticity (GARCH) models under mild conditions. We give some counterexamples for the parameter identifiability result in Jeantheau (1998) and provide a better necessary and sufficient condition. We prove the ergodicity of the conditional variance process on an application of theorems by Meyn and Tweedie (2009). Under those conditions, the consistency and asymptotic normality of the QMLE can be proved by the standard com-

Ce travail étudie la théorie asymptotique de l'estimateur de maximum de quasi-vraisemblance (EMQV) pour des modèles multivariés d'hétéroscédasticité conditionnelle autorégressive généralisés (GARCH) sous des conditions faibles. Nous donnons quelques contre-exemples du résultat d'identifiabilité des paramètres de Jeantheau (1998) et nous fournissons une meilleure condition nécessaire et suffisante. Nous prouvons l'ergodicité du processus de variance conditionnelle en appliquant les théorèmes de Meyn et Tweedie (2009). Dans ces conditions, la convergence et la normalité asymptotique de l'EMQV peuvent être démontrées au moyen d'un argument de compacité standard et d'un dé-

1D-I3: Linear and Nonlinear Time Series Modèles linéaires et non linéaires de séries chronologiques

pactness argument and Taylor expansion of the score function, under the assumption of finite sixth moment of the observed sequence.

veloppement en série de Taylor de la fonction score, sous l'hypothèse de l'existence du moment d'ordre six.

RAFAL KULIK (University of Ottawa), **PHILIPPE SOULIER** (Université Paris X, France)

[Monday June 4 / lundi 4 juin, 16:15]

Stochastic Volatility Models with Heavy Tails and Long Memory: Structure and Estimation

Modèles de volatilité stochastique à mémoire longue et à ailes lourdes : structure et estimation

We consider long memory stochastic volatility models with heavy tails. We show that the tail empirical process with random levels never suffers from long memory. This is very desirable from a practical point of view, since such a process may be used to construct the Hill estimator of the tail index. Furthermore, we discuss limit theorems for sample covariances. To prove our results we need to establish new results for regularly varying distributions, which may be of independent interest.

Nous considérons les processus de volatilité stochastique à mémoire longue et à ailes lourdes. Nous montrons que le processus empirique caudal à niveaux aléatoires n'est jamais affecté négativement par la longue mémoire du modèle. Cette caractéristique est très enviable d'un point de vue pratique car un tel processus peut servir à construire l'estimateur de Hill de l'indice caudal. Nous discutons également de théorèmes limites pour la covariance empirique. Pour prouver nos résultats, nous devons prouver de nouveaux résultats pour les distributions à variation régulière, ce qui constitue une contribution d'intérêt en soi.

1D-I4: Recent Advances in Small Area Estimation Progrès récents en estimation pour les petits domaines

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

Rozanski Hall, room/local 103

Abstracts/Résumés

GAURI DATTA (University of Georgia), **PETER HALL** (University of Melbourne, Australia), **LILY WANG** (University of Georgia)

[Monday June 4 / lundi 4 juin, 15:15]

Estimation of Small Area Means under Semi-Parametric Measurement Error Model

Estimation des moyennes de petits domaines en fonction d'un modèle d'erreur de mesure semi-paramétrique

Due to resource consideration, model-based approach has been extensively used by the government agencies to produce reliable small area estimates. Existing methods in small area estimation are mostly parametric, and they usually treat the explanatory variables are measured without error. However, explanatory variables are often subject to measurement error. A few authors have addressed the measurement error problem in small area estimation through parametric approach based on the normality assumption. Resulting estimates are usually sensitive to the distributional assumptions. In this talk, we consider a semi-parametric approach to produce reliable point estimates and prediction intervals for small area means.

En raison de considérations pour les ressources, l'approche basée sur un modèle a été largement utilisée par les organismes gouvernementaux pour produire des estimations fiables pour les petits domaines. Les méthodes d'estimation de petits domaines existantes sont principalement paramétriques, et les variables explicatives sont supposées être mesurées sans erreur. Cependant, les variables explicatives sont souvent sujettes à des erreurs de mesure. Quelques auteurs ont corrigé les problèmes d'erreur de mesure dans les estimations pour petits domaines par une approche paramétrique basée sur l'hypothèse de normalité. Les estimateurs en découlant sont généralement sensibles aux hypothèses de distribution. Dans cet exposé, nous considérons une approche semi-paramétrique afin de produire des estimations ponctuelles et des intervalles de prévision fiables pour les moyennes de petits domaines.

JIMING JIANG (University of California, Davis), **THUAN NGUYEN** (Oregon Health and Science University), **J. SUNIL RAO** (University of Miami)

[Monday June 4 / lundi 4 juin, 15:45]

Observed Best Prediction via Nested-Error Regression

Meilleure prévision observée par la régression à erreur emboîtée

We consider the observed best prediction (OBP; Jiang, Nguyen & Rao 2011) for small area estimation under the nested-error regression model, where both the mean and variance functions may be misspecified. We show via a simulation study that the OBP may significantly outperform the empirical best linear unbiased prediction method not just in the overall mean squared prediction error (MSPE) but also in the area-specific MSPE for every one of the small areas. Estimation of the

Nous considérons la meilleure prévision observée (Jiang, Nguyen & Rao 2011) pour une estimation sur des petits domaines en fonction du modèle de régression à erreur emboîtée, dans lequel les fonctions de moyennes et de variance pourraient être mal spécifiées. Grâce à une étude de simulation, nous montrons que la meilleure prévision observée peut considérablement surpasser la méthode de prévision linéaire sans biais empirique, non seulement pour l'erreur quadratique moyenne de prévision (EQMP) globale, mais également pour l'EQMP spécifique aux petits domaines sur

1D-I4: Recent Advances in Small Area Estimation Progrès récents en estimation pour les petits domaines

area-specific MSPE of OBP is considered. We introduce a method for obtaining functional expressions, called partial derivation (PD), that avoids tedious derivations.

chacun d'entre eux. On considère l'estimation de l'EQMP de la meilleure prévision observée pour chaque domaine. Nous présentons une méthode pour obtenir des expressions fonctionnelles, appelées dérivées partielles, qui permettent d'éviter des calculs fastidieux.

PARTHA LAHIRI (University of Maryland), **SANTANU PRAMANIK** (NORC, University of Chicago)

[Monday June 4 / lundi 4 juin, 16:15]

Estimation of Average Design-Based Mean Squared Error of Synthetic Small Area Estimators

Estimation de la moyenne des écarts quadratiques moyens selon le plan d'échantillonnage d'estimateurs synthétiques pour petits domaines

Many survey organizations use synthetic methods to produce small area estimates because of their simplicity and wide applicability. In this paper, we propose a new design-based average mean square error (AMSE) estimator of a set of synthetic estimators of small area means. Using data analyses and Monte Carlo simulations, we compare our proposed AMSE estimator with two other AMSE estimators available in the literature. The results suggest that the proposed AMSE estimator is a sensible choice.

De nombreuses firmes de sondage utilisent des méthodes synthétiques pour produire des estimations pour petits domaines en raison de leur simplicité et de leur vaste champ d'applications. Dans cet exposé, nous proposons un nouvel estimateur basé sur le plan d'échantillonnage de la moyenne des écarts quadratiques moyens d'un ensemble d'estimateurs synthétiques pour les moyennes de petits domaines. En utilisant des analyses de données et des simulations de Monte Carlo, nous comparons l'estimateur proposé à deux autres estimateurs du même type existant dans la littérature. Les résultats suggèrent que l'estimateur proposé est un bon choix.

1D-15: Statistical Methods for Clinical Trials and Causal Inference Méthodes statistiques pour les essais cliniques et l'inférence causale

Organizer and Chair / Responsable et président: Lawrence McCandless (Simon Fraser University)

Alexander Hall, room/local 200

Abstracts/Résumés

JULI ATHERTON (Université du Québec à Montréal), **DAVID WOLFSON** (McGill University), **EDENS THADDEUS**
[Monday June 4 / lundi 4 juin, 15:15]

An Overview for Bayesian Optimal Design for Clinical Trials with an Unknown Delay to Treatment Effect
Survol de la planification optimale bayésienne d'essais cliniques avec un délai inconnu à l'effet du traitement

In a clinical trial it is rare that the intervention effect will be immediately measurable. The literature on the design of clinical trials has not addressed this issue of “delay-to-effect”. Multipath changepoint models provide a flexible framework for optimal design for trials that entail longitudinal follow-up, and where a delay to effect is anticipated. We shall first introduce the multipath changepoint model and show how a Bayesian approach to optimal design allows for subject-specific delays to effect as well as non-response. Beginning with the simplest setting, we shall discuss various extensions and difficulties, including some challenging computational hurdles.

Dans un essai clinique, il est rare que l'effet de l'intervention soit immédiatement mesurable. La littérature sur la planification des essais cliniques n'a, jusqu'à maintenant, pas abordé cette question de « délai à l'effet ». Les modèles de point de rupture à trajets multiples fournissent un cadre flexible pour les essais qui ont un suivi longitudinal, et où un délai est prévu. Nous présenterons le modèle et montrerons comment une approche bayésienne pour la planification optimale prend en compte des délais à l'effet et la non-réponse. En commençant par une situation simple, nous discuterons de diverses extensions et de quelques difficultés.

GENEVIÈVE LEFEBVRE & JULI ATHERTON (Université du Québec à Montréal)

[Monday June 4 / lundi 4 juin, 15:45]

On the Role of the Prior Distribution in the “Bayesian Adjustment for Confounding” Algorithm
Sur le rôle de la loi a priori de l'algorithme “Bayesian Adjustment for Confounding”

A methodology for Bayesian model averaging in a causal inference framework, the Bayesian Adjustment for Confounding (BAC) algorithm, has recently been suggested by Wang et al. (2011). BAC tweaks model averaging by favouring models for which the treatment effect of interest has a causal interpretation. This is made possible by considering an outcome and an exposure model, and a prior distribution that controls the probability of inclusion of the covariates in these two models. In this presentation, we will shed further light on the prior and show that it may have a considerable effect on the estimation.

Une méthodologie de moyennage de modèles bayésien en inférence causale, l'algorithme Bayesian Adjustment for Confounding (BAC), a récemment été suggérée par Wang et coll. (2011). BAC altère la stratégie de moyennage régulière en favorisant les modèles qui ont une interprétation causale. Ceci est rendu possible par la considération d'un modèle de réponse et de traitement, et d'une spécification a priori qui contrôle l'inclusion des covariables dans ces deux modèles. Dans cette présentation, nous allons effectuer un retour sur le rôle de la loi a priori et montrer que celle-ci peut avoir un impact considérable sur l'estimation.

1D-I5: Statistical Methods for Clinical Trials and Causal Inference Méthodes statistiques pour les essais cliniques et l'inférence causale

SAMAN MUTHUKUMARANA (University of Manitoba)

[Monday June 4 / lundi 4 juin, 16:15]

Assessing Noninferiority in Multi Arm Clinical Trials

Analyse de non-infériorité dans des essais cliniques comportant plusieurs groupes

Recently, there has been a growing interest in clinical trials to demonstrate whether a new treatment is not worse than an active control by more than a specified margin. This is known as two-arm noninferiority trial and has been discussed in numerous articles. We develop generalized p-value and generalized confidence intervals procedures for the noninferiority analysis. Two generalized test functions are constructed and Monte Carlo simulations are used to compute the p-value. A Bayesian meta-analysis approach for multi arm trials will also be discussed. This is based in part on joint work with Mark Gamalo, Pulak Ghosh and Ram Tiwari.

Récemment, un intérêt croissant a été porté aux essais cliniques permettant de démontrer qu'un contrôle actif n'est pas mieux qu'un nouveau traitement par plus d'une marge spécifiée. Ces essais de non-infériorité à deux groupes ont été discutés dans de nombreux articles. Nous créons une valeur p et des procédures d'intervalles de confiance généralisées pour les analyses de non-infériorité. Deux fonctions de test généralisées sont construites et des simulations de Monte Carlo sont utilisées pour calculer la valeur p. Nous discuterons également d'une approche bayésienne de méta-analyse pour des essais comportant plusieurs groupes. Cet article est partiellement basé sur des travaux conjoints avec Mark Gamalo, Pulak Ghosh et Ram Tiwari.

1D-C1: Density Approximation and Estimation Approximation et estimation de densité

Chair/Président: Mikelis Bickis (University of Saskatchewan)

Rozanski Hall, room/local 105

Abstracts/Résumés

ANAS RAMADAN & ALAN KER (University of Guelph)

[Monday June 4 / lundi 4 juin, 15:15]

Empirical Bayes Nonparametric Density Estimation of Crop Yield Densities: Rating of Crop Insurance Estimation non paramétrique bayésienne empirique de la densité des rendements des cultures

I examine a newly proposed density estimator in order to evaluate its usefulness for government crop insurance programs confronted by the problem of adverse selection. While the Federal Crop Insurance Corporation (FCIC) offers multiple insurance programs, what is needed is a more accurate method of estimating actuarially fair premium rates in order to eliminate adverse selection. The Empirical Bayes Nonparametric Kernel Density Estimator (EBNKDE) showed a substantial efficiency gain in estimating crop yield densities, as opposed to other estimators. The data of 97 Illinois counties' corn yield from 1955 to 2010 was taken from the National Agricultural Statistics Service (NASS).

J'examine un nouvel estimateur de la densité pour évaluer l'utilité dans les programmes d'assurance-récolte confrontés au problème de l'antisélection. Le Federal Crop Insurance Corporation (FCIC) offre divers programmes d'assurance, mais il ne dispose d'aucune méthode efficace pour estimer des primes actuariellement équitables et éliminer ainsi l'antisélection. L'estimateur non paramétrique bayésien empirique obtenu par la méthode du noyau de la densité (EBNKDE) est nettement plus efficace pour estimer la densité des rendements des cultures que les autres estimateurs. Les données sur le rendement des cultures de maïs de 97 comtés de l'Illinois entre 1955 et 2010 proviennent du National Agricultural Statistics Service (NASS).

SHYAM DADIMUNI (Wilfrid Laurier University)

[Monday June 4 / lundi 4 juin, 15:30]

Large Forest Fire Risk Density Estimation Estimation de la densité du risque de grands feux de forêt

Spatial forest fire risk mapping is of interest to fire management, especially for large fire events. Kernel Density Estimation is used to estimate a risk map in Ontario, Canada, using historical records on 1420 large forest fires that occurred between 1959 and 1999. The Hall, Sheather, Jones and Marron plug-in method is used to select the optimal bandwidth. The resulting density shows strong spatial heterogeneity with Northwestern Ontario having the highest risk.

La cartographie des risques de feux de forêt est d'intérêt pour la gestion des incendies, en particulier pour les grands incendies. L'estimation de la densité par noyau est utilisée pour estimer une carte du risque en Ontario (Canada) à l'aide de données historiques sur 1420 grands feux de forêts ayant eu lieu entre 1959 et 1999. L'estimateur de remplacement de Sheather, Jones et Marron est utilisé pour choisir la fenêtre optimale. La densité résultante montre une forte hétérogénéité, le nord-ouest de l'Ontario ayant le risque le plus élevé.

SERGE PROVOST (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 15:45]

Polynomial Representations of Differentiated Logdensity Approximants Sur l'approximation de la dérivée du logarithme d'une densité de probabilité par polynômes

1D-C1: Density Approximation and Estimation Approximation et estimation de densité

A density approximation methodology, whereby the derivative of the logarithm of a density approximant is expressed as a polynomial, is being proposed. An explicit solution is provided for the polynomial coefficients, which are determined by a moment-matching technique. The existence and unicity of this solution is established assuming that the distribution under consideration is uniquely determined by its moments. An extension of this approach, wherein differentiated logdensity approximants are represented as ratios of polynomials, shall be discussed as well. All the resulting approximants turn out to be bona fide density functions. Several illustrative examples will be presented.

Nous proposons une méthodologie pour approximer une densité de probabilité selon laquelle la première dérivée du logarithme de l'approximation est représentée par un polynôme. Les coefficients polynomiaux sont explicitement obtenus en résolvant un système d'équations linéaires qui est basé sur les moments de la distribution approximée. L'existence et l'unicité de cette solution sont démontrées sous l'hypothèse que cette distribution est uniquement spécifiée par ses moments. Le cas où les approximations sont représentées par un rapport de polynômes est également considéré. Chaque densité de probabilité ainsi obtenue est en fait bona fide. Plusieurs exemples seront présentés à titre illustratif.

ALIAKBAR MOHSENIPOUR & SERGE PROVOST (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 16:00]

The Distribution of Quadratic forms in Elliptically Contoured Vectors

La distribution des formes quadratiques pour des vecteurs aléatoires elliptiques

It is shown in this talk that a certain scale mixture representation of the density function of a central and non-central elliptically contoured vector yield computable expressions for the moments, the density and the distribution function of quadratic forms in elliptically contoured vectors. The same approach can readily be used to determine the distribution of other function of elliptically contoured vectors in terms of their Gaussian counterparts. A method of approximating the distribution function of quadratic forms in elliptically contoured vectors is considered as well.

Il est montré dans cette présentation qu'une certaine représentation de la fonction de densité d'un vecteur aléatoire elliptique centré et non-centré permet des expressions calculables pour les moments, la densité et la fonction de répartition. La même approche peut directement être utilisée pour déterminer la distribution d'une autre fonction des vecteurs elliptiques en termes d'équivalences gaussiennes. Une méthode pour estimer la fonction de répartition des formes quadratiques des vecteurs elliptiques est aussi considérée.

SUSAN ZHE SHENG (University of Western Ontario)

[Monday June 4 / lundi 4 juin, 16:15]

Polynomially Adjusted Saddlepoint Approximations

Approximations des points de selle avec ajustement polynomial

Given the cumulant-generating function of a random variable, saddlepoint density approximations are evaluated over its support and an interpolating spline is determined. This function is then normalized and adjusted by means of a polynomial of degree n whose coefficients are such that the first n moments of the resulting approximation agree with those of the distribution being considered. Several illustrative examples, including the case of mix-

Étant donné la fonction génératrice des cumulants d'une variable aléatoire, les approximations de densité des points de selle sont évaluées sur leur support et une fonction spline d'interpolation est déterminée. Cette fonction est ensuite normalisée et ajustée par un polynôme de degré n dont les coefficients sont tels que les n premiers moments de l'approximation résultante concordent avec ceux de la distribution étudiée. Divers exemples, dont des mélanges de fonctions de densité beta, gamma et Gaussiennes, indiquent

1D-C1: Density Approximation and Estimation Approximation et estimation de densité

tures of beta, gamma and Gaussian density functions, indicate that the adjustment consistently provides more accurate approximations over the entire range of the distribution.

que l'ajustement offre toujours des approximations plus précises concernant l'étendue de la distribution.

1D-C2: Impact and Comparison of Cancer Screening Strategies Impact et comparaison des examens préventifs pour le cancer

Chair/Présidente: Wendy Lou (University of Toronto)

MacNaughton Building, room/local 113

Abstracts/Résumés

CECILIA COTTON & KRITHIKA SURESH (University of Waterloo), **DAVID HODGSON** (University of Toronto), **CRYSTAL HANN & LINDA MAYHEW** (Princess Margaret Hospital)

[Monday June 4 / lundi 4 juin, 15:15]

Breast Cancer Screening in Female Survivors of Childhood Hodgkin Lymphoma: A Simulation Study

Dépistage du cancer du sein chez les survivantes du lymphome hodgkinien de l'enfant : étude de simulation

Female survivors of childhood Hodgkin Lymphoma (HL) who received chest radiation therapy are at high risk of developing breast cancer in early to mid adulthood. Thus, beginning annual MRI and/or mammogram screening at a younger age might improve early-stage breast cancer detection and reduce breast cancer mortality. We present a simulation model of 100,000 women diagnosed with HL at age 15 years and followed from age 25 to 50. The natural history of breast cancer is simulated using an existing stochastic model, various screening strategies are considered and the benefits and costs are compared.

Les survivantes du lymphome hodgkinien (LH) qui ont reçu une radiothérapie thoracique présentent un risque élevé de cancer du sein au début et au milieu de l'âge adulte. Un IRM et/ou mammogramme annuel précoce pourrait donc améliorer la détection précoce et réduire la mortalité du cancer du sein. Nous présentons un modèle de simulation de 100 000 femmes atteintes de LH à 15 ans et suivies de 25 à 50 ans. L'évolution naturelle du cancer du sein est simulée par un modèle stochastique existant. Diverses stratégies de dépistage sont comparées selon leurs avantages et leurs coûts.

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

[Monday June 4 / lundi 4 juin, 15:30]

Measuring Impact of Cancer Screening without Data-Holders Having to Share Trial-Specific Data

Mesure des répercussions du dépistage du cancer sans que les détenteurs de données doivent partager des informations spécifiques au sujet de leur essai clinique

We have developed a 3-parameter model describing the mortality reductions produced by one round of screening. Parameter estimation requires the counts of deaths in each Lexis rectangle in each screening trial (the corresponding design matrices are known). Data-holders are reluctant to share data even at this aggregate level. We propose that each data-holder send us the 3-D grid of log-likelihood values generated by the counts from their trial. The summed log-likelihood functions allow for parameter-estimation, without data-holders having to disclose their data, or the complexity of the DataSHIELD approach.

Nous avons conçu un modèle à trois paramètres décrivant la réduction de mortalité produite par un cycle de dépistage. L'estimation de paramètres nécessite le nombre de décès observés dans chaque rectangle de Lexis, pour chaque essai de dépistage (les matrices de plan d'expérience correspondantes sont connues). Les détenteurs de données sont réticents à partager leurs données, même à un tel niveau d'agrégation. Nous proposons que chaque détenteur de données nous envoie une grille tridimensionnelle des valeurs de la log-vraisemblance générées par les résultats de leurs essais. Le total des fonctions de log-vraisemblance permet une estimation des paramètres, sans que les détenteurs n'aient eu à dévoiler leurs données, et sans la complexité de l'approche "DataSHIELD".

1D-C2: Impact and Comparison of Cancer Screening Strategies Impact et comparaison des examens préventifs pour le cancer

AMIDU OLALEKAN RAIFU & STEPHEN D. WALTER (McMaster University), **EDUARDO FRANCO & RAMANA-KUMAR AGNIHOTRAM** (McGill University)

[Monday June 4 / lundi 4 juin, 15:45]

Effects of Covariates on the Performance of Cervical Cancer Screening Tests: Logistic Regression and Latent Class Models

Effets des covariables sur la performance des tests de dépistage du cancer du col de l'utérus : régression logistique et modèles de catégorie latente

To investigate how covariates affect the performance of two cervical screening tests - visual inspection with acetic acid (VIA) and lugol iodine (VILI) tests. The sensitivity and specificity of VIA and VILI tests were estimated empirically and using an extension of a logistic model proposed by Coughlin et al. (1992). Latent class models and Kappa statistic were also applied. Age, parity, and HPV DNA influenced the performance of the two tests. The false positive and false negative error rates for the tests measured by physician and nurse were different. By dichotomizing the outcomes, there was agreement between the raters measurements.

Nous étudions comment les covariables affectent la performance de deux tests de dépistage du cancer du col de l'utérus, à savoir l'inspection visuelle avec acide acétique (IVA) et le soluté de Lugol (IVL). La sensibilité et la spécificité des tests IVA et IVL est estimée empiriquement et via l'extension d'un modèle logistique proposé par Coughlin et coll. (1992). Des modèles de catégorie latente et des statistiques Kappa sont aussi appliqués. L'âge, la parité et l'ADN de VPH influencent la performance des deux tests. Les taux de faux positifs et faux négatifs des tests mesurés par des médecins et des infirmières sont différents. En dichotomisant les résultats, les mesures des évaluateurs concordent.

ZHIHUI (AMY) LIU, ERIN STRUMPF, NANDINI DENDUKURI, MAURICE MCGREGOR & JAMES A. HANLEY (McGill University)

[Monday June 4 / lundi 4 juin, 16:00]

Measuring the Mortality Reductions Produced by Breast Cancer Screening

Mesurer la réduction de mortalité due au dépistage du cancer du sein

Current guidelines on breast cancer screening are based on data-analyses that ignore key principles of cancer screening. Consequently they underestimate the mortality reductions that would result from a sustained program. We formulate a statistical model that characterizes the delayed impact of each round of screening. A reanalysis of the year-specific data from five commonly cited screening trials suggests that a program of regular mammography screening, from age 50 to 69, would result in yearly breast cancer mortality reductions of at least 40%, rather than the approximately 20% reported by the 2011 Canadian Task Force (and other) meta-analyses.

Les lignes directrices pour le dépistage du cancer du sein sont basées sur des analyses qui ignorent des principes clés de la détection du cancer. Ces analyses sous-estiment en conséquence les réductions de mortalité qui seraient causées par un programme soutenu. Nous formulons un modèle statistique qui caractérise l'impact différé de chaque ronde de dépistage. Une ré-analyse des données annuelles de cinq essais cliniques souvent cités suggère qu'un programme de dépistage par mammographie entre 50 et 69 ans résulterait en une réduction de la mortalité annuelle d'au moins 40 %, au lieu du 20 % habituellement publié par le Groupe d'étude canadien 2011, ainsi que par d'autres méta-analyses.

1D-C3: Monte Carlo Methods and Inference for Stochastic Processes Méthodes de Monte Carlo et inférence pour les processus stochastiques

Chair/Président: Rob Deardon (University of Guelph)

Rozanski Hall, room/local 102

Abstracts/Résumés

CHRISTIANE LEMIEUX (University of Waterloo)

[Monday June 4 / lundi 4 juin, 15:15]

A Study of Quasi-Monte Carlo Methods Via Dependence Concepts

Concepts de dépendence et méthodes quasi-Monte Carlo

Quasi-Monte Carlo methods are multidimensional numerical integration methods that often provide more accurate estimators than the naive Monte Carlo method. The constructions underpinning these methods are designed to provide a form of structured sampling that can exploit certain characteristics of the integrand under study. These methods are typically studied using function decompositions, for example based on Fourier, Walsh or Haar series. In this talk, we propose an alternative way of studying these methods based on dependence concepts such as those introduced by Lehmann in 1966. This allows us to gain some new insight on these methods.

Les méthodes quasi-Monte Carlo sont des méthodes d'intégration numérique multidimensionnelles qui fournissent souvent des estimateurs plus précis que la méthode de Monte Carlo usuelle. Les constructions à la base de ces méthodes sont définies de façon à produire une forme d'échantillonnage structurée qui peut exploiter certaines caractéristiques de la fonction à intégrer. Ces méthodes sont habituellement étudiées en utilisant des décompositions en séries de Fourier, Walsh ou Haar. Dans cette présentation, nous proposons une façon différente d'étudier ces méthodes, basée sur les concepts de dépendence tels qu'introduits par Lehmann. Cette approche alternative nous permet d'apporter un nouvel éclairage sur ces méthodes.

JOURDAN GOLD, ROB DEARDON & ZENY FENG (University of Guelph)

[Monday June 4 / lundi 4 juin, 15:30]

An Investigation of Multi-stage MCMC Approaches for Sampling from Correlated, Discrete Target Distributions

Une étude de différentes approches dans la méthode de Monte Carlo par chaîne de Markov multi-étapes pour la simulation de distributions discrètes corrélées

An investigation of the effectiveness of various MCMC algorithms for sampling from highly correlated, discrete target distributions will be discussed. The relative effectiveness of various multi-stage MCMC approaches, including algorithms based upon random-walk and independence samplers, as well as hybrid combinations of these, will be considered. Kullback-Liebler divergence and effective sample size will be used to assess algorithm quality in this comparison process.

On discutera d'une étude sur l'efficacité des divers algorithmes de type Monte Carlo par chaîne de Markov pour l'échantillonnage de distributions discrètes très corrélées. On considérera l'efficacité relative de diverses approches multi-étapes, y compris les algorithmes basés sur l'indépendance ou sur une marche aléatoire, ainsi que des combinaisons hybrides de ceux-ci. La divergence de Kullback-Liebler et la taille équivalente de l'échantillon seront utilisées pour analyser la qualité de l'algorithme dans ce processus de comparaison.

1D-C3: Monte Carlo Methods and Inference for Stochastic Processes Méthodes de Monte Carlo et inférence pour les processus stochastiques

LOTFI KHRIBI & MARC FREDETTE (HEC Montréal), **BRENDA MACGIBBON** (Université du Québec à Montréal)

[Monday June 4 / lundi 4 juin, 15:45]

The Poisson Maximum Entropy Model for Homogeneous Poisson Processes

Application du principe d'entropie maximale dans la prédiction des évènements récurrents dans le cas des processus de Poisson

We suggest a Bayesian model with the maximum entropy prior distribution to predict the number of future events for subjects already under observation. The intensity function used to model these events is the one corresponding to a homogeneous Poisson process with unknown parameter rates. The prior distribution for these unknown rates is obtained by maximizing the entropy subject to the condition that the first two moments equal the empirical ones. We find from a simulation study and from a warranty dataset from the automobile industry that the maximum entropy prior is preferable to the noninformative Jeffreys prior.

Nous suggérons un modèle bayésien afin de prévoir le nombre futur d'événements pour des sujets déjà sous observation où la distribution a priori est obtenue par le principe d'entropie maximale. La fonction d'intensité que nous emploierons pour modéliser ces événements correspond à un processus de Poisson homogène où les taux sont des paramètres. La loi a priori est obtenue en maximisant l'entropie sujette à la condition que les deux premiers moments théoriques égalent leurs équivalents empiriques. Nous avons trouvé, à l'aide d'une étude de simulation, que la loi a priori d'entropie maximale est préférable à la loi a priori de Jeffreys.

LARISSA VALMY & JEAN VAILLANT (Université des Antilles et de la Guyane)

[Monday June 4 / lundi 4 juin, 16:00]

Statistical Inference for Point Processes Associated with a Dirichlet Process

Inférence statistique pour des processus ponctuels associés à un processus de Dirichlet

We consider events of a spatio-temporal point process which intensity is associated with a hidden process. Both of the following situations are studied. Firstly, the data consist of maps of new occurrence spatial positions between two consecutive observation dates. Secondly, we have count data in spatial units from a systematic random sampling. MCMC techniques are developed for statistical inference on a Cox process partially directed by a Dirichlet process.

Nous nous intéressons aux occurrences d'un processus ponctuel spatio-temporel dont l'intensité est associée à un processus caché. Deux situations sont considérées : d'abord, nous disposons à chaque date d'observation de toutes les nouvelles positions des occurrences du processus depuis la précédente date d'observation sans information sur les dates précises d'occurrence ; ensuite, nous disposons de comptages dans des unités spatiales obtenues par échantillonnage aléatoire systématique. Une approche inférentielle bayésienne utilisant des algorithmes MCMC est proposée dans les deux cas quand le processus observé est un processus de Cox dirigé partiellement par un processus de Dirichlet.

LIWEN ZOU (North Carolina State University)

[Monday June 4 / lundi 4 juin, 16:15]

Fitting Nonstationary General-time-reversible Models to Obtain Edge-lengths and Frequencies for the Barry-Hartigan Model

Ajustement de modèles temporels réversibles généraux non stationnaires pour l'obtention des fréquences et longueurs des frontières dans le modèle Barry-Hartigan

1D-C3: Monte Carlo Methods and Inference for Stochastic Processes Méthodes de Monte Carlo et inférence pour les processus stochastiques

The Barry and Hartigan (BH) model is very flexible. Due to an identifiability problem, the parameters of the BH model cannot be expected to consistently estimate the actual pairwise frequencies. We define a nonstationary GTR (NSGTR) model for each edge and fit the NSGTR model by minimizing the distance between the estimates of transition probabilities under the NSGTR and BH models. With the best-fitting NSGTR estimates, the internal node frequency vector is interpretable as well as edge-length estimates that are otherwise not yielded by the BH model. These edge-lengths are interpretable as the expected number of substitutions along edges.

Le modèle de Barry-Hartigan (BH) est très flexible. En raison d'un problème d'identifiabilité, nous ne pouvons pas espérer que les paramètres du modèle BH mènent à un estimateur convergent des fréquences actuelles par paire. Nous définissons un modèle temporel réversible général non stationnaire (TRGNS) pour chaque frontière et nous ajustons ce modèle en minimisant la distance entre les estimations des probabilités de transition sous les modèles TRGNS et BH. Avec les estimateurs les mieux ajustés au modèle TRGNS, il est possible d'interpréter le vecteur de fréquence du nœud interne de même que les estimateurs des longueurs de frontières, qui ne sont pas produits par le modèle BH. Ces longueurs de frontières peuvent être interprétées comme le nombre attendu de substitutions le long des frontières.

ANNALIZA MCGILLIVRAY & ABBAS KHALILI (McGill University)

[Monday June 4 / lundi 4 juin, 16:30]

A Penalized Quasi-Likelihood Approach for Estimating the Number of States in a Hidden Markov Model

Une approche de la quasi-vraisemblance pénalisée pour estimer le nombre d'états dans une chaîne de Markov cachée

In applications of hidden Markov models (HMMs), one may have no knowledge of the number of states (or order) of the model needed to represent the underlying process of the data. We present a penalized quasi-likelihood method for order estimation in HMMs which utilizes the fact that the marginal distribution of the HMM observations is a finite mixture. The method starts with a HMM with a large number of states and obtains a model of lower order by clustering and merging states through two penalty functions. Its performance is assessed theoretically, via simulation, and with the help of a real data application.

Dans les applications des chaînes de Markov cachées (CMC), il se peut que le statisticien n'ait pas d'information sur le nombre d'états (ou ordre) nécessaires pour représenter le processus. Nous présentons une méthode de la quasi-vraisemblance pénalisée pour estimer l'ordre dans des CMC. Cette méthode utilise le fait que la distribution marginale des observations CMC est un mélange fini. La méthode débute avec un CMC avec un grand nombre d'états et obtient un modèle d'ordre inférieur en regroupant et fusionnant les états à l'aide de deux fonctions de pénalité. La performance est évaluée théoriquement, via des simulations, et à l'aide d'une application sur un vrai jeu de données.

1D-C4: Optimal Design Planification optimale d'expériences

Chair/Président: François Pageau (General Dynamics Produits de défense et Systèmes tactiques - Canada Inc.)

Thornbrough Building, room/local 1307

Abstracts/Résumés

YIN LI & KEUMHEE CHOUGH CARRIERE (University of Alberta)

[Monday June 4 / lundi 4 juin, 15:15]

Optimal Designs in the N of 1 Trials

Plans optimaux pour essais N de 1

The N of 1 trial is a randomized experiment in a single patient which is designed to study the response of an individual to the treatments and assist to make personal best clinical decision. To compare two treatments, A and B, a popular design in N of 1 trials is a single sequence which consists of AB pairs and BA pairs. In this study, we consider models with carryover effects. The efficiency of designs are compared and the optimal N of 1 trials are presented under the models.

Un essai N de 1 est une expérience randomisée avec un seul patient dont le but est d'étudier la réponse de cet individu aux traitements pour aider à prendre une meilleure décision clinique pour celui-ci. Pour comparer deux traitements A et B, un plan populaire pour l'essai N de 1 est une seule séquence qui implique les paires AB et BA. Dans cette étude, nous considérons des modèles avec effets résiduels de traitement. L'efficacité des plans est comparée et les essais N de 1 optimaux pour ce modèle sont présentés.

SAUMEN MANDAL & CHAO ZHU (University of Manitoba)

[Monday June 4 / lundi 4 juin, 15:30]

Some Iterative Procedures for Constructing Optimal Designs

Des processus itératifs pour la construction de plans optimaux

We consider the problem of computing designs which optimize some design criteria subject to the basic constraints on design weights. Examples of such problems include parameter estimation, adaptive design and stratified sampling. In order to construct the optimal designs we use a class of algorithms, indexed by a function which depends on the derivatives of the criterion function. These algorithms possess some suitable properties. We first investigate the performance of the algorithm by constructing D-optimal designs. We then explore the properties of partial and directional derivatives of the criterion function and develop strategies for improving convergence rates of the algorithms.

Nous considérons le problème de la construction de plans qui optimisent un critère de plan sujet à des contraintes de base sur les poids du plan. Des exemples de ce type de problème incluent l'estimation de paramètres, les plans adaptatifs et l'échantillonnage stratifié. Pour construire des plans optimaux, nous utilisons une classe d'algorithmes indexés par une fonction qui dépend des dérivées de la fonction de critère. Ces algorithmes ont des propriétés avantageuses. Nous étudions d'abord la performance de l'algorithme en construisant des plans D-optimaux. Puis, nous explorons les propriétés des dérivées partielles et directionnelles de la fonction de critère et développons des stratégies pour améliorer la vitesse de convergence des algorithmes.

1D-C4: Optimal Design Planification optimale d'expériences

YANQING YI & YUAN YUAN (Memorial University of Newfoundland)

[Monday June 4 / lundi 4 juin, 15:45]

An Optimal Randomization Rule for Response Adaptive Designs

Règle de randomisation optimale pour les plans adaptatifs à la réponse

Response adaptive designs have the ethical advantage of allocating more patients to the potential treatment. However, the adaptation of treatment allocation may reduce statistical power. We propose an optimal adaptive randomization rule and compare this rule with the existing randomization rules in literature. The proposed randomization rule is demonstrated in theory and through simulation to have the property of assigning higher proportion of patients to the better treatment. Simulation results indicate that the statistical power under this rule is similar as that under the Neyman allocation.

Les plans adaptatifs à la réponse présentent l'avantage éthique d'affecter plus de patients au traitement potentiel. Cependant, l'adaptation de l'affectation au traitement peut réduire la puissance de la procédure. Nous proposons une règle de randomisation adaptative optimale et nous la comparons aux règles de randomisation existantes dans la littérature. Il est démontré de façon théorique et par des simulations que la règle proposée permet d'affecter une proportion plus élevée de patients au meilleur traitement. Les résultats des simulations indiquent que la puissance avec cette règle est similaire à celle de la méthode d'affectation de Neyman.

XUAN LI & XIKUI WANG (University of Manitoba)

[Monday June 4 / lundi 4 juin, 16:00]

Variance-penalized Response-adaptive Randomization with Mismeasurement

Randomisation adaptée à la réponse avec pénalisation de la variance et erreur de mesure

We consider response-adaptive design of clinical trials under a variance-penalized criterion in the presence of mismeasurement. An explicit expression for the variance-penalized criterion with misclassified dichotomous responses is derived for response-adaptive design and some properties are discussed. A new target proportion of treatment allocation is proposed under the criterion and related simulation results are presented.

Nous considérons un plan pour essais cliniques adapté à la réponse avec un critère de pénalisation de la variance et en présence d'erreur de mesure. Une expression explicite pour le critère de pénalisation de la variance avec réponses binaires et erreur de mesure est développée pour le plan adapté à la réponse, et certaines propriétés sont présentées. Une nouvelle proportion cible pour l'allocation au traitement est proposée sous ce critère et des résultats de simulation sont présentés.

1D-C5: Survival Analysis and Accelerated Life Testing Analyse de survie et tests accélérés de durée de vie

Chair/Président: Christian Boudreau (University of Waterloo)

MacNaughton Building, room/local 105

Abstracts/Résumés

AMMAR SARHAN & JOSEPH APALOO (St. Francis Xavier University)

[Monday June 4 / lundi 4 juin, 15:15]

Exponentiated Modified Weibull Extension Distribution

La distribution de Weibull modifiée généralisée et exponentiée

A new modified Weibull extension distribution is proposed by Xie et al. (2002). El-Gohary et al. (2012) proposed a new distribution referred to as the generalized Gompertz distribution. In this paper, we propose a new model of a life time distribution that mainly generalizes these two distributions, in addition to many others. We refer to this new distribution as the exponentiated modified Weibull extension (EMWE) distribution. Estimation of its parameters are studied. Real datasets are analyzed using the new distribution, which show that the EMWE distribution can be used quite effectively in fitting and analyzing real lifetime data.

Xie et coll. (2002) proposent une nouvelle généralisation de la distribution de Weibull modifiée, alors que El-Gohary et coll. (2012) proposent une distribution de Gompertz généralisée. Nous présentons la distribution de Weibull modifiée généralisée et exponentiée (EWME) pour la distribution de durées de vie qui généralise ces deux lois, et bien d'autres d'ailleurs. Nous étudions quelques estimateurs de ses paramètres et l'utilisons pour analyser des ensembles de données réelles. Nos résultats montrent que la distribution EWME permet d'ajuster et d'analyser efficacement des données de durée de vie.

DR. PETER ADAMIC & SYLVAIN CARON (Laurentian University)

[Monday June 4 / lundi 4 juin, 15:30]

A Model for Time-Dependent Competing Risks

Un modèle sujet aux risques concurrents lié au temps

In this presentation, we consider generalizations of Turnbull's classic Self-Consistent Algorithms for both interval-censored as well as double-censored data, and show how they can be applied in a competing risks environment. Recent work in this area has led to the development of very flexible time-dependent models, where each competing risk is allowed to have its own unique range of possible failure times for each individual failure. A case study involving the modeling of breast cancer mortality rates will also be furnished to illustrate the theory.

Dans cette présentation, nous allons discuter de la généralisation des algorithmes cohérents de Turnbull en montrant comment les algorithmes de censure par intervalle et double censurés peuvent être adaptés aux risques concurrents. Des recherches récentes ont permis de créer des modèles très flexibles qui nous permettent de modifier les intervalles de temps de chaque échec dans n'importe quel risque. Un exemple portant sur la mortalité de femmes affectées par le cancer du sein sera présenté pour illustrer la théorie.

1D-C5: Survival Analysis and Accelerated Life Testing Analyse de survie et tests accélérés de durée de vie

SCOTT HUNT & XIAOJIAN XU (Brock University)

[Monday June 4 / lundi 4 juin, 15:45]

Robust Designs for an Accelerated Life Testing Model with Simple Step-Stress Plans

Plans d'expériences robustes pour un modèle d'essai de vie accéléré avec un plan simple en contraintes échelonnées

An accelerated life testing model with a simple step-stress plan is presented to obtain the optimum hold time for changing the stress level. A Weibull failure time distribution is assumed at any constant stress level. The scale parameter of the distribution is assumed to be a log-linear function of the stress level. Robust designs are obtained to protect against possible departure from the constant shape parameter, and possible misspecification of the life-stress relationship for both censored and complete data. The experimental test minimizes both the asymptotic squared bias and the asymptotic mean squared error of the reliability estimate.

Un modèle d'essai de vie accéléré est présenté pour obtenir le temps de maintien optimal pour le changement du niveau de stress. Une distribution de défaillance de Weibull est supposée pour tout niveau de stress constant. Le paramètre d'échelle de la distribution est supposé être une fonction log-linéaire du niveau de stress. Des plans robustes sont obtenus pour protéger contre une dérogation à l'hypothèse d'un paramètre d'échelle constant et d'erreurs de spécification de la relation entre la survie et le stress pour les données complètes et censurées. Le test empirique minimise le biais au carré et l'erreur moyenne au carré asymptotiques par rapport à la valeur estimée de la fiabilité.

LONGLONG HUANG (University of Calgary), **KAREN KOPCIUK** (Alberta Health Services), **XUEWEN LU** (University of Calgary)

[Monday June 4 / lundi 4 juin, 16:00]

Jackknife Empirical Likelihood for Smoothed Weighted Rank Regression with Censored Data

Vraisemblance empirique avec jackknife pour régression lisse pondérée basée sur les rangs avec données censurées

A smoothed weighted rank estimating function is used to estimate the regression parameter vector in an accelerated failure time model with right censored data. However, since the inference is based on a normal approximation approach, it may suffer poor performances when sample sizes are small. To increase inference accuracy and robustness, we propose a jackknife empirical likelihood method. The jackknife empirical likelihood ratio is shown to be a standard Chi-squared statistic. Simulations are conducted to compare the proposed method with the normal approximation based method. The new method gives better coverage probabilities and smaller confidence regions for small samples.

Une fonction d'estimation lisse, pondérée et basée sur les rangs est utilisée pour estimer un vecteur de paramètres pour une régression dans un modèle accéléré de données de survie avec données censurées à droite. Toutefois, puisque l'inférence est basée sur une approximation gaussienne, les performances peuvent être faibles pour de petits échantillons. Pour améliorer l'exactitude et la robustesse, nous proposons une méthode jackknife pour la vraisemblance empirique. Nous prouvons que le ratio de vraisemblance empirique du jackknife est une statistique du chi-carré. Des simulations comparent la méthode proposée avec la méthode d'approximation gaussienne. La nouvelle méthode a une meilleure couverture et de plus petites régions de confiance pour de petits échantillons.

2A-A: Gold Medal Address
Allocution du récipiendaire de la médaille d'or

Organizer and Chair / Responsable et président: Roman Viveros-Aguilera (McMaster University)

Rozanski Hall, room/local 104

Abstract/Résumé

CHRISTIAN GENEST (McGill University)

[Tuesday June 5 / mardi 5 juin, 8:45]

How to Model Dependence if You Must

Comment modéliser la dépendance s'il le faut

Modeling dependence between random variables is a central issue in statistics. A survey of the copula approach to this problem will be presented. Copula model construction, selection, validation, interpretation, and use for prediction purposes will be considered. The key role of rank-based inference methods for the study of dependence between continuous variables will be highlighted. The emerging role of copulas in extreme-value theory and time series analysis will also be described. Data from actuarial science, finance, and hydrology will be used for illustration purposes.

La modélisation de la dépendance entre aléas est une préoccupation centrale en statistique. On verra comment l'approche par copules permet d'aborder ce problème. On décrira des méthodes de construction, de sélection, de validation et d'interprétation de modèles de copules, ainsi que leur usage à des fins prédictives. On mettra en lumière le rôle clef des méthodes d'inférence fondées sur les rangs pour l'étude de la dépendance entre des aléas continus. On décrira aussi l'importance croissante des copules en théorie des valeurs extrêmes et en séries chronologiques. Le propos sera illustré au moyen de données actuarielles, financières et hydrologiques.

2B-V: An Overview of Models and Methods for Spatio-temporal Data Analysis Un survol des modèles et méthodes d'analyse de données spatio-temporelle

Organizer and Chair / Responsable et président: Alexandre Leblanc (University of Manitoba)

Rozanski Hall, room/local 104

Abstract/Résumé

JIM ZIDEK (University of British Columbia)

[Tuesday June 5 / mardi 5 juin, 10:15]

An Overview of Models and Methods for Spatio-temporal Data Analysis

Un survol des modèles et méthodes d'analyse pour données spatio-temporelles

This presentation will describe the elements of spatio-temporal analysis for the three major types of processes, geostatistical, lattice and point, on which spatio-temporal methods are based, with an emphasis on first. These types cover the different types of data encountered in practice and their different statistical frameworks will be described. In particular the very successful Bayesian hierarchical approach, which can be used in all these areas, will be introduced. Examples will be presented along with emerging frontiers. The presentation will be given at a fairly non-technical level and it should be accessible to anyone with a reasonable working knowledge of statistics, gained either through theory or practice.

Cette présentation décrira les éléments de l'analyse spatio-temporelle pour les trois types de processus sur lesquels la plupart des méthodes spatio-temporelles sont fondées : géostatistique, treillis et point, en mettant l'accent sur le premier cas. Ces trois types couvrent les différents types de données observées en pratique et le cadre statistique associé à chacun sera décrit. Plus particulièrement, nous présenterons la très fructueuse approche bayésienne hiérarchique qui peut être utilisée pour tous ces cas. Des exemples seront présentés, ainsi que des frontières émergentes. La présentation sera à un faible niveau technique et devrait être accessible à quiconque ayant une connaissance pratique raisonnable des statistiques, obtenue théoriquement ou par la pratique.

2B-I1: Applications of Statistics in Medicine Applications de la statistique à la médecine

Organizer and Chair / Responsable et présidente: Juli Atherton (Université du Québec à Montréal)

Session sponsored by the Committee on New Investigators
Séance parrainée par le Comité des nouveaux chercheurs

Rozanski Hall, room/local 101

Abstracts/Résumés

JIGUO CAO (Simon Fraser University), **WEI LUO**, **MARCUS GALLAGHER & JANET WILES** (University of Queensland, Australia)

[Tuesday June 5 / mardi 5 juin, 10:15]

Estimating the Intensity of Ward Admission and its Effect on Emergency Department Access Block

Estimation de l'intensité des admissions en unité de soins et son effet sur le blocage d'accès aux services d'urgence

Emergency department access block is an urgent problem faced by many public hospitals today. When access block occurs, patients in need of acute care cannot be transferred from the emergency room to inpatient wards within an optimal time frame. We study the effect of admission intensity on access block, using data from a metropolitan hospital in Australia. Several modern statistical methods in functional data analysis are applied to analyze the data. New insights have been gained from this novel application of functional data analysis to health services research.

Le blocage d'accès aux services d'urgence constitue un problème crucial pour plusieurs hôpitaux publics aujourd'hui. Lorsqu'un blocage d'accès survient, les patients qui nécessitent des soins de courte durée ne peuvent pas être transférés de la salle d'urgence aux unités de patients hospitalisés dans un délai optimal. Nous étudions l'effet de l'intensité des admissions sur le blocage d'accès à l'aide de données d'un hôpital métropolitain d'Australie. L'analyse des données fait appel à plusieurs méthodes statistiques modernes d'analyse de données fonctionnelles. Cette nouvelle application d'analyse de données fonctionnelles a jeté un nouvel éclairage sur la recherche dans le domaine des services de santé.

ABBAS KHALILI, YETING DU, JOHANNA NEŠLEHOVÁ & RUSSELL STEELE (McGill University)

[Tuesday June 5 / mardi 5 juin, 10:45]

The Use of Mixture of Linear Mixed Effect Models in identifying Clinical Characteristics of Systemic Sclerosis Disease

L'utilisation d'un mélange de modèles linéaires à effets mixtes pour l'identification des caractéristiques cliniques de la sclérose systémique

The focus of this talk is a dataset from patients enrolled in the Canadian Scleroderma Research Group registry between 2003 and 2010. It is a longitudinal study that aims to identify the clinical characteristics of systemic sclerosis disease that best relate to the patients' health status and quality of life. A finite mixture of linear mixed effect (FMLME) models is used to account for both within-subject correlation and between-subject heterogeneity in

Cet exposé est axé sur un jeu de données de patients inscrits au registre du Groupe de recherche canadien sur la sclérodémie (GRCS) entre 2003 et 2010. Cette étude longitudinale cherche à identifier les caractéristiques cliniques de la sclérose systémique qui sont les plus reliées à la condition de santé des patients et à leur qualité de vie. Un mélange fini de modèles linéaires à effets mixtes (MFMLEM) est utilisé pour tenir compte de la corrélation entre les sujets et de l'hétérogénéité entre les sujets dans l'analyse de

2B-I1: Applications of Statistics in Medicine Applications de la statistique à la médecine

analyzing such data. A new penalized likelihood approach is proposed to fit an FMLME model to this data that allows us to investigate the aforementioned scientific question of interest.

telles données. Nous proposons une nouvelle approche de vraisemblance pénalisée pour ajuster un MFMLEM à ces données qui nous permet d'étudier la question scientifique d'intérêt susmentionnée.

MAHMOUD TORABI (University of Manitoba)

[Tuesday June 5 / mardi 5 juin, 11:15]

Likelihood Inference in Spatio-Temporal Modeling

Inférence à l'aide de la vraisemblance en modélisation spatio-temporelle

The analysis of incidence rates over space and time has received considerable attention because of the desire to provide reliable maps of disease rates. The Poisson mixed model is commonly used for the analysis of disease cases. In some applications, spatial and temporal count data may contain excess zeros. We propose a zero-inflated model to capture true rates over space and time, and reduce the variations yielded from excess zeros. We use data cloning, which yields maximum likelihood, to conduct frequentist analysis of zero-inflated models. The performance of the proposed approach is evaluated with a real dataset and a simulation study.

L'analyse des taux d'incidence dans l'espace et dans le temps a bénéficié d'une attention considérable en raison de la volonté de fournir des cartes fiables des taux de maladies. Le modèle mixte de Poisson est fréquemment utilisé pour l'analyse de cas de maladie. Dans certaines applications, les données de dénombrement spatiales et temporelles peuvent contenir une surabondance de zéros. Nous proposons un modèle avec surreprésentation de zéros pour déterminer les véritables taux dans l'espace et dans le temps et réduire les variations provenant de la surabondance de zéros. Nous utilisons une approche de clonage de données, qui mène au maximum de vraisemblance, pour réaliser une analyse fréquentiste de modèles à surreprésentation de zéros. Nous évaluons la performance de l'approche proposée à l'aide d'un jeu de données réelles et d'une étude par simulation.

2B-I2: On Assessing Liking and Intensity in Sensory Science Évaluation du goût et de l'intensité en sciences sensorielles

Chair/Président: Paul McNicholas (University of Guelph)

Organizers/Responsables: Paul McNicholas (University of Guelph) and/et 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

Alexander Hall, room/local 200

Abstracts/Résumés

CHRIS FINDLAY (Compusense Inc.)

[Tuesday June 5 / mardi 5 juin, 10:15]

You Know What You Like, But What About Everyone Else? A Case Study on Incomplete Block Segmentation of White-Bread Consumers.

Vous savez ce que vous aimez, mais qu'en est-il des autres? Une étude de cas sur la segmentation en blocs incomplets de consommateurs de pain blanc.

“One man’s meat is another man’s poison.” There will always be a wide range of consumer liking response across any product category. Cluster analysis can provide consumer segments based upon common liking that reflect underlying sensory preferences. To determine valid population segments requires a large sample of consumers. As the number of products tested by each consumer increases, experimental bias degrades the data quality. A limited number of well-designed incomplete sample sets can provide cleaner responses. The challenge in this study was to collect 200 valid consumer responses for each of 12 commercial white breads whilst minimizing fatigue.

« Tous les goûts sont dans la nature. » Il y aura toujours une grande diversité de réaction des consommateurs face à plusieurs produits d’une même catégorie. L’analyse de regroupement peut identifier des segments de consommateurs ayant une inclinaison commune sous-jacente qui reflète leurs préférences. Il faut de grands échantillons de consommateurs pour obtenir une segmentation valide de la population. Plus le nombre de produits testés par chaque consommateur est grand, plus la qualité des données décroît, à cause de biais expérimentaux. Un nombre limité de tests sur des échantillons incomplets, mais bien conçus, de produits peut fournir de meilleures données. Le défi de notre étude était ainsi de recueillir des réponses valides pour 200 consommateurs pour 12 marques de pains blancs, tout en minimisant la fatigue des répondants.

RYAN BROWNE & PAUL MCNICHOLAS (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 10:35]

*Design and Analysis of Incomplete Block Designs
Conception et analyse de plans en blocs incomplets*

We provide results from a case study on segmentation of white-bread consumers from an incomplete block design under the assumption of underlying latent factors. We assess ways to improve efficiency of the incomplete block design. Several

Nous présentons les résultats d’une étude de cas sur la segmentation des consommateurs de pain blanc menée à partir d’un plan en blocs incomplets selon l’hypothèse de facteurs latents sous-jacents. Nous évaluons des manières d’améliorer l’efficacité du plan en blocs incomplets. Plusieurs ques-

2B-I2: On Assessing Liking and Intensity in Sensory Science Évaluation du goût et de l'intensité en sciences sensorielles

questions can be proffered but they depend on our criteria; these include, for example, the effect on clustering and effect on variability of the estimators. Using either of these criteria, we wish to assess the effect of changing the block size and the number of treatments. We can quantify the effect of 'contrasting' when building a design versus using a standard incomplete block design.

tions peuvent être soulevées, mais elles dépendent de nos critères, qui comprennent notamment l'effet sur les grappes et l'effet sur la variabilité des estimateurs. En utilisant l'un de ces critères, nous souhaitons évaluer l'effet d'un changement de la taille des blocs et du nombre de traitements. Nous pouvons quantifier « l'effet de contraste » en concevant un plan d'expérience par rapport à un plan en blocs incomplets classique.

JOHN CASTURA (Compusense Inc.)

[Tuesday June 5 / mardi 5 juin, 10:55]

Sweet and Powerful: A Two-attribute Time Intensity Study of Fruit Liqueur

Doux et puissant : une étude de deux attributs de l'intensité temporelle de la liqueur de fruits

Time intensity (TI) methods are used to measure perceived changes in the intensity of one, or sometimes two, sensory attributes over time in a product or stimuli. Evaluations are performed by a panel of well-trained human assessors. Sweetness taste and alcohol flavour were evaluated independently in a 4-product studies on fruity liqueurs and a 4-product study on ready-to-drink coolers, each using a testing protocol established according to business objectives. Data were represented as TI curves, with the goal of capturing the temporal sensory information about the products. Outcomes informed subsequent product development and re-formulation decisions.

Les mesures d'intensité temporelle (IT) sont utilisées pour analyser les changements observés dans les attributs au fil du temps. Les évaluations sont réalisées par un panel. Le goût sucré et la saveur d'alcool ont été évalués isolément dans deux études de quatre produits chacune : une qui traitait des liqueurs fruitées, l'autre qui concernait les boissons rafraîchissantes. Dans le but d'illustrer les données temporelles et sensorielles, nous avons visualisé les résultats en formes de courbe d'IT. Les résultats se sont avérés utiles pour guider le développement subséquent de produits et et les décisions de modifier leur formulation.

MOYI LI, RYAN BROWNE & PAUL MCNICHOLAS (University of Guelph), **JOHN CASTURA** (Compusense Inc.)

[Tuesday June 5 / mardi 5 juin, 11:15]

Analysis of Time Intensity Data: a Fruit Liqueur Study

Analyse de données temporelles d'intensité : étude d'une liqueur de fruits

Using statistical methods we analyze the temporal sensory characteristics of two fruit liqueurs and two ready-to-drink coolers, using data from 13 panelists. We apply EM algorithm to estimate the mean and standard deviation at each time point for each panellist. A simulation study is conducted to test the performance of the estimation method. A fitted time intensity curve exhibits signature pattern for each panellist. The variation across panellists and the comparison between two products are discussed.

Nous analysons les caractéristiques sensorielles temporelles de deux liqueurs de fruits et deux boissons prêtes-à-boire à l'aide de méthodes statistiques et des données de 13 panélistes. Nous utilisons l'algorithme EM pour estimer la moyenne et l'écart-type à chaque temps et pour chacun des panélistes. Une étude par simulation est menée pour évaluer la performance de la méthode d'estimation. Une courbe temporelle d'intensité ajustée montre le motif caractéristique de chaque panéliste. Nous discutons de la variation entre les panélistes et de la comparaison de deux produits.

2B-I3: Outlier-Robust Estimation in Survey Sampling Estimation robuste et méthodes d'échantillonnage

Chair/Présidente: Julie Trépanier (Statistics Canada / Statistique Canada)

Organizer/Responsable: Jean-François Beaumont (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Alexander Hall, room/local 100

Abstracts/Résumés

MALAY GOSH (University of Florida)

[Tuesday June 5 / mardi 5 juin, 10:15]

Outlier Adjusted Robust Bayes Estimators in Survey Sampling

Estimateurs bayésiens robustes ajustés pour les valeurs aberrantes pour les méthodes d'enquête

The paper proposes a general class of Bayes estimators in survey sampling which are either exact or asymptotically design-unbiased. Further an adjustment is proposed to these Bayes estimators to protect against possible outliers.

Cet article propose une classe générale d'estimateurs bayésiens pour données d'enquêtes qui sont soit exactement, soit asymptotiquement sans biais par rapport au plan. Un ajustement des estimateurs est aussi proposé pour protéger contre les valeurs aberrantes.

DAVID HAZIZA (Université de Montréal), **JEAN-FRANÇOIS BEAUMONT** (Statistics Canada / Statistique Canada),
ANNE RUIZ-GAZEN (Université de Toulouse)

[Tuesday June 5 / mardi 5 juin, 10:45]

A Unified Approach to Robust Estimation in Finite Population Sampling

Une approche unifiée d'estimation robuste en échantillonnage

In classical statistics, the influence function has an approximate relationship with the conditional bias of an observation. This suggests that the conditional bias might be a useful measure of influence since it can be easily extended to any inferential framework unlike the influence function. We use the conditional bias to derive robust estimators in finite population sampling. We show that our approach leads to estimators closely related to those proposed in Chambers (1986) and Kocic and Bell (1994). For design-based inference, our approach possesses the desirable feature of handling arbitrary sampling designs. Results of a limited simulation study are presented.

La fonction d'influence en statistique classique est étroitement liée au biais conditionnel d'une observation. Cela suggère que le biais conditionnel pourrait être une mesure d'influence utile puisqu'il peut être facilement dérivé pour toute approche inférentielle, contrairement à la fonction d'influence. Nous utilisons le biais conditionnel pour construire des estimateurs robustes en échantillonnage. Nous montrons que nous obtenons des estimateurs étroitement liés à ceux de Chambers (1986) et Kocic et Bell (1994). Pour l'inférence fondée sur le plan de sondage, notre approche possède la caractéristique désirable d'être applicable pour des plans arbitraires. Nous présentons des résultats d'une étude par simulation.

2B-I3: Outlier-Robust Estimation in Survey Sampling Estimation robuste et méthodes d'échantillonnage

NIKOLAS TZAVIDIS (University of Southampton, U.K.), **SABINE KRIEG** (Netherlands Central Bureau of Statistics), **CHIARA BOCCI** (IRPET and University of Florence, Italy / IRPET et University of Florence, Italy), **MARC SMEETS & VIRGINIE RAYMOND-BLAESS** (Netherlands Central Bureau of Statistics)

[Tuesday June 5 / mardi 5 juin, 11:15]

Outlier Robust Small Domain Estimation for Business Surveys in the Netherlands

Estimation robuste aux valeurs aberrantes pour petits domaines dans les enquêtes auprès des entreprises aux Pays-Bas

The Netherlands Central Bureau of Statistics recently investigated the use of small domain estimation methods in business surveys with emphasis on outlier robust methodologies. In this talk we discuss some of the approaches we considered for this purpose. In particular, we present outlier robust domain estimation using random effects and M-quantile models as well as robust regression synthetic and GREG-type estimators. Attention is paid to the use of design weights, their role in protecting against model misspecification and their relation to influence weights. The alternative methodologies are applied to data from one industry sector of the Dutch Structural Business Survey.

Le Netherlands Central Bureau of Statistics a récemment étudié l'utilisation de méthodes d'estimation pour petits domaines dans les enquêtes auprès des entreprises en mettant l'accent sur les méthodologies robustes aux valeurs aberrantes. Dans cet exposé, nous discutons quelques-unes des approches considérées à cette fin. En particulier, nous présentons une méthode d'estimation robuste aux valeurs aberrantes des domaines utilisant des effets aléatoires et les modèles M-quantile, ainsi que les estimateurs de régression robustes synthétiques et de type GREG. Une attention est accordée à l'utilisation de poids déterminés par le plan d'échantillonnage, leurs rôles dans la protection contre les mauvaises spécifications du modèle et leurs relations avec les poids d'influence. Ces méthodologies alternatives sont appliquées aux données d'un secteur industriel du Dutch Structural Business Survey.

2B-I4: Trend Analysis of Time Series Data and Applications Analyse de tendances en séries chronologiques et applications

Chair/Présidente: Debbie Dupuis (HEC Montréal)
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

Rozanski Hall, room/local 103

Abstracts/Résumés

XIN GAO (York University), **GRACE Y. YI** (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 10:15]

Simultaneous Model Selection and Estimation for Mean and Association Structures with Clustered Binary Data
Sélection de modèle et estimation simultanés pour la moyenne et les structures d'associations avec données binaires regroupées

This talk investigates the property of the penalized estimating equations when both the mean and association structures are modelled. To select variables for the mean and association structures sequentially, we propose a hierarchical penalized generalized estimating equations (HPGEE2) approach. The hierarchical approach is designed to accommodate model constraints relating the inclusion of covariates into the mean and the association models. This two-step penalization strategy enjoys a compelling advantage of easing computational burdens compared to solving the two sets of penalized equations simultaneously. HPGEE2 with SCAD penalty is shown to have the oracle property for the mean and association models.

Cet exposé explore les propriétés des équations d'estimation pénalisées pour modéliser la moyenne et des structures d'association. Pour choisir les variables pour la moyenne et les structures d'association de façon séquentielle, nous proposons une approche d'équations d'estimation généralisées hiérarchique (HPGEE2). L'approche hiérarchique est conçue pour accommoder les contraintes du modèle reliant l'inclusion de covariables dans les modèles pour la moyenne et les associations. Cette stratégie de pénalisation en deux étapes a l'avantage de faciliter les calculs comparée à la résolution simultanée des deux ensembles d'équations pénalisées. Nous montrons que HPGEE2 avec pénalité SCAD a la propriété d'oracle pour la moyenne et les modèles d'association.

YULIA GEL (University of Waterloo), **ABDEL EL SHAARAWI** (National Water Research Institute), **VYACHESLAV LYUBCHICH** (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 10:45]

Nonparametric Tests to Detect Parametric Patterns in Time Series with ARCH Effects
Tests non-paramétriques pour détecter des tendances paramétriques dans les séries chronologiques avec effets ARCH

While trend detection traditionally refers only to detecting monotonic trends, nowadays there exists an increasing interest in diagnostic tools for general (non)-monotonic forms of trends, especially in studies related to climate change, finance and en-

Alors que la détection de tendances fait habituellement uniquement référence à la détection de tendances monotones, il y a un intérêt croissant pour les outils diagnostiques de tendances générales de forme (non)-monotones, particulièrement dans les études sur les changements climatiques, la

2B-14: Trend Analysis of Time Series Data and Applications Analyse de tendances en séries chronologiques et applications

vironmental sciences, to name only a few. In this talk we present a sieve bootstrap-based ANOVA-type nonparametric test for assessing parametric assumptions of trends in conditionally heteroscedastic time series. We discuss asymptotic properties and finite-sample performance of the test and provide a comparative analysis with other trend tests, using simulations and case studies.

finance et les sciences environnementales pour n'en nommer que quelques-unes. Dans cet exposé, nous présentons un test non-paramétrique de type ANOVA basé sur le bootstrap pour évaluer les hypothèses paramétriques des tendances de séries chronologiques conditionnellement hétéroscédastiques. Nous discutons les propriétés asymptotiques et la performance du test pour des échantillons finis et les comparons analytiquement à celles d'autres séries chronologiques à l'aide de simulations et d'études de cas.

YING ZHANG & PAUL CABILIO (Acadia University)

[Tuesday June 5 / mardi 5 juin, 11:15]

A Generalized Jonckheere Test for Testing Trend in Repeated Measures in Randomized Blocks and Single Group Short Time Series

Un test généralisé de Jonckheere pour tester la tendance dans des mesures répétées en blocs aléatoires ou pour un seul groupe de séries chronologiques courtes

Motivated by statistical methodology research in patient reported outcomes, we propose and develop a generalized Jonckheere test against ordered alternatives for testing trend in repeated measures in a randomized block design or a single group of short time series. We derive its asymptotic null distribution properties, describe methods for estimating the null distribution for testing the hypothesis, and propose a procedure of multiple comparisons. A numerical example on a score of pain is presented to illustrate the test procedure.

Cette étude est motivée par la recherche en méthodologie statistique sur les résultats déclarés par les patients. Nous proposons et développons un test généralisé de Jonckheere contre des alternatives ordonnées pour tester la tendance de mesures répétées dans un plan d'expériences en blocs aléatoires ou un seul groupe de courtes séries chronologiques. Nous dérivons le comportement asymptotique de la distribution de la statistique de test sous l'hypothèse nulle et décrivons des méthodes pour l'estimer. Nous proposons aussi une procédure pour les comparaisons multiples. Un exemple numérique sur le niveau de douleur est présenté pour illustrer la procédure du test.

2B-C1: Biostatistics: Inference Methods Biostatistique: méthodes inférentielles

Chair/Présidente: Lisa Lix (University of Saskatchewan)

MacNaughton Building, room/local 105

Abstracts/Résumés

LAJMI LAKHAL-CHAIEB & THIERRY DUCHESNE (Université Laval)

[Tuesday June 5 / mardi 5 juin, 10:15]

Association Measures in the Presence of a Bivariate Cure Fraction

Mesure d'association en présence d'une fraction de guérison bivariable

In this talk, we discuss association measures between two continuous time-to-event random variables in the presence of a bivariate cure fraction. We introduce a semi-parametric model to express the dependence. We derive an EM-like algorithm to estimate the parameters of our model. The proposed methodology is illustrated via a motivating study.

Dans cette présentation, on présente des mesures d'association entre deux variables aléatoires en présence d'une fraction de guérison. On introduit un modèle semi-paramétrique pour exprimer la dépendance. On propose une procédure inspirée de l'algorithme EM pour estimer les paramètres du dit modèle. La méthodologie proposée est illustrée avec un jeu de données réel.

JIANAN PENG (Acadia University)

[Tuesday June 5 / mardi 5 juin, 10:30]

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

Inférence simultanée pour l'estimation du risque à faible dose avec données continues pour l'analyse d'un banc d'essais

Risk assessment studies is challenging as study information is limited to high dose levels of a toxic agent. Piegorsch et al. (2005) developed the hyperbolic band for low dose inference in quantitative risk analysis. However, the shape of the constant width band may be more desirable than that of the hyperbolic band in some applications as a confidence band should be narrower especially near the lower end of low doses. In this talk, we will show that a constant width band should do better than the hyperbolic band considered by Piegorsch et al. (2005).

Les études d'évaluation du risque sont compliquées car l'information est limitée aux hautes doses d'agent toxique. Piegorsch et coll. (2005) ont développé une bande hyperbolique pour l'inférence à faible doses pour l'analyse quantitative du risque. Toutefois, une forme de bande à largeur constante est plus souhaitable qu'une forme hyperbolique pour certaines applications, car une bande de confiance devrait être plus étroite, particulièrement pour les plus faibles doses. Dans cet exposé, nous montrerons qu'une bande constante devrait donner de meilleurs résultats que la hyperbolique suggéré par Piegorsch et coll. (2005).

WANHUA SU (Grant MacEwan University), **MU ZHU** (University of Waterloo), **HUGH CHIPMAN** (Acadia University)

[Tuesday June 5 / mardi 5 juin, 10:45]

Precision-Recall Curve or ROC Curve?

Courbe précision-rappel ou courbe ROC?

2B-C1: Biostatistics: Inference Methods Biostatistique: méthodes inférentielles

The Receiver Operating Characteristic (ROC) curve is a popular performance measure in evaluating different rating systems. For unbalanced data (e.g., information retrieval), however, precision-recall curve is advocated. In this talk, we show that precision-recall curve has a higher resolution than ROC curve in comparing the performance of different ranking systems for rare target detection problems. That means precision-recall curve is able to distinguish two ranking algorithms that ROC curve fails to.

La courbe de la fonction d'efficacité du récepteur (ou courbe ROC) est une mesure de performance communément utilisée pour évaluer différents systèmes de classification. Pour les données déséquilibrées (par exemple la recherche d'informations), on préfère toutefois la courbe précision-rappel. Nous montrons que la courbe précision-rappel présente une meilleure résolution que la courbe ROC pour comparer la performance de différents systèmes de classification pour la détection des cibles rares. Cela signifie que la courbe précision-rappel permet de distinguer entre deux algorithmes de classification, contrairement à la courbe ROC.

LIHUI ZHAO (Northwestern University), **LU TIAN** (Stanford University), **TIANXI CAI**, **BRIAN CLAGGETT** & **LEE-JEN WEI** (Harvard University)

[Tuesday June 5 / mardi 5 juin, 11:00]

Effectively Selecting a Target Population with Desired Treatment Benefit Using the Data from a Randomized Comparative Study

Choix efficace d'une population cible sensible à l'effet bénéfique voulu d'un traitement au moyen des données provenant d'une étude comparative randomisée

The success of a randomized clinical trial heavily depends on the choice of the target population. In this research, we show a systematic and effective way to identify a promising population, for which the new treatment is expected to have a desired benefit, using the data from a current study involving similar comparator treatments. Specifically, we first create a parametric scoring system using multiple covariates to estimate subject-specific treatment differences. We then use a nonparametric method to estimate the average treatment difference for the "promising" subgroups selected by the score. Our method is illustrated with a cardiovascular clinical trial.

Le succès d'un essai clinique randomisé dépend largement du choix de la population cible. Dans cette recherche, nous proposons une façon systématique et efficace d'identifier une population prometteuse sur laquelle il est prévu qu'un nouveau traitement ait l'effet bénéfique voulu, en utilisant les données d'une étude en cours comportant des traitements comparatifs similaires. Plus précisément, nous créons d'abord un système de scores paramétriques en utilisant des covariables multiples pour estimer les différences de traitement spécifiques aux sujets. Nous utilisons ensuite une méthode non paramétrique pour estimer la différence moyenne de traitement pour les sous-groupes « prometteurs » sélectionnés par le score. Notre méthode est illustrée au moyen d'un essai clinique en médecine cardiovasculaire.

2B-C2: Hazard Models and Censored Data Modèles de risque et données censurées

Chair/Président: Pierre-Jérôme Bergeron (University of Ottawa)

MacNaughton Building, room/local 113

Abstracts/Résumés

XIN XIN, JULIE HORROCKS & GERARDA DARLINGTON (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 10:15]

Ties Between Event Times and Jump Times in the Cox Model

Égalités entre les temps d'événements et les temps de sauts dans le modèle de Cox

The partial likelihood of the Cox-PH model is affected by ties between event times and moments where discrete time-varying covariates change. The default behaviour of current softwares will be discussed and proposals for simple methods for dealing with such ties will be outlined. Results of a simulation study show that current methods found in commonly used software can lead to biased estimates of the regression coefficient of a binary time-varying covariate and that two proposed methods reduce estimation bias. Illustrations of these methods will be presented.

La vraisemblance partielle du modèle à risques proportionnels de Cox est affectée par des égalités entre les temps d'événements et les moments où les variables temporelles discrètes changent. Nous aborderons le comportement par défaut des logiciels actuels et nous proposerons des méthodes simples pour gérer de telles égalités. Les résultats d'une étude de simulation montrent que les méthodes actuelles utilisées dans les logiciels courants peuvent donner lieu à des estimations biaisées du coefficient de régression d'une covariable temporelle binaire. Deux des méthodes proposées, pour lesquelles nous présentons des illustrations, réduisent le biais d'estimation.

THARSHANNA NADARAJAH & ASOKAN MULAYATH VARIYATH (Memorial University of Newfoundland)

[Tuesday June 5 / mardi 5 juin, 10:30]

Penalized Empirical Likelihood Based Variable Selection for Cox's Proportional Hazards Model

Sélection de variables pour le modèle de Cox à risque proportionnel basée sur la vraisemblance empirique pénalisée

Variable selection is an important problem in survival analysis. In practice, an investigator may take many covariates as potential risk factors at the initial stage of the modeling and the selection of significant risk factors plays a crucial role. We propose a nonparametric version of the penalized-likelihood variable selection method for Cox's proportional hazards model with right censoring, by replacing the parametric likelihood by the empirical likelihood. Our simulation studies shows that our proposed method is consistent and when a parametric model is available its performance is comparable with existing methods.

La sélection de variables est un problème important pour l'analyse de survie. En pratique, le chercheur inclut plusieurs covariables en tant que facteurs de risque potentiels au premier stade de la modélisation et la sélection des facteurs de risque significatifs joue un rôle important dans la modélisation. Nous proposons une version non-paramétrique de la méthode de sélection de variables de la vraisemblance pénalisée pour le modèle de Cox à risque proportionnel avec censure à droite en remplaçant la vraisemblance paramétrique par la vraisemblance empirique. Nos simulations montrent que notre méthode est convergente et que sa performance est comparable à celle des méthodes existantes lorsqu'un modèle paramétrique est disponible.

2B-C2: Hazard Models and Censored Data Modèles de risque et données censurées

KATRINA TOMANELLI, ABDULKADIR HUSSEIN & SÉVÉRIEN NKURUNZIZA (University of Windsor)

[Tuesday June 5 / mardi 5 juin, 10:45]

Improved Estimation of Aalen's Additive Hazards Model

Estimation raffinée dans le modèle de risque additif d'Aalen

We consider Aalen's additive hazards model for censored data under the assumption that the time-dependent regression coefficients restricted to a prior subspace. If the restriction information is uncertain, then it is unclear whether to employ the restricted or the unrestricted least squared estimators (LSEs). We propose a class of Stein-type estimators of the parameters which combine the restricted and unrestricted LSEs. We show both analytically and by simulations that such estimators have relatively superior performance. We illustrate the proposed methodology by using data from a clinical trial on primary biliary cirrhosis of the liver.

Nous considérons le modèle de risques additif d'Aalen pour des données censurées sous l'hypothèse que les paramètres dépendent du temps, en plus d'être restreints dans un sous-espace a priori. Si la restriction est incertaine, alors le choix entre l'estimateur des moindres carrés restreint et sans restriction n'est pas clair. En combinant lesdits estimateurs, nous proposons une classe d'estimateurs de type-Stein pour les coefficients de régression. Nous montrons analytiquement et par simulations que de tels estimateurs sont relativement meilleurs. Nous illustrons l'application de la méthode proposée à l'aide des données d'un essai clinique sur la cirrhose biliaire primitive du foie.

PORDELI POONEH, XUEWEN LU & MURRAY BURKE (University of Calgary)

[Tuesday June 5 / mardi 5 juin, 11:00]

Estimation in Partially Linear Single-index Additive Hazards Regression with Current Status data

Estimation dans la régression des risques additive à index unique partiellement linéaire avec des données d'état actuel

We propose a partially linear single-index additive hazards regression model for current status data. The linear covariates are time-dependent and the nonlinear covariates are time-independent. The proposed model can model both linear and nonlinear covariate effects on the hazard and it avoids "curse of dimensionality". We use B-splines to model the nonparametric covariate functions. Asymptotic properties of the estimators are derived using the theory of counting processes. Simulation studies are presented to compare the new method with the standard linear additive hazards regression model.

Nous proposons un modèle de régression des risques additif à index unique partiellement linéaire pour des données d'état actuel. Les covariables linéaires sont dépendantes du temps, contrairement aux covariables non linéaires. Le modèle proposé peut modéliser les effets sur le risque des covariables linéaires et non linéaires et évite la « malédiction de la dimensionnalité ». Nous utilisons des B-splines pour modéliser les fonctions à covariables non paramétriques et dérivons les propriétés asymptotiques des estimateurs par la théorie des processus de comptage. Nous présentons des études de simulation pour comparer la nouvelle méthode avec le modèle de régression des risques additif linéaire standard.

XUEWEN LU (University of Calgary), **HENG LIAN** (Nanyang Technological University, Singapore), **HUA LIANG** (University of Rochester)

[Tuesday June 5 / mardi 5 juin, 11:15]

Variable Selection for Censored Partially Linear Single-Index Models

Sélection de variables pour modèles à un index partiellement linéaires et censurés

We consider variable selection for partially linear single-index models with randomly censored sam-

Nous considérons la sélection de variables pour des modèles à un index partiellement linéaires avec des échan-

2B-C2: Hazard Models and Censored Data Modèles de risque et données censurées

ples. We adopt a weighted profile least-squares procedure for estimation of regression coefficients. We invoke the smoothly clipped absolute deviation penalty (SCAD) approach for simultaneous variable selection and estimation. We show that the resultant SCAD estimators are consistent and hold the oracle property. We modify the tuning parameter selector BIC for the complete data case and show that the modified BIC is able to identify the true model consistently. We present simulation results for illustration.

tillons aléatoirement censurés. Nous adoptons une procédure pondérée et profilée des moindres carrés pour l'estimation des coefficients de régression. Nous invoquons l'approche de la pénalité lisse coupée de la déviation absolue (SCAD) pour la sélection de modèles et l'estimation simultanée. Nous montrons que les estimateurs SCAD qui en résultent sont convergents et possèdent la propriété "oracle". Nous modifions le paramètre de réglage du sélecteur BIC pour le cas des données complètes et montrons que le BIC modifié peut identifier le vrai modèle de façon consistante. Nous présentons des résultats de simulation pour illustrer le propos.

2B-C3: Hypothesis Testing and R Statistical Software Tests d'hypothèses et logiciel statistique R

Chair/Présidente: Erica Moodie (McGill University)

Rozanski Hall, room/local 105

Abstracts/Résumés

MARTA PADILLA & DAVID R. BICKEL (University of Ottawa)

[Tuesday June 5 / mardi 5 juin, 10:15]

Empirical Bayes Method Corrected for Small Number of Tests

Méthode bayésienne empirique corrigée pour un petit nombre de tests

Histogram-based empirical Bayes methods developed for analyzing data for large numbers of biological features tend to have large biases when applied to data with a smaller number of features, such as those measured conventionally. To analyze small-scale data, we introduce a correction of the maximum likelihood estimator (MLE) of the local false discovery rate (LFDR). The correction lies in excluding each feature when estimating the required parameters. Both methods are compared using simulated data. Results show that the corrected MLE has smaller negative bias, although both have strong negative biases when the proportion of unaffected features is greater than 90%.

Les méthodes bayésiennes basées sur des histogrammes qui ont été développées pour analyser des données avec un grand nombre de caractéristiques biologiques tendent à être largement biaisées quand elles sont appliquées à un nombre de caractéristiques plus petit, comme on en mesure habituellement. Pour analyser des données à petites échelles, nous introduisons une correction à l'estimateur du maximum de vraisemblance (EMV) du taux local de fausses découvertes (LFDR). La correction est créée en excluant chaque caractéristique lors de l'estimation des paramètres. Les deux méthodes sont comparées à l'aide de données simulées. Les résultats montrent que l'EMV corrigé a un biais négatif plus petit, quoique les deux méthodes aient de grands biais négatifs quand la proportion de caractéristiques non-affectées est plus grande que 90%.

MIN CHEN, CHANGBAO WU & MARY THOMPSON (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 10:30]

Empirical Likelihood Method for Pretest-Posttest Studies

Méthode de vraisemblance empirique pour les études prétest/post-test

Pretest-posttest studies are very popular to assess the treatment effect in clinical trials. Huang, Qin and Follmann (2008) analyzed the treatment effect in pretest-posttest studies using the empirical likelihood (EL) method. Their proposed method, however, deals with the mean responses of treatment and control separately, and confidence intervals were constructed through bootstrap. In this talk, we will explore alternative EL formulations directly over the difference of the mean responses between the two groups, using an approach similar to Wu and Yan (2012). Confidence intervals on the parameter of interest are constructed through

Les études prétest/post-test sont largement utilisées pour évaluer l'effet d'un traitement dans les essais cliniques. Huang, Qin et Follmann (2008) ont analysé cet effet par la méthode de vraisemblance empirique (VE). Cependant, leur méthode traite les réponses moyennes du traitement et du contrôle séparément et dérivent les intervalles de confiance par bootstrap. Nous explorons d'autres formulations de VE directement sur l'écart entre les réponses moyennes des deux groupes, par une approche similaire à Wu et Yan (2012). Nous construisons les intervalles de confiance pour le paramètre d'intérêt via les statistiques de ratio de VE. Enfin, nous présentons des résultats de simulations.

2B-C3: Hypothesis Testing and R Statistical Software Tests d'hypothèses et logiciel statistique R

the EL ratio statistics. Simulation results are provided.

G. ARTHUR MIHRAM (Princeton, NJ), **DANIELLE MIHRAM** (University of Southern California)

[Tuesday June 5 / mardi 5 juin, 10:45]

Confronting Quinn: Our Regression and Hypothesis-testing Methodologies in Modern Science
Face à Quinn: nos méthodologies, régression et tests d'hypothèses en science moderne.

Mathematician Frank Quinn concludes [NOTICES, AMS: 59: 31 (2012)] that our mathematics is not science. Science is that human activity devoted to the search for the very explanation for (i.e., for the truth about) any particular naturally occurring phenomenon. Cotes exemplarily described Modern Science [Preface, 2nd Edition, Newton's "Philosophiae Naturalis Principia Mathematica", ca 1713]: correctly implemented natural philosophy (viz., observation, then mental reflection thereon). Our statistical regression methodology indeed focuses on structured data-collection (observations), miming this natural philosophy initially. Our statistical hypothesis-testing methodology employs a "size", α , implying that $100\alpha\%$ of proclaimed (published) statistically significant differences are in error.

Le mathématicien Quinn conclut [NOTICES, AMS (2012) : 31] que nos mathématiques ne sont pas une science. La Science est l'activité humaine consacrée à la recherche de l'explication même (de la vérité) d'un phénomène produit naturellement. Cotes décrit de façon exemplaire la science moderne [Preface, 2nd Edition, Newton's "Philosophiae Naturalis Principia Mathematica" (1713)] : une philosophie naturelle correctement implémentée (observer, puis y penser activement). Notre méthodologie de régression statistique se concentre sur une saisie organisée de données (observations), mimant au départ cette philosophie naturelle. Notre méthodologie statistique de tests d'hypothèses emploie un niveau de signification α : alors $100\alpha\%$ des différences proclamées/publiées statistiquement significatives sont erronées.

ERIC CAI (University of Toronto)

[Tuesday June 5 / mardi 5 juin, 11:00]

A Pathological Monte Carlo Simulation of the Power Function of the Pearson Correlation Test
Simulation de Monte Carlo problématique pour la fonction puissance du test de corrélation de Pearson

The original intent of this project - done in a graduate statistical computing course - was to compare the power functions of the Pearson correlation test and the Spearman correlation test using Monte Carlo simulations. This was specifically done for data from two distributions: the bivariate normal distribution and the bivariate t-distribution. However, pathological power functions of the Pearson correlation test were encountered for data from the bivariate t-distribution. This pathology and the solution to overcome it to obtain the correct power function will be discussed in detail.

Réalisé dans le cadre d'un cours d'informatique statistique de niveau supérieur, ce projet visait à comparer les fonctions puissance des tests de corrélation de Pearson et de Spearman au moyen de simulations de Monte Carlo. Cette comparaison a été effectuée spécifiquement sur les données de distributions normale bivariée et de Student bivariée. Cependant, des fonctions puissance problématiques ont été obtenues pour le test de corrélation de Pearson avec les données de la distribution de Student bivariée. Nous expliquons en détail le problème en question ainsi que la solution proposée pour le régler et obtenir la fonction puissance correcte.

2B-C3: Hypothesis Testing and R Statistical Software Tests d'hypothèses et logiciel statistique R

ADRIAN WADDELL & R. WAYNE OLDFORD (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 11:15]

Extending RnavGraph to Use the R Plotting Facility

Extension de RnavGraph pour utiliser la fonctionnalité graphique de R

Navigation graphs define an infrastructure to link a “You are here” bullet, constrained to be on the graph, to some view onto the data. As one example, one may link the nodes of the graph to 2d scatterplots and the graph-edges to smooth 3d or 4d transitions between the two scatterplots. In this talk, we show how to create navigation graphs with arbitrary linking. We use the RnavGraph R package to create the navigation graphs and link them to different visualization systems in R, such as the base R graphics system, grid, ggplot2 and rgl.

Les graphiques de navigation définissent une infrastructure reliant un point « Vous êtes ici », restreint au graphique, à une certaine vision des données. À titre d'exemple, il est possible de lier les nœuds du graphique à des nuages de points 2D, et les arêtes du graphe à des transitions lisses 3D et 4D entre les deux nuages de points. Dans cet exposé, nous montrons comment créer des graphiques de navigation à partir de liens arbitraires. Nous utilisons la librairie R RnavGraph pour créer les graphiques et les lier à différents systèmes de visualisation dans R, notamment le système graphique de base de R, grid, ggplot2 et rgl.

DUNCAN MURDOCH (University of Western Ontario)

[Tuesday June 5 / mardi 5 juin, 11:30]

tables: An R Package to Produce Tables

tables : Une librairie R pour produire des tableaux

Tables are an old, boring topic: everything there is to say about them was said by Ehrenberg (JRSS A, 1977). But they are under attack, e.g. Gelman's tongue-in-cheek “Why tables are really much better than graphs” (JCGS, 2011), because there are so many badly produced tables in published papers. The `tables` package is an effort to counteract this, to make it easy to produce good tables. It uses a formula interface to describe the table (reminiscent of `PROC TABULATE` in SAS), and produces ASCII, CSV or \LaTeX output.

Les tableaux, un sujet vieux et ennuyant : tout ce qu'il y a à dire à leur sujet a été dit par Ehrenberg (JRSS A, 1977). Mais, ils sont sous attaque (e.g. l'article ironique de Gelman “Pourquoi les tableaux sont vraiment mieux que les graphiques”, JCGS, 2011), et cela parce qu'il y a tant de tableaux mal conçus publiés dans la littérature. La librairie `tables` a été créée afin d'enrayer ce phénomène et de rendre la création de bons tableaux plus facile. Cette librairie utilise une interface avec des formules pour décrire le tableau (comme pour `PROC TABULATE` dans SAS) et produit des résultats ASCII, CSV ou \LaTeX .

2B-C4: Model and Variable Selection Sélection de modèle et de variables

Chair/Président: Patrice Mathieu (Statistics Canada / Statistique Canada)

Rozanski Hall, room/local 102

Abstracts/Résumés

DENIS TALBOT, GENEVIÈVE LEFEBVRE & JULI ATHERTON (Université du Québec à Montréal)

[Tuesday June 5 / mardi 5 juin, 10:15]

Bayesian Adjustment for Confounding : A Look at its Theoretical Basis

Ajustement bayésien pour la confusion : un regard sur la base théorique de la méthode

It is well known that causal inference and predictive inference are two different concepts. Wang and his collaborators proposed a data-driven method for model selection in a causal inference framework that accounts for uncertainty associated with model selection (*Bayesian Adjustment for Confounding*). We will discuss this method and examine its theoretical basis. We also show that the method can be justified more rigorously using the graphical framework advocated by Pearl and deriving sufficient conditions to avoid confounding.

Il est bien connu que l'inférence causale et l'inférence prédictive sont deux concepts différents. Wang et ses collaborateurs proposent une méthode d'exploration des données qui permet d'effectuer de la sélection de modèle dans un cadre d'inférence causale en tenant compte de l'incertitude associée au choix de modèle (*Bayesian Adjustment for Confounding*). Nous effectuerons une présentation de leur méthode en apportant un regard critique sur la justification théorique proposée. Nous montrerons également que la méthode peut être justifiée de façon plus rigoureuse en se référant à l'approche graphique proposée par Pearl en dérivant des conditions suffisantes pour éviter la confusion.

SHINPEI IMORI, HIROKAZU YANAGIHARA & HIROFUMI WAKAKI (Hiroshima University, Japan)

[Tuesday June 5 / mardi 5 juin, 10:30]

Theoretical Comparisons with AIC and TIC

Comparaisons théoriques de l'AIC et du TIC

In many fields, Akaike's information criterion (AIC) and Takeuchi's information criterion (TIC) are frequently used for selecting the best model among the candidate models. The AIC was derived under the assumption that the candidate model is correctly specified, and the TIC was developed under the assumption that the candidate model is not necessarily correctly specified. In the present study, we discuss the theoretical difference between AIC and TIC through the asymptotic probabilities of selecting the true model by AIC and TIC. Especially, we focus on selections of mean and covariance structures.

Dans plusieurs domaines, le critère d'information d'Akaike (AIC) et le critère d'information de Takeuchi (TIC) sont fréquemment utilisés pour sélectionner le meilleur modèle parmi une liste de candidats. L'AIC est dérivé sous l'hypothèse que le modèle suspecté est correctement spécifié, et le TIC a été développé en supposant que le modèle suspecté n'est pas nécessairement spécifié correctement. Dans cette étude, nous traitons des différences théoriques entre l'AIC et le TIC par le biais des probabilités asymptotiques de sélection du vrai modèle avec l'AIC et le TIC. Nous nous attardons particulièrement à la sélection des structures de moyennes et covariance.

2B-C4: Model and Variable Selection Sélection de modèle et de variables

AHMED FARAHAT, ALI GHODSI & MOHAMED KAMEL (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 10:45]

An Efficient Greedy Method for Unsupervised Variable Selection

Une méthode gloutonne efficace de sélection de variables non supervisée

The selection of relevant variables is a crucial task in data analysis. This work proposes a novel method for unsupervised variable selection, which efficiently selects variables in a greedy manner. We first define an effective criterion for unsupervised variable selection which measures the reconstruction error of the data matrix based on the selected variables, and then present a novel algorithm for greedily minimizing the reconstruction error based on the variables selected so far. Experiments on real datasets demonstrate the effectiveness of the proposed algorithm in comparison to the state-of-the-art methods for unsupervised variable selection.

La sélection des variables pertinentes est une tâche cruciale dans l'analyse de données. Cet article propose une nouvelle méthode de sélection non supervisée qui permet de choisir efficacement des variables d'une manière gloutonne. Nous définissons d'abord un critère efficace de sélection de variables non supervisée qui mesure l'erreur de reconstruction de la matrice de données en fonction des variables sélectionnées. Ensuite, nous présentons un nouvel algorithme qui permet de minimiser d'une manière gloutonne l'erreur de reconstruction en fonction des variables sélectionnées jusqu'ici. Des expériences sur des jeux de données réelles démontrent l'efficacité de l'algorithme proposé en comparaison avec les techniques de pointe de sélection des variables non supervisée.

CHRISTIAN LÉGER & PIERRE LAFAYE DE MICHEAUX (Université de Montréal)

[Tuesday June 5 / mardi 5 juin, 11:00]

A Law of the Single Logarithm for Weighted Sums of Arrays Applied to Bootstrap Model Selection in Regression

Loi du simple logarithme de sommes pondérées d'une suite triangulaire avec application au bootstrap en sélection de modèle

We generalize a law of the single logarithm obtained by Qi (1994) and Li et al. (1995) to the case of weighted sums of triangular arrays of random variables. We apply this result to bootstrapping the all-subsets model selection problem in regression, where we show that the popular Bayesian Information Criterion of Schwarz (1978) is no longer asymptotically consistent. Indeed, at the bootstrap level, the weighted sum of random variables used to select the variables is of order $(n \log n)^{1/2}$ instead of $(n \log \log n)^{1/2}$ as with the original variables.

Nous généralisons une loi du simple logarithme obtenue par Qi (1994) et Li et al. (1995) au cas de sommes pondérées d'une suite triangulaire de variables aléatoires. Nous appliquons ce résultat au cas du bootstrap dans le problème de la sélection de modèle en régression où nous démontrons que le populaire critère de l'information bayésienne (BIC) de Schwarz (1978) n'est plus asymptotiquement convergent. Ainsi, au niveau du bootstrap, la somme pondérée de variables aléatoires utilisée pour sélectionner les variables est de l'ordre de $(n \log n)^{1/2}$ plutôt que $(n \log \log n)^{1/2}$ comme c'est le cas pour les variables originales.

HIROKAZU YANAGIHARA, HIROFUMI WAKAKI & YASUNORI FUJIKOSHI (Hiroshima University, Japan)

[Tuesday June 5 / mardi 5 juin, 11:15]

A Consistency Property of AIC for Multivariate Linear Models when the Dimension and the Sample Size are Large

Propriété de convergence de l'AIC pour les modèles linéaires multivariés de grandes dimensions et avec une grande taille d'échantillon

It is common knowledge that Akaike's information criterion (AIC) is not consistent in the model selec-

Il est connu que le critère d'information d'Akaike (AIC) n'est pas convergent pour la sélection de modèles. Ceci

2B-C4: Model and Variable Selection Sélection de modèle et de variables

tion. This fact has been confirmed from an asymptotic selection-probability based on a large-sample asymptotic framework. However, we note that when a high-dimensional asymptotic framework such that the dimension and the sample size are large is used for evaluating a selection-probability, a consistency property of AIC for selecting variables in multivariate linear models can be proved. This means that a selection-probability of selecting the true model by AIC goes to 1 as the sample size and the dimension are approaching to infinity simultaneously.

est confirmé par les probabilités de sélection asymptotique avec une grande taille d'échantillon. Cependant, nous avons constaté que lorsque l'évaluation de la probabilité de sélection est effectuée au moyen d'un cadre asymptotique de grande dimension, c'est-à-dire pour une dimension et une taille d'échantillon grandes, une propriété de convergence de l'AIC pour la sélection de variables dans des modèles linéaires multivariés peut être prouvée. Ainsi, la probabilité de sélectionner le vrai modèle grâce à l'AIC tend vers 1 lorsque la taille d'échantillon et la dimension tendent simultanément vers l'infini.

MARIKO YAMAMURA & HIROKAZU YANAGIHARA (Hiroshima University, Japan), **MUNI SRIVASTAVA** (University of Toronto)

[Tuesday June 5 / mardi 5 juin, 11:30]

Variable Selection by C_p Statistic in Multiple Responses Regression with Fewer Sample Size than the Dimension

Sélection de variables avec la statistique C_p dans des régressions à réponses multiples dont la taille d'échantillons est inférieure à la dimension

We consider the problem of selecting q explanatory variables out of k ($q \leq k$), when the dimension p of the response variables is larger than the sample size n in the multiple responses regression. We consider C_p statistic which is an estimator of the sum of standardized mean squared errors. The standardization uses the inverse of the variance-covariance matrix of p response variables and thus the estimator of the inverse of the sample variance-covariance matrix. However, since $n < p$, such an inverse matrix cannot be used. Thus, we use the Moore-Penrose inverse and define the C_p statistic.

Nous étudions le problème de la sélection de q variables explicatives parmi k ($q \leq k$), lorsque la dimension p des variables réponses est supérieure à la taille d'échantillon n dans la régression à réponses multiples. Nous considérons la statistique C_p qui représente un estimateur de la somme des erreurs quadratiques moyennes normalisées. La normalisation fait appel à l'inverse de la matrice de variance-covariance des variables réponses p et l'estimateur utilise donc l'inverse de la matrice de variance-covariance de l'échantillon. Cependant, puisque $n < p$, la matrice inverse ne peut pas être utilisée. Nous utilisons plutôt l'inverse de Moore-Penrose pour définir la statistique C_p .

2B-C5: Theory and Applications of Copulas Théorie et applications des copules

Chair/Présidente: Johanna Nešlehová (McGill University)

Thornbrough Building, room/local 1307

Abstracts/Résumés

ELIF FIDAN ACAR (McGill University), **RADU V. CRAIU & FANG YAO** (University of Toronto)

[Tuesday June 5 / mardi 5 juin, 10:15]

Statistical Testing for Conditional Copulas

Tests statistiques pour copules conditionnelles

In conditional copula models, the copula parameter is deterministically linked to a covariate via the calibration function. The latter is of central interest for inference and is usually estimated nonparametrically. However, when a parametric model for the calibration function is appropriate, the resulting estimator exhibits significant gains in statistical efficiency. We develop methodology for testing a parametric formulation of the calibration function against a general alternative and propose a generalized likelihood ratio-type test that enables conditional copula model diagnostics. We derive the asymptotic null distribution of the proposed test and study its finite sample performance using simulations.

Dans les modèles de copules conditionnelles, le paramètre des copules est lié de manière déterministe à une covariable au moyen de la fonction d'étalonnage. Cette dernière est d'un intérêt central pour l'inférence et est habituellement estimée de façon non paramétrique. Cependant, lorsqu'un modèle paramétrique pour la fonction d'étalonnage est approprié, l'estimateur qui en découle présente une efficacité statistique significativement plus élevée. Nous élaborons une méthodologie pour tester une formulation paramétrique de la fonction d'étalonnage par rapport à une alternative générale et nous proposons un test du rapport de vraisemblance généralisé qui permet le diagnostic de modèles de copules conditionnelles. Nous obtenons la distribution asymptotique du test proposé sous l'hypothèse nulle et nous étudions la performance du test sur des échantillons de taille finie au moyen de simulations.

TOUNKARA FODE, LOUIS-PAUL RIVEST & BELKACEM ABDOUS (Université Laval)

[Tuesday June 5 / mardi 5 juin, 10:30]

A Copula-based Estimator for the Intracluster Correlation Coefficient

Estimation du coefficient de corrélation intraclasse pour des données binaires basée sur des copules

Exchangeable copulas are used to model an extra-binomial variation in Bernoulli experiments with a variable number of trials. Maximum likelihood inference procedures for the intra-cluster correlation coefficient are proposed. Profile confidence intervals are constructed for several specifications of the copula family. Alternative models where the probability of success depends on the number of trials are investigated. The sensitivity of the inference to the specification of the copula family is investigated in a simulation study. Numerical examples are presented.

Des copules échangeables sont utilisées pour modéliser la variation extra-binomiale dans une expérience de Bernoulli, avec un nombre variable d'essais. La méthode du maximum de vraisemblance est utilisée pour faire de l'inférence sur le coefficient de corrélation intraclasse. Des intervalles de confiance sont construits selon plusieurs spécifications de familles de copules. Des modèles alternatifs où la probabilité de succès dépend du nombre d'essais sont également considérés. Des études de simulation sont faites pour mesurer la sensibilité de l'inférence à la copule choisie. Des exemples numériques seront présentés.

2B-C5: Theory and Applications of Copulas Théorie et applications des copules

HÉLA ROMDHANI, LOUIS-PAUL RIVEST & LAJMI LAKHAL-CHAIEB (Université Laval)

[Tuesday June 5 / mardi 5 juin, 10:45]

Measuring and Testing the Intraclass Dependence for Non-normally Distributed Clustered Data
Mesurer et tester la dépendance intra-classe pour des données groupées non normales

We are interested in measuring the intraclass dependence for non-normal clustered data. We first propose a generalization of the one way random effect ANOVA model valid for the larger family of elliptical distributions. Under this model, we give the asymptotic properties of the moment estimator of the intraclass correlation coefficient. Then we propose an estimator of Kendall's tau adapted to exchangeable data. Under a copula model, we derive its asymptotic distribution and give an estimator for its variance. We also give an alternative estimator of the intraclass correlation coefficient for the class of elliptical distributions and derive its asymptotic properties.

Nous nous intéressons à mesurer la dépendance intra-classe pour des données groupées non-normales. Nous proposons, d'abord, une généralisation du modèle ANOVA à un facteur aléatoire valide pour la famille des distributions elliptiques. Sous ce modèle, nous dérivons les propriétés asymptotiques de l'estimateur des moments du coefficient de corrélation intra-classe. Nous proposons, ensuite, un estimateur du tau de Kendall adapté aux données échangeables. Sous un modèle de copules, nous donnons sa distribution asymptotique ainsi qu'un estimateur de sa variance. Nous proposons aussi, pour la famille de distributions elliptiques, un deuxième estimateur du coefficient de corrélation intra-classe et donnons ses propriétés asymptotiques.

BIN DONG & JUDY-ANNE W. CHAPMAN (NCIC Clinical Trials Group, Queen's University), **PAUL E. GOSS** (Massachusetts General Hospital Cancer Center, Harvard University), **MICHAEL N. POLLAK** (Lady Davis Institute, Jewish General Hospital, McGill University), **LOIS E. SHEPHERD** (NCIC Clinical Trials Group, Queen's University)

[Tuesday June 5 / mardi 5 juin, 11:00]

Copula Estimation for Censored Data via Empirical Likelihood: NCIC CTG MA.27 and MA.14 Endocrine Breast Cancer Trials

Estimation par copules pour données censurées à l'aide d'une vraisemblance empirique : essais endocriniens NCIC CTG MA.27 et MA.14 sur le cancer du sein

In multivariate survival analysis, copulas have become a popular tool for modeling the dependence in a vector of continuous time-to-event random variables subject to censoring. Many authors have investigated goodness-of-fit (gof) tests for copulas. In the presence of censoring, most of the established work is restricted to a particular class of copulas. In order to develop a gof test procedure that can be used in more general situations, I use an empirical likelihood (EL) approach to estimate copula non-parametrically. With this EL-based estimator of a copula, I can derive a gof test for assessing a specific parametric copula model for censored data.

Pour l'analyse de données de survie multivariées, les copules sont maintenant un outil populaire pour modéliser la dépendance dans un vecteur de variables aléatoires de temps de survie continues avec de la censure. Plusieurs auteurs ont étudié les tests d'adéquations pour les copules. En présence de censure, la majorité des travaux se restreignent à une classe spécifique de copules. Pour développer un test d'adéquation pouvant être utilisé dans des situations plus générales, j'utilise l'approche de la vraisemblance empirique pour estimer la copule de façon non-paramétrique. Avec cet estimateur de vraisemblance empirique de la copule, je peux dériver un test d'adéquation pour évaluer un modèle de copule paramétrique particulier pour données censurées.

2B-C5: Theory and Applications of Copulas Théorie et applications des copules

BEILEI WU & ALEXANDER R. DE LEON (University of Calgary)

[Tuesday June 5 / mardi 5 juin, 11:15]

Flexible Random Effects Copula Models for Clustered Mixed Outcomes – Application in Toxicology

Modèles de copules flexibles à effets aléatoires pour des résultats mixtes regroupés – application en toxicologie

We employ random effects models to analyze clustered mixed-outcomes and account for associations between discrete and continuous outcomes within clusters. The regression parameters in models for both outcomes are marginally meaningful; in addition, by assuming a latent variable framework to describe discrete outcomes, complications of copulas to discrete variables are avoided. The marginal distributions can be chosen from any distributions. Maximum likelihood estimation is implemented using SAS. Results of simulations concerning the bias and efficiency of the estimates are reported. The proposed methodology is motivated by and illustrated using a developmental toxicity study of ethylene glycol in mice.

Nous employons des modèles à effets aléatoires pour analyser des résultats mixtes regroupés en tenant compte des associations entre les résultats discrets et continus à l'intérieur des grappes. Les paramètres de régression des modèles pour les deux types de résultats sont marginalement probants. De plus, en assumant un cadre de variables latentes pour décrire les résultats discrets, on évite les complications dues à l'usage des copules en présence de variables discrètes. Les distributions marginales peuvent provenir de n'importe quelle distribution. Le maximum de vraisemblance est estimé à l'aide de SAS. Les résultats des simulations relatives aux biais et à l'efficacité des estimateurs sont présentés. La méthodologie proposée est motivée et illustrée au moyen d'une étude sur la toxicité de l'éthylène glycol sur le développement de la souris.

2C-A: Isobel Loutit Invited Address Allocution Isobel Loutit

Organizer and Chair / Responsable et président: Boxin Tang (Simon Fraser University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Rozanski Hall, room/local 103

Abstract/Résumé

ROGER HOERL (GE Global Research)

[Tuesday June 5 / mardi 5 juin, 13:15]

The Next Frontier in Business and Industrial Statistics: Attacking Complex, Unstructured Problems

La prochaine frontière de la statistique industrielle et d'entreprise: s'attaquer aux problèmes complexes et non structurés

Statistical problems presented at conferences or in journals typically correspond to a recognizable textbook chapter. However, a class of problems faced by statisticians in business and industry that appears to be becoming more and more strategic is complex and unstructured problems. This session will propose that statistical engineering can provide the needed framework from which to effectively attack large, complex, unstructured problems. Further, development of statistical engineering as a discipline can provide the sound theoretical basis and body of literature needed to guide statisticians in the integration of various methods into overall strategies needed to solve these types of problems.

Les problèmes statistiques présentés lors de conférences ou dans des revues ressemblent habituellement à un chapitre de livre scolaire. Cependant, une catégorie de problèmes se présentant aux statisticiens en entreprise et dans l'industrie, et semblant devenir de plus en plus stratégique, porte sur des problèmes complexes et non structurés. Cette séance proposera que l'ingénierie statistique peut fournir le cadre nécessaire pour s'attaquer efficacement à ces problèmes. Le développement de l'ingénierie statistique en tant que discipline pourrait fournir la structure nécessaire pour poser les bases théoriques et contribuer au développement de la littérature scientifique nécessaires à la résolution de ce type de problèmes.

2C-I1: Absolute Penalty Estimation and Variable Selection Estimation à l'aide de la perte absolue et sélection de variables

Organizer and Chair / Responsable et président: Ejaz Ahmed (Brock University)

MacNaughton Building, room/local 105

Abstracts/Résumés

YANG FENG (Columbia University), **JIANQING FAN & XIN TONG** (Princeton University)

[Tuesday June 5 / mardi 5 juin, 13:15]

A Nonparametric Optimal Decision Rule in High Dimensional Space

Une règle de décision non-paramétrique optimale dans un espace à plusieurs dimensions

A Nonparametric Optimal Decision Rule (NOD) is proposed to tackle the high-dimensional binary classification problem. We assume that the log-likelihood of class conditional densities can be written as a linear combination of log-likelihoods of marginal class conditional densities. After we plug-in estimates of marginal class conditional densities, to solve the classification problem is equivalent to finding the best linear classification rule in a nonparametrically-transformed feature space. Oracle inequalities are developed regarding the classification error. A vast array of simulation and real data analysis demonstrate the outstanding performance of NOD.

Nous proposons une règle de décision non-paramétrique optimale (DNO) pour aborder le problème de classification binaire en plusieurs dimensions. Nous supposons que la log-vraisemblance des densités conditionnelles par classe peut s'écrire comme une combinaison linéaire des log-vraisemblances des densités marginales par classe. Après avoir imputé les densités marginales par un estimé, la résolution du problème de classification équivaut à trouver la meilleure règle de classification linéaire dans un espace de caractéristiques transformé non-paramétriquement. Des inégalités oracle sont développées pour les erreurs de classification. Un grand ensemble de simulations et d'analyses de données réelles démontrent la performance remarquable des DNO.

M. SHAKHAWAT HOSSAIN (University of Winnipeg), **EJAZ AHMED** (Brock University), **KJELL A. DOKSUM** (University of Wisconsin-Madison)

[Tuesday June 5 / mardi 5 juin, 13:45]

Shrinkage and Penalty Estimators in Generalized Linear Models

Estimateurs par rétrécissement et pénalité pour les modèles linéaires généralisés

We consider the shrinkage and penalty estimators in generalized linear models when it is suspected that the regression coefficients may be restricted to a subspace. We develop the asymptotic bias and risk of a shrinkage estimator. A Monte Carlo simulation is conducted for comparing the relative performance of penalty and shrinkage estimators with respect to MLE. This indicates that shrinkage method is an effective model selection technique and is comparable to the LASSO, adaptive LASSO, and SCAD when there are moderate or large numbers of inactive predictors in the model. A real data example is presented to illustrate the proposed methods.

Nous considérons les estimateurs par rétrécissement et pénalité pour les modèles linéaires généralisés lorsque nous croyons que les coefficients de régression peuvent être restreints à un sous-espace. Nous développons le biais asymptotique et le risque d'un estimateur par rétrécissement. Une simulation de Monte-Carlo est menée pour comparer la performance relative des estimateurs par pénalité et rétrécissement par rapport à l'estimateur au maximum de vraisemblance. Ceci indique que la méthode par rétrécissement est une technique efficace pour sélectionner le modèle et elle se compare au LASSO, au LASSO adaptatif et au SCAD lorsqu'il y a un nombre moyen ou important de variables explicatives inactives dans le modèle. Un exemple concret permet d'illustrer les méthodes proposées.

2C-I1: Absolute Penalty Estimation and Variable Selection Estimation à l'aide de la perte absolue et sélection de variables

PAUL MCNICHOLAS & SAKYAJIT BHATTACHARYA (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 14:15]

A Modified BIC for Mixture Model Selection

Un CIB modifié pour la sélection d'un modèle de mélange

In the context of Gaussian mixture model-based clustering analysis, we propose a penalized likelihood approach with a penalty function akin to the LASSO. We derive an EM algorithm to fit our proposed model with some pre-defined priors. A modified BIC is proffered for selection of the number of components and the penalization parameter. Our approach is especially suitable for high dimension, low sample size settings. We show that this modified BIC has oracle properties, i.e. that the criterion is consistent and the estimated parameter is asymptotically normal. Examples are used for illustration.

Dans le cadre d'une analyse de regroupement basée sur un modèle de mélange gaussien, nous proposons une approche de vraisemblance pénalisée semblable au LASSO. Nous dérivons un algorithme EM pour ajuster le modèle proposé avec quelques lois a priori prédéfinies. Un CIB modifié est présenté pour la sélection du nombre de composantes et du paramètre de pénalisation. Notre approche s'applique particulièrement bien dans un contexte de grandes dimensions et de faible taille d'échantillon. Nous montrons que ce CIB modifié possède des propriétés oracle, c'est-à-dire que le critère est convergent et que le paramètre estimé est asymptotiquement normal. Des exemples sont employés à titre illustratif.

2C-I2: New Statistical Design and Analysis for Epidemiological Investigations Nouvelles méthodes de conception et d'analyse pour les enquêtes épidémiologiques

Organizer and Chair / Responsable et président: Yutaka Yasui (University of Alberta)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Rozanski Hall, room/local 101

Abstracts/Résumés

MASOUD ASGHARIAN, DAVID WOLFSON & CHRISTINA WOLFSON (McGill University)

[Tuesday June 5 / mardi 5 juin, 13:15]

Prevalent Cohort Design and Analysis: A Case Study

Plan et analyse de cohorte prévalente : une étude de cas

Logistic or other constraints often preclude the possibility of conducting incident cohort studies. A feasible alternative in such cases is a prevalent cohort study. When the interest lies in estimating the lifespan between an initiating event, experienced before the start of the study, and a terminating event, the prevalent subjects may be followed prospectively until the terminating event or loss to follow-up. These prevalent cases have, on average, longer lifespans and hence do not constitute a representative sample from the target population. I discuss the challenges in analyzing such data and illustrate recently developed methodologies using data from the CSHA.

Les contraintes logistiques ou autres excluent souvent la possibilité de mener des études sur une cohorte incidente. Dans de telles situations, une étude de cohorte prévalente constitue une alternative envisageable. Lorsque l'intérêt réside dans l'estimation de la durée de vie entre un événement initial survenu avant le début de l'étude et un événement de fin, les sujets prévalents peuvent être suivis de manière prospective jusqu'à l'événement de fin ou la perte du suivi. Ces cas prévalents ont, en moyenne, une plus longue durée de vie et ne constituent donc pas un échantillon représentatif de la population cible. Je discute les défis dans l'analyse de telles données et illustre les développements méthodologiques récents à l'aide de données de l'Étude sur la santé et le vieillissement au Canada (ESVC).

JUXIN LIU (University of Saskatchewan), **PAUL GUSTAFSON** (University of British Columbia)

[Tuesday June 5 / mardi 5 juin, 13:45]

Interactions and Average Predictive Comparison

Interactions et comparaison prédictive moyenne

In a regression model with interactions, an individual regression coefficient itself cannot provide insightful explanation regarding the effect of a particular predictor. Therefore, we consider an average predictive comparison (APC, Gelman and Pardoe 2007) as a target of inference. Particularly, in light of the difficulties in dealing with interaction terms in regression models, we examine inferences about APC when additive models are fitted to relationships truly involving pairwise interaction terms. We consider two different versions of APC and two different regression contexts.

Dans un modèle de régression avec interactions, un coefficient de régression ne peut fournir à lui seul une explication éclairée sur l'effet d'une variable explicative en particulier. Par conséquent, nous considérons comme cible d'inférence une comparaison prédictive moyenne (CPM, Gelman et Pardoe, 2007). En particulier, compte tenu des difficultés de traiter les termes d'interaction dans les modèles de régression, nous examinons les inférences de la CPM lorsque des modèles additifs sont ajustés à des relations comprenant des termes d'interaction par paire. Nous considérons deux versions différentes de la CPM et deux contextes différents de régression.

2C-I2: New Statistical Design and Analysis for Epidemiological Investigations Nouvelles méthodes de conception et d'analyse pour les enquêtes épidémiologiques

JOSE MIGUEL MARTINEZ (Universitat Pompeu Fabra, Spain), **JOSÉ MARIA VELARDE** (CiSAL, Universitat Pompeu Fabra, Spain), **RAMON CLÉRIES** (Cancer Registry of Catalonia and University of Barcelona, Spain / Cancer Registry of Catalonia et University of Barcelona, Spain), **JOAN BENACH** (Health Inequalities Research Group, GREDS-EMCONET, Spain), **FERNANDO G. BENAVIDES** (CiSAL, Universitat Pompeu Fabra and CIBERESP, Spain / CiSAL, Universitat Pompeu Fabra et CIBERESP, Spain), **YUTAKA YASUI** (University of Alberta)

[Tuesday June 5 / mardi 5 juin, 14:15]

Improving Multilevel Analyses: the Integrated Epidemiologic Design and its Extension

Amélioration des analyses multiniveaux : le plan d'expérience épidémiologique intégré et son extension

The “Integrated Design” combines population-level data and individual-level data into a unified analysis, and is an alternative to the “Multilevel Design” for assessing both “contextual” and “individual” covariate effects on outcomes. Its inference can be based on population-based estimating equations (PBEEs), or a Bayesian approach. Simulation studies show its efficiency advantage, especially for contextual covariates. In this talk, I will discuss the general framework of the Integrated Design for both risk and censored survival/rate data and its extension to handle missing data using weighted PBEEs. Advantages of this new design over the Multilevel Design will be illustrated with real data.

Le « plan intégré » combine des données aux niveaux de la population et des individus en une analyse unifiée, et se veut une alternative au « plan multiniveaux » pour l'évaluation des effets « contextuels » et « individuels » des covariables sur les résultats. Son inférence peut se fonder sur des équations d'estimation basées sur la population (EEBP) ou sur une approche bayésienne. Des études par simulation démontrent son avantage en termes d'efficacité, en particulier pour les covariables contextuelles. Dans cet exposé, je présente le cadre général de ce plan intégré tant pour les données de risque que censurées des taux de survie et son extension pour gérer les données manquantes en utilisant des EEBP pondérées. Les avantages de ce nouveau plan par rapport au plan multiniveaux sont illustrés à l'aide de données réelles.

2C-I3: Session in Honour of Jon N.K. Rao's 75th Birthday
Séance en l'honneur du 75e anniversaire de Jon N.K. Rao

Chair/Président: Louis-Paul Rivest (Université Laval)
Organizer/Responsable: David Haziza (Université de Montréal)

Rozanski Hall, room/local 104

Abstracts/Résumés

MIKE HIDIROGLOU & YONG YOU (Statistics Canada / Statistique Canada)

[Tuesday June 5 / mardi 5 juin, 13:15]

Confidence Intervals for Unit and Area Level Small Area Estimators

Estimation par intervalles de confiance pour petits domaines au niveau de l'unité ou du petit domaine

We focus on the estimation of small area means. We compare, via simulation, how their associated confidence intervals behave. Unit level estimators of the mean were computed, for each selected sample within the small areas, using EBLUP and pseudo-EBLUP estimators. Area level estimates were computed using the Fay-Herriot model as applied to the aggregated weighted or non-weighted small area estimates. We were particularly interested in comparing confidence interval coverage under model misspecification. Our simulation showed that, as expected, unit level estimates are more efficient than area level estimates. However, the pseudo-EBLUP estimates are robust under model misspecification.

Nous nous concentrons sur l'estimation des moyennes régionales. Nous comparons, par simulation, comment les intervalles de confiance leur étant associées se comportent. Des estimateurs de la moyenne des unités ont été calculés, pour chaque échantillon sélectionné à l'intérieur des petits domaines, en utilisant des estimateurs MPLSBE et pseudo-MPLSBE. Les estimations au niveau des petits domaines ont été calculées en utilisant le modèle de Fay-Herriot et ont été appliquées soit aux estimations agrégées pondérées ou non pondérées des petits domaines. Nous nous sommes particulièrement intéressés à comparer la couverture de l'intervalle de confiance étant donné que le modèle était mal spécifié.

JAE-KWANG KIM (Iowa State University), **SEUNGHWAN PARK** (Seoul National University, South Korea)

[Tuesday June 5 / mardi 5 juin, 13:45]

Fractional Imputation for Regression Analysis Combining Two Independent Surveys

Imputation fractionnaire pour des analyses de régression dans le cas de deux enquêtes indépendantes

We consider the problem of combining two survey data for regression analysis, where each of the two samples contains only partial items. The classical approach of matching based on demographic variables can lead to biased analyses. Using a surrogate variable that is observed for both surveys, we can obtain a consistent estimator of the regression coefficients. Implementation using fractional imputation and its variance estimation are discussed. Results from a limited simulation study confirm the validity of the proposed method.

Nous considérons le problème d'analyse de régression dans le contexte de deux ensembles combinés de données d'enquêtes, où les deux échantillons contiennent de l'information partielle. L'approche classique d'appariement sur des variables démographiques peut mener à des analyses biaisées. Au moyen d'une variable observée dans les deux enquêtes, nous obtenons un estimateur convergent des coefficients de régression. L'implantation au moyen de l'imputation fractionnaire et l'estimation de la variance sont discutées. Les résultats d'une étude par simulation confirment la validité de la méthode proposée.

2C-I3: Session in Honour of Jon N.K. Rao's 75th Birthday Séance en l'honneur du 75e anniversaire de Jon N.K. Rao

MARY THOMPSON (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 14:15]

Bootstrap Methods in Complex Surveys

Méthodes bootstrap dans les enquêtes complexes

This talk presents a review of the bootstrap in the survey sampling context, with emphasis on a paper published in the *Journal of the American Statistical Association* in 1988 which introduced the Rao-Wu bootstrap, and more recent work of Rao and others on bootstrapping with estimating functions. Some interesting problems, addressable in part by bootstrapping, arise in inference for the parameters of multilevel models when the sampling design is a multistage design and the inclusion probabilities are possibly informative, and in inference for network parameters under network sampling.

Cet exposé présente un examen de la méthode bootstrap dans le contexte d'enquêtes par sondage, en mettant l'accent sur un article publié en 1988 dans le *Journal of the American Statistical Association* qui présentait l'approche de Rao-Wu, ainsi que sur le travail plus récent mené par Rao et d'autres collègues sur le bootstrap de fonctions d'estimation. Certains problèmes intéressants, que le bootstrap peut résoudre partiellement, surviennent lors de l'inférence des paramètres de modèles à plusieurs niveaux dans un plan de sondage à plusieurs étapes, et dont les probabilités d'inclusion sont possiblement informatives, de même que lors de l'inférence des paramètres dans un échantillonnage en réseau.

2C-I4: Statistical Analysis of Network Graph Data Analyse statistique de réseaux et graphes

Organizer and Chair / Responsable et président: Shojaeddin Chenouri (University of Waterloo)

Alexander Hall, room/local 100

Abstracts/Résumés

CHRIS GROENDYKE (Robert Morris University), **DAVID WELCH** (University of Auckland, New Zealand), **DAVID R. HUNTER** (Pennsylvania State University)

[Tuesday June 5 / mardi 5 juin, 13:15]

A Network-Based Analysis of the 1861 Hagelloch Measles Data

Une analyse réseau des données sur l'épidémie de rougeole d'Hagelloch de 1861

Here we demonstrate a statistical method for fitting the parameters of a sophisticated network and epidemic model to disease data. The pattern of contacts between hosts is described by a class of dyadic independence ERGMs while the transmission process that runs over the network is modeled as a stochastic SEIR epidemic. We fit these models to very detailed data from the 1861 measles outbreak in Hagelloch, Germany. Using RJMCMC, we produce samples from the joint posterior distribution of all the parameters of this model and perform Bayesian model selection to find the best-fitting network model.

Dans cet exposé, nous illustrons une méthode statistique qui permet d'ajuster les paramètres d'un modèle sophistiqué de réseau et d'épidémie à des données. La prise de contact entre les hôtes est décrite par une classe dyadique de modèles graphiques exponentiels aléatoires indépendants, alors que le processus de transmission qui fonctionne sur un réseau est modélisé comme un modèle épidémique stochastique « susceptible - exposé - infecté - rétabli ». Nous ajustons ces modèles aux données très détaillées sur l'épidémie de rougeole de 1861 à Hagelloch (Allemagne). En utilisant la méthode Monte Carlo par chaîne de Markov à sauts réversible, nous produisons des échantillons de la loi conjointe a posteriori de tous les paramètres de ce modèle et nous effectuons une sélection de modèle bayésienne pour trouver le modèle de réseau optimal.

PATRICK PERRY & NIKOLAY KULMATITSKIY (New York University, Stern), **PATRICK WOLFE** (University College London, U.K.)

[Tuesday June 5 / mardi 5 juin, 13:45]

A Parametric View of Some Nonparametric Network Algorithms

Une vision paramétrique de certains algorithmes réseau non paramétriques

Methods for analyzing network data are not always grounded in rigorous statistical theory. They do not make explicit modeling assumptions, and they do not come with consistency guarantees. Here we analyze two popular network tools, the expected degree model and modularity-based community detection. We show that the expected degree model is the maximum likelihood estimate under a certain log-linear model, and we show that modularity-based community detection is approximate maximum likelihood estimation under a differ-

Les méthodes d'analyse de données réseau ne sont pas toujours fondées sur une théorie statistique rigoureuse. Les hypothèses du modèle ne sont pas énoncées explicitement et la convergence n'est pas garantie. Dans cet exposé, nous analysons deux outils populaires pour les données réseau : le modèle de degré espéré et la détection de communautés basée sur la modularité. Nous montrons que le modèle de degré espéré correspond à l'estimateur du maximum de vraisemblance d'un certain modèle log-linéaire. Nous montrons également que la détection de communautés basée sur la modularité est une approximation du maximum de

2C-I4: Statistical Analysis of Network Graph Data Analyse statistique de réseaux et graphes

ent model. These results enable us to extend the analyzed ad hoc network algorithms by applying standard statistical model selection and estimation devices.

vraisemblance sous un autre modèle. Ces résultats nous permettent d'étendre les algorithmes réseau ad hoc par l'application de dispositifs de sélection et d'estimation de modèles statistiques standards.

ALI SHOJAIE (University of Washington), **BASU SUMANTA & MICHAILIDIS GEORGE** (University of Michigan)

[Tuesday June 5 / mardi 5 juin, 14:15]

Estimation of Network Granger Causal Models with Inherent Group Structures

Estimation des modèles de réseaux de liens de causalité au sens de Granger avec des structures de groupe inhérentes

The problem of estimating high-dimensional network structures arises naturally in the analyses of many physical, biological and socio-economic systems. In many of these applications, the variables have inherent grouping structures, that when incorporated, can result in improved estimation and prediction. I introduce different variants of group lasso penalty to estimate the structure of the network by employing the framework of graphical Granger causal models, under the assumptions of sparsity of its edges. Asymptotic results on the consistency of the new estimation procedure are developed. The performance of the proposed methodology is assessed through extensive numerical studies.

Le problème d'estimation d'une structure de réseau de grande dimension survient naturellement dans l'analyse de nombreux systèmes physiques, biologiques et socio-économiques. Dans un grand nombre de ces applications, les variables ont des structures inhérentes de regroupement, qui lorsqu'elles sont incorporées, peuvent améliorer l'estimation et la prévision. Je présente différentes variantes des pénalités de type lasso pour estimer la structure du réseau au moyen de modèles graphiques de liens de causalité au sens de Granger, sous les hypothèses de l'harmonie de leurs arêtes. Des résultats asymptotiques sur la convergence de la nouvelle procédure d'estimation sont produits. Le rendement de la méthodologie proposée est analysé avec des études numériques approfondies.

2C-I5: Stochastic Methods in Finance Méthodes stochastiques en finance

Organizer and Chair / Responsable et président: Cody Hyndman (Concordia University)

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

Alexander Hall, room/local 200

Abstracts/Résumés

ALEXANDRU BADESCU (University of Calgary), **JUAN-PABLO ORTEGA** (Université de Franche-Comté, France)
[Tuesday June 5 / mardi 5 juin, 13:15]

Hedging GARCH Options with Generalized Innovations
Couverture d'options GARCH avec innovations généralisées

We study the performance of different hedging schemes when the asset return process is modeled by a general class of GARCH models. Since the minimal martingale measure fails to produce a probability measure in this setting, we construct local risk minimization (Irm) hedging strategies with respect to a risk neutral measure. Using the Esscher transform and the Girsanov principle as our martingale measure candidates, we construct Irm delta hedges based on different distributional assumptions regarding the GARCH innovations. A numerical experiment is conducted to compare these hedges to the standard stochastic volatility hedges for different option maturities and hedging frequencies.

Nous étudions la performance de différents schémas de couverture lorsque le processus de retour sur l'investissement est modélisé par une classe générale de modèles GARCH. Puisque la mesure minimale de martingale ne peut pas produire une mesure de probabilité dans ce cadre, nous construisons des stratégies de minimisation locale du risque (MLR) de couverture sous une mesure neutre au risque. En employant la transformation d'Esscher et le principe de Girsanov comme candidats pour la martingale définissant la mesure, nous construisons des couvertures delta de MLR qui s'appuient sur différentes hypothèses de distribution des innovations GARCH. Une simulation numérique est réalisée pour comparer ces couvertures aux couvertures traditionnelles de volatilité stochastique pour différentes échéances et différentes fréquences de couvertures d'options.

SEBASTIAN JAIMUNGAL (University of Toronto), **ALVARO CARTEA** (University College London, U.K.)
[Tuesday June 5 / mardi 5 juin, 13:45]

Risk Measures and Fine Tuning of High Frequency Trading Strategies
Mesures de risque et ajustement fin de stratégies de négociation à haute fréquence

We propose risk measures to assess the performance of High Frequency (HF) trading strategies that seek to maximize profits from making the realized spread. The HF trader is risk-neutral and maximizes expected terminal wealth but is constrained by both capital and inventories. The risk measures enable the HF trader to tune her strategies by trading off different measures of inventory risk against expected profits. The asset midprice is driven by a Brownian motion plus jumps when market orders

Nous proposons des mesures de risque pour évaluer la performance de stratégies de négociation à haute fréquence (HF) qui cherchent à maximiser les profits en négociant les titres où un écart est observé. Le négociant HF est neutre au risque et maximise la richesse prévue à l'échéance, mais il est contraint par le capital et les inventaires. Les mesures de risque permettent au négociant HF d'ajuster ses stratégies en échangeant différentes mesures de risque d'inventaire par rapport aux profits espérés. Le prix moyen de l'actif évolue selon un mouvement brownien avec sauts lorsque les

2C-I5: Stochastic Methods in Finance Méthodes stochastiques en finance

arrive. We set up and demonstrate how to solve the optimal control problem and illustrate features of the optimal strategy through examples.

ordres du marché arrivent. Nous définissons le problème de contrôle optimal, nous démontrons comment le résoudre et nous illustrons les caractéristiques de la stratégie optimale à l'aide d'exemples.

ADAM METZLER, R. MARK REESOR & JINGYA LI (University of Western Ontario)

[Tuesday June 5 / mardi 5 juin, 14:15]

Information, Ambiguity and the Cost of Contingent Capital

Information, ambiguïté et coût du capital conditionnel

A contingent capital bond (CCB) is a security which “begins life” as subordinated debt, and converts to equity if the issuing firm becomes financially distressed. Recently they have been suggested as a means to shift the cost of bailouts from taxpayers to a failed institution’s owners. Previous theoretical analyses have indicated that the cost of contingent capital to the issuing institution would be surprisingly cheap. In this talk we investigate the extent to which this holds under various levels of ambiguity. Time permitting we will discuss the impacts of parameter uncertainty, discretionary/ambiguous conversion triggers and incomplete accounting information.

Une obligation de capital conditionnel est un titre commençant par une dette subordonnée, puis convertie en capitaux propres, au besoin. Il a été proposé que ces bons soient utilisés pour transférer les coûts d’un plan de sauvetage des contribuables vers les propriétaires de la firme. Les analyses précédentes ont indiqué que le coût du capital conditionnel pour l’entreprise émettrice serait faible. Lors de cet exposé, nous étudierons dans quelle mesure ce résultat est véridique sous ambiguïté. Si le temps le permet, nous discuterons des impacts de l’incertitude à propos des paramètres, des déclencheurs d’une conversion (ambiguë ou à la discrétion des actionnaires) et de l’incomplétude de l’information comptable.

2C-C1: Models Accounting for Measurement Error Modèles avec erreurs de mesure

Chair/Présidente: Michelle Qian Zhou (Harvard School of Public Health)

Rozanski Hall, room/local 102

Abstracts/Résumés

YING YAN & GRACE Y. YI (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 13:15]

Methods for Additive Hazards Models with Covariate Measurement Error

Méthodes pour les modèles à risques additifs avec erreur de mesure dans les covariables

In this talk, we propose several consistent estimators for additive hazards model with mismeasured covariates. Starting with the bias analysis of the naive method that ignores measurement error, we establish an explicit relationship between the asymptotic limit of the naive estimator and the true estimator. This analysis not only shows that ignoring covariate measurement error would yield biased result, but also provides a correction method to the naive estimator. Different from this “estimator-correction” method, we explore several other methods that are derived based on various formulations of unbiased estimating functions. These methods are justified both theoretically and empirically.

Dans cet exposé, nous proposons plusieurs estimateurs convergents pour les modèles à risques additifs avec mesures erronées des covariables. En commençant par analyser le biais de la méthode naïve qui ignore l’erreur de mesure, nous établissons un lien explicite entre la limite asymptotique de l’estimateur naïf et l’estimateur réel. D’une part, cette analyse montre que des résultats biaisés sont obtenus en ignorant l’erreur de mesure des covariables et, d’autre part, elle fournit une méthode pour corriger l’estimateur naïf. En plus de cette méthode de “correction de l’estimateur”, nous explorons plusieurs autres méthodes dérivées de diverses formules de fonctions d’estimation non biaisées. Ces méthodes sont justifiées de façon théorique et empirique.

XUE YAO, LISA LIX & LIU JUNXIN (University of Saskatchewan)

[Tuesday June 5 / mardi 5 juin, 13:30]

Multiple Imputation to Correct Measurement Error: Application to Chronic Disease Case Ascertainment in Administrative Databases

Imputation multiple pour la correction de l’erreur de mesure : application à la confirmation des cas de maladie chronique dans les bases de données administratives

Measurement error can be treated as a missing data problem to which multiple imputation (MI) methods can be applied. We use MI methods to correct for measurement error in response and covariates of a predictive model and apply them to accurately ascertain disease status in administrative health databases. Using simulation, we investigate the effects of magnitude of misclassification in the response, amount of measurement error in model covariates, proportion of missing data and number of imputations on performance of MI meth-

L’erreur de mesure peut être considérée comme un problème de données manquantes auquel des méthodes d’imputation multiple (IM) peuvent être appliquées. Nous utilisons ces méthodes d’IM pour corriger l’erreur de mesure de variables réponses et de covariables d’un modèle prédictif pour les appliquer à l’établissement précis du statut de maladies dans les bases de données administratives en santé. À l’aide de simulations, nous examinons les effets de l’ampleur du classement erroné dans la réponse, de la magnitude de l’erreur de mesure dans les covariables, de la proportion de valeurs manquantes et du nombre d’imputations

2C-C1: Models Accounting for Measurement Error Modèles avec erreurs de mesure

ods. Case ascertainment rates were substantially improved, but precision depended on the prediction model characteristics and number of imputations.

sur la performance des méthodes d'IM. Les taux de confirmation des cas sont considérablement améliorés, mais la précision dépend des caractéristiques du modèle de prédiction et du nombre d'imputations.

TARANEH ABARIN, BRAJENDRA SUTRADHAR & GUANG SUN (Memorial University of Newfoundland)

[Tuesday June 5 / mardi 5 juin, 13:45]

Gene-Environment Interaction Studies with Measurement Error

Études d'interaction gène-environnement avec erreur de mesure

It is now well-known that obesity leads to a number of metabolic and chronic diseases that affect a large portion of the population. This complex trait is determined by multiple genetic and environmental factors that interact with one another in complicated ways. The existing studies examine such factors under the assumption that they are measured accurately. However, unobserved or error-prone environmental factors and/or misclassification in genotyping are unavoidable. Motivated by the Complex Diseases in the Newfoundland population: Environment and Genetics (CODING) study, we propose a Generalized Quasi-Likelihood estimation method when both environmental and genetic factors are subject to error.

Il est bien connu que l'obésité engendre de nombreuses maladies métaboliques et chroniques qui affectent une grande partie de la population. Ce trait complexe est déterminé par plusieurs facteurs génétiques et environnementaux qui interagissent mutuellement de manière complexe. Les études existantes examinent ces facteurs sous l'hypothèse qu'ils sont mesurés de façon précise. Cependant, des facteurs environnementaux non-observés ou liés à des erreurs et/ou à de mauvaises classifications en génotypage sont inévitables. Motivés par la complexité des maladies de la population de Terre-Neuve dans l'étude environnementale et génétique "CODING", nous proposons une estimation par la quasi-vraisemblance généralisée lorsque les facteurs environnementaux et génétiques sont tous deux sujets à l'erreur.

WEI LIU (York University), **LANG WU** (University of British Columbia)

[Tuesday June 5 / mardi 5 juin, 14:00]

Simultaneous Inference for Semiparametric Nonlinear Mixed-effects Models with Covariate Measurement Errors and Missing Responses

Inférence simultanée pour modèles semi-paramétriques non-linéaires à effets mixtes avec erreur de mesure sur les covariables et données manquantes

Semiparametric nonlinear mixed-effects (NLME) models are flexible for modeling complex longitudinal data. Covariates are usually introduced in the models to partially explain inter-individual variations. Some covariates, however, may be measured with substantial errors. Moreover, the responses may be missing and the missingness may be nonignorable. We propose two approximate likelihood methods for semiparametric NLME models with covariate measurement errors and nonignorable missing responses. The methods are illustrated in a real data example. Simulation results show that both methods perform well and are much better than the commonly used naive method.

Les modèles semi-paramétriques non-linéaires à effet mixtes (NLEM) sont flexibles pour la modélisation de données longitudinales complexes. Des covariables sont habituellement incluses dans le modèle pour expliquer une partie de la variation entre les individus. Certaines covariables sont toutefois mesurées avec des erreurs substantielles. De plus, certaines variables dépendantes sont parfois manquantes et la non-réponse peut ne pas être ignorable. Nous proposons deux méthodes de vraisemblance approximée pour des modèles semi-paramétriques NLEM avec erreur de mesure sur les covariables et données manquantes. Les méthodes sont illustrées à l'aide d'un jeu de données réelles. Des résultats de simulation montrent que les deux méthodes performant bien et sont grandement supérieures aux méthodes naïves

2C-C1: Models Accounting for Measurement Error Modèles avec erreurs de mesure

populaires.

ANNE MICHELE MILLAR & REBECCA ROMO (Mount Saint Vincent University), **BETHANY WHITE** (University of Western Ontario), **CANDACE SCHAU** (CS Consultants, LLC, New Mexico)

[Tuesday June 5 / mardi 5 juin, 14:15]

Measuring Students' Attitudes to Statistics in the Presence of Regression-to-the-Mean

Mesurer l'attitude d'étudiants envers la statistique en présence de régression à la moyenne

Measuring change in students' attitudes from the beginning of a statistics course to the end is not as simple as it might appear. Typically, the gain in attitude (post-score – pre-score) depends on the pre-score: negative attitudes tend to have a higher gain score, while more positive attitudes tend to have a lower one. Since pre/post-scores include measurement error, the “true” relationship between the gain and the pre-score is difficult to estimate due to regression-to-the-mean. We compare methods for parameter estimation and confidence interval construction to accurately estimate the change in attitude for a given pre-score, allowing for measurement error.

Mesurer le changement d'attitude des étudiants entre le début et la fin d'un cours de statistique n'est pas aussi simple qu'il peut sembler. Habituellement, l'amélioration de l'attitude (post-test – pré-test) dépend du résultat au départ : les attitudes négatives ont tendance à avoir de plus grands gains, alors que les attitudes positives en ont de plus faibles. Puisque ces résultats sont sujets à une erreur de mesure, la vraie relation entre le gain et le résultat au départ est difficile à estimer à cause du phénomène de la régression à la moyenne. Nous comparons des méthodes pour l'estimation de paramètres et la construction d'intervalles de confiance pour estimer de façon exacte le changement d'attitude pour un certain résultat avant le cours, en tenant compte de l'erreur de mesure.

2C-C2: Nonparametric Methods Méthodes non-paramétriques

Chair/Présidente: Elif Fidan Acar (McGill University)

Rozanski Hall, room/local 105

Abstracts/Résumés

MUHAMMAD ISLAM (Grant MacEwan University)

[Tuesday June 5 / mardi 5 juin, 13:15]

Comparing Distributions Using Data Driven Statistics

Comparaison de distributions à l'aide de tests adaptatifs

When comparing two distributions, tests for location do not give any information on how the two distributions disagree on other characteristics such as variance, skewness or kurtosis when the null hypothesis is rejected. Data driven smooth test statistics and empirical distribution function test statistics can often be decomposed into component statistics reflecting the differences between the distributions. In this talk I will present an informative statistical analysis using data driven smooth tests and EDF tests with gene expression data.

Lorsqu'on compare deux distributions, le rejet de l'hypothèse nulle d'un test sur le paramètre de position ne fournit aucune information sur la façon dont les distributions diffèrent selon d'autres caractéristiques telles que la variance, l'asymétrie ou l'aplatissement. Les tests statistiques lisses adaptatifs et les tests basés sur la fonction de répartition empirique peuvent souvent être morcelés en plusieurs composantes statistiques qui reflètent les différences entre les distributions. Dans cet exposé, je vais présenter une analyse statistique sur des données d'expression de gènes à l'aide de tests lisses adaptatifs et de tests basés sur la fonction de répartition empirique.

THIERRY DUCHESNE, LAJMI LAKHAL-CHAIEB & BELKACEM ABDOUS (Université Laval)

[Tuesday June 5 / mardi 5 juin, 13:30]

Estimation of the Conditional Survival Function Under Dependent Censoring

Estimation de la fonction de survie conditionnelle en présence de censure dépendante

In this talk we consider a nonparametric estimation method for the conditional survival function when the covariate is continuous and when the response is subject to dependent right censoring. This occurs, for instance, when the response and covariate are parallel failure times, successive failure times, or semi-competing risks. The method that we propose is based on local polynomial smoothing and yields valid estimators in all the contexts listed above. We show how the bandwidth parameter can be efficiently selected with the double-kernel method. We apply the proposed methodology to the analysis of a few real datasets.

Dans cet exposé nous nous intéressons à l'estimation non-paramétrique de la fonction de survie conditionnelle avec covariable continue et réponse sujette à une censure à droite dépendante. Ceci peut survenir lorsque la réponse et la covariable sont des temps de défaillance parallèles, des temps de défaillance successifs ou des risques semi-concurrents. La méthode que nous proposons est basée sur le lissage polynomial local et donne des estimateurs valides dans les situations ci-dessus. Nous montrons comment la fenêtre de lissage peut être choisie efficacement par la méthode du double noyau. Nous appliquons cette méthodologie à l'analyse de jeux de données réels.

2C-C2: Nonparametric Methods Méthodes non-paramétriques

MEHRDAD GANGEH, ALI GHODSI & MOHAMED KAMEL (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 13:45]

Kernelized Supervised Dictionary Learning

Méthode du noyau pour l'apprentissage supervisé par dictionnaire

Signals can usually be represented by a linear combination of some predefined (such as wavelets) or learned (called dictionary) bases. However, the response variables are not usually taken into account on computing the learned bases, i.e. dictionary is computed unsupervised. We propose a supervised approach in learning dictionary by incorporating the information on class labels. To this end, a transformation is introduced on the data such that the dependency between the data and their corresponding labels is maximized. One of the main advantages of this novel approach is that it can be easily kernelized by incorporating a kernel into the transformation.

Les signaux peuvent être représentés par une combinaison linéaire de certaines bases prédéfinies (comme les ondelettes) ou apprises (formant un dictionnaire). Cependant, les variables dépendantes ne sont normalement pas utilisées dans le calcul des bases apprises, i.e. le dictionnaire est calculé de façon non supervisée. Nous proposons une approche supervisée pour l'apprentissage du dictionnaire en incorporant de l'information sur des étiquettes de classes. À cette fin, les données sont transformées pour que leur dépendance avec les étiquettes correspondantes soit maximisée. L'un des principaux avantages de cette nouvelle approche est que l'on peut facilement incorporer un noyau dans la transformation.

MOHAMMAD NOURMOHAMMADI, MOHAMMAD JAFARI JOZANI & BRAD JOHNSON (University of Manitoba)

[Tuesday June 5 / mardi 5 juin, 14:00]

Nonparametric Confidence Intervals of Quantiles of a Finite Population using Randomized Nomination Sampling

Intervalle de confiance non-paramétriques pour quantiles dans une population finie à l'aide de l'échantillonnage randomisé par nomination

We develop nonparametric confidence intervals for quantiles of finite populations based on randomized maxima-minima nomination sampling (RMMNS) design. Three variations of the RMMNS design are discussed and recursive algorithms are proposed for computing the coverage probabilities of their corresponding confidence intervals. To achieve (approximate) coverage probabilities of interest interpolated symmetric intervals are proposed. The expected length of the intervals is assessed and compared with the corresponding one based on simple random sampling (SRS). We show that, symmetric nonparametric confidence intervals under RMMNS design are shorter than their corresponding ones under SRS. Theoretical results are augmented by numerical and simulation studies.

Nous développons des intervalles de confiance non-paramétriques pour les quantiles de populations finies basés sur un plan randomisé par nomination de minima et maxima (RMMNS). Trois variantes du plan RMMNS sont présentées et des algorithmes récursifs sont proposés pour calculer la probabilité de couverture de leurs intervalles de confiance. Pour obtenir la couverture (approximative) voulue, des intervalles symétriques interpolés sont proposés. L'espérance de la longueur des intervalles est calculée et comparée avec celle des intervalles basés sur l'échantillonnage aléatoire simple (EAS). Nous montrons que les intervalles de confiance symétriques non-paramétriques sont plus courts sous le plan RMMNS que sous EAS. Les résultats théoriques sont accompagnés de résultats numériques et de simulations.

2C-C3: Spatial Statistics Statistique spatiale

Chair/Président: Douglas Woolford (Wilfrid Laurier University)

Thornbrough Building, room/local 1307

Abstracts/Résumés

OLAF BERKE (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 13:15]

The 350th Anniversary of Statistical Epidemiology: A Spatial Re-analysis of John Graunt's London Plague Data
350e anniversaire de l'épidémiologie statistique : analyse spatiale des données sur la peste de Londres de John Graunt

John Graunt was dubbed father of modern statistics and founder of modern epidemiology for his Observations about the Bills of Mortality published in 1662. A major objective for Graunt was the comparison of plague mortality at parish level. His data were re-analyzed using modern spatial epidemiological methods. Comparisons were based on the proportional mortality ratio (PMR). Choropleth mapping was used to visualize the spatial distribution of the PMR. Moran's I does not indicate clustering. The spatial scan test detected a high-risk cluster. Logistic regression identifies population density and poverty as risk factors.

John Graunt a été surnommé le père de la statistique et de l'épidémiologie modernes pour ses observations sur les listes de décès publiées en 1662. Son objectif principal était de comparer les taux de mortalité par paroisse. Nous avons ré-analysé ses données à l'aide de méthodes épidémiologiques spatiales modernes, en basant nos comparaisons sur la mortalité proportionnelle (MP). Une cartographie choroplèthe permet de visualiser la distribution spatiale de la MP. L'indice de Moran n'indique aucun groupage. Le balayage spatial détecte un groupe à haut risque. La régression logistique identifie la densité démographique et la pauvreté comme facteurs de risque.

MASUD RANA, SHAHEDUL A. KHAN & LONGHAI LI (University of Saskatchewan), **JOEL A. DUBIN** (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 13:30]

Monitoring and Understanding Atmospheric CFC Decline with the Spatial-Longitudinal Bent-Cable Model
Surveillance et compréhension du déclin des CFC atmosphériques par le modèle "bent-cable" longitudinal

The recent steady decline in atmospheric concentration of chlorofluorocarbon (CFC) could be a direct result of the Montreal Protocol's ban on CFC products. CFC has long atmospheric residence times, and as a consequence has been distributed globally. The bent cable is a regression model to study data that exhibits change. To study the spatial effects and extent of the decline, we extend the existing longitudinal bent-cable regression in a hierarchical Bayesian framework by allowing the error terms to be correlated across space. The methodology is illustrated with applications to CFC data obtained from several measuring stations across the globe.

La diminution récente et constante de la concentration atmosphérique des chlorofluorocarbones (CFC) est peut-être le résultat direct de l'interdiction des CFC par le Protocole de Montréal. Les CFC, qui séjournent longtemps dans l'atmosphère, sont distribués à l'échelle mondiale. Le "bent-cable" est un modèle de régression permettant l'étude de données évolutives. Pour étudier les effets spatiaux et l'étendue du déclin, nous élargissons la régression "bent-cable" longitudinale existante dans un cadre bayésien hiérarchique en permettant une corrélation spatiale des termes d'erreur. Nous illustrons notre méthodologie par des applications aux données sur les CFC obtenues de diverses stations de mesure dans le monde.

2C-C3: Spatial Statistics Statistique spatiale

JONATHAN LEE, REG KULPERGER & HAO YU (University of Western Ontario)

[Tuesday June 5 / mardi 5 juin, 13:45]

Large-scale Spatial Simulation and Computation Using Parallel Computing
Simulation spatiale à grande échelle et calcul parallèle

Spatial simulation models such as stochastic lattice models as well as spatial descriptive statistics such as Ripley's K-function have wide applicability in spatial statistics but in general do not scale well for large datasets. Parallel computing (high performance computing) is one solution that can provide almost linear scalability to these applications. Implementation, optimizations, and complications that arise from parallelizing existing algorithms will be discussed. Two R packages that implement the discussed methods will be introduced.

Les modèles de simulation spatiale tels que les modèles stochastiques en treillis et les statistiques descriptives spatiales telle que la fonction K de Ripley ont beaucoup d'applications en statistique spatiale, mais s'adaptent mal à de grands jeux de données. Les calculs parallèles (calculs de haute performance) permettent d'obtenir une augmentation presque simplement linéaire de la quantité de calcul pour ces applications. L'implémentation, l'optimisation et les complications causées par la parallélisation des algorithmes existants seront abordés. Deux bibliothèques R employant les méthodes discutées seront présentées.

ALISHA ALBERT-GREEN (Princess Margaret Hospital), **CHARMAINE DEAN** (University of Western Ontario), **DOUGLAS WOOLFORD** (Wilfrid Laurier University), **DAVID MARTELL** (University of Toronto)

[Tuesday June 5 / mardi 5 juin, 14:00]

A Nonparametric Framework for Quantifying Temporal Trends in the Seasonality of Forest Fire Risk
Cadre non-paramétrique de quantification des tendances temporelles de la saisonnalité du risque d'incendies forestiers

Investigating the seasonality of forest fires and how this is changing over time is of interest to fire managers and researchers. We develop flexible models for describing the temporal variation in lightning-caused fire risk. These models produce daily smoothed estimates of fire risk when fit to historical records from Alberta and Ontario. Inverse calculations provide interval estimates for the start and end of the fire season each year, defined by the crossing of a risk threshold. Permutation-based methods are employed to test for significant trends. Results suggest changes to the timing of the fire season in Alberta, but not Ontario.

Les gestionnaires d'incendies et les chercheurs s'intéressent à la saisonnalité des incendies forestiers et à l'évolution de celle-ci. Nous élaborons des modèles souples pour décrire la variation temporelle du risque d'incendies allumés par la foudre. Ces modèles produisent des estimations quotidiennes lissées du risque d'incendie, qui correspondent aux données historiques pour l'Alberta et l'Ontario. Par calcul inverse, nous estimons des intervalles pour le début et la fin de chaque saison des incendies, définies par le franchissement d'un seuil de risque. Des méthodes de permutation permettent de rechercher des tendances significatives. Nos résultats indiquent que la saison des feux a évolué en Alberta, mais pas en Ontario.

DEVAN BECKER & DOUGLAS WOOLFORD (Wilfrid Laurier University), **DAVID MARTELL** (University of Toronto)

[Tuesday June 5 / mardi 5 juin, 14:15]

Exploring the Lifetime of Forest Fires
Exploration de la durée de vie des incendies forestiers

During fire events, there are often tough decisions to be made concerning the optimal allocation of resources for fire management. A small fire that

Pendant les incendies, il faut souvent prendre des décisions difficiles concernant la répartition optimale des ressources. Un petit feu sous-estimé peut devenir catastrophique, alors

2C-C3: Spatial Statistics Statistique spatiale

is underestimated can become potentially catastrophic, and a large fire that is overestimated can become a drain on resources that may be used more efficiently elsewhere. In this project, we will investigate how statistical methods can be used to model the future lifetimes of forest fires by analyzing data from historical forest fire records. This talk presents exploratory work along with some preliminary modelling results.

qu'un grand incendie surestimé peut épuiser des ressources qui pourraient être utilisées plus efficacement ailleurs. Dans ce projet, nous étudions comment utiliser des méthodes statistiques pour modéliser les durées de vie d'incendies en analysant des données historiques. Nous présentons ici nos travaux exploratoires et quelques résultats préliminaires de modélisation.

2C-C4: Studying Gene Expression/Association Étude de l'expression et de l'association des gènes

Chair/Présidente: Shelley Bull (Samuel Lunenfeld Research Institute)

MacNaughton Building, room/local 113

Abstracts/Résumés

FARNOOSH ABBAS AGHABABAZADEH, MAYER ALVO & DAVID R. BICKEL (University of Ottawa)

[Tuesday June 5 / mardi 5 juin, 13:15]

Statistical Methods to Analyze Genome-Wide Association Data

Méthodes statistiques pour analyser des données d'association sur tout le génome

The theory of this research project is to identify the single nucleotides (SNPs) that are associated with disease. We applied our methods to coronary artery disease data in order to estimate the local false discovery rate (LFDR) and computed the proportion of SNPs associated with disease (π_1). We are interested in determining appropriate reference classes to improve the proportion of SNPs associated with disease. Therefore, we considered two reference classes to estimate the LFDR of low frequency SNPs: (1) low-frequency SNPs and (2) both common and low-frequency SNPs together, and we present considerable results based on bias and variance properties.

La théorie de ce projet de recherche permet d'identifier les polymorphismes d'un seul nucléotide qui sont associés à une maladie. Nous avons appliqué notre méthode à des données de maladies coronariennes afin d'estimer le taux local de fausses découvertes et avons calculé la proportion des polymorphismes d'un seul nucléotide associés à la maladie (π_1). Nous cherchons à déterminer les classes de référence appropriées pour améliorer la proportion de ces polymorphismes associés à la maladie. Ainsi, nous avons considéré deux classes de référence pour estimer le taux local de fausses découvertes des polymorphismes d'un seul nucléotide à basse fréquence : (1) les polymorphismes d'un seul nucléotide à basse fréquence et (2) les polymorphismes d'un seul nucléotide commun et ceux à basse fréquence. Nous présentons des résultats importants basés sur les propriétés de biais et de variance.

GUN HO JANG & RUI FENG (University of Pennsylvania)

[Tuesday June 5 / mardi 5 juin, 13:30]

A Statistical Model for Copy Number Polymorphism Calling Using Haplotypes

Un modèle statistique pour l'identification du nombre de copies de polymorphismes à l'aide d'haplotypes

Both single nucleotide polymorphism (SNP) and copy number variation (CNV) are widespread characteristic of the human genome that have been found to be related to various diseases. Limited by the available statistical methods, SNP genotypes and CNVs were often called separately on common genotyping platforms. Few methods are available for integrating SNP and CNV information. We propose a haplotype-based maximum likelihood method for simultaneously calling SNP and CNV and incorporating linkage disequilibrium information among multiple markers in the population.

Les polymorphismes de nucléotide simples (SNP) et la variation du nombre de copies (CNV) sont des caractéristiques répandues du génome humain qui ont été associées à plusieurs maladies. Dû à l'absence de méthodes statistiques appropriées, les SNP génotype et CNVs étaient souvent obtenus séparément sur des plateformes de génotypage communes. Peu de méthodes sont disponibles pour intégrer l'information des SNP et des CNV. Nous proposons une méthode de maximum de vraisemblance basée sur l'haplotype pour identifier les SNP et CNV simultanément et incorporer l'information sur le déséquilibre des liaisons entre plusieurs marqueurs dans la population. À l'aide de simulations, nous

2C-C4: Studying Gene Expression/Association Étude de l'expression et de l'association des gènes

Through simulation studies, we demonstrated our model is more sensitive and accurate in detecting CNV regions, compared with commonly-used CNV calling methods.

démontrons que notre modèle est plus sensible et plus précis pour détecter les régions de CNV par rapport à d'autres méthodes courantes de détection de CNV.

AHMED HOSSAIN (McMaster University), **ANDREW WILLAN** (University of Toronto), **JOSEPH BEYENE** (McMaster University)

[Tuesday June 5 / mardi 5 juin, 13:45]

A Flexible Nonparametric Approach for Identifying Correlated Differentially Expressed Genes in Microarray Experiments

Une approche non-paramétrique flexible pour identifier des gènes corrélés et exprimés différemment dans des expériences de micropuces

Finding genes in the same biological pathway corresponding to a particular gene may enhance further understanding of its biological function. We are interested in finding other candidate genes whose expression values are highly correlated with that of a "seed" gene. The "seed" gene, which is associated to a disease, is used as a reference to extract candidate genes. We propose a nonparametric test statistic that compares two Area Under Receiver Operating Characteristic Curves (AUC) for gene pairs taking implicit correlation between two AUCs into account. The performance of our method is investigated through simulation and real data analysis.

Identifier les gènes dans le même parcours biologique correspondant à un gène particulier pourrait améliorer notre compréhension du fonctionnement biologique de ce gène. Nous sommes intéressés à trouver d'autres gènes candidats dont les valeurs d'expression sont fortement corrélées avec celle d'un gène souche. Le gène souche, qui est associé à une maladie, est utilisé comme référence pour obtenir d'autres gènes candidats. Nous proposons un test statistique non-paramétrique qui compare l'aire sous la courbe sensibilité/spécificité (courbe ROC) pour des paires de gènes en tenant compte de façon implicite de la corrélation entre les deux courbes ROC en question. La performance de notre méthode est évaluée à l'aide de simulations et de l'analyse d'un jeu de données.

THIERRY CHEKOUGO TEKOUANG & ALEJANDRO MURUA (Université de Montréal)

[Tuesday June 5 / mardi 5 juin, 14:00]

The Gibbs-plaid Biclustering Model

Le modèle Gibbs-plaid de biclustering

Within the context of gene expression, biclustering refers to the simultaneous clustering of genes and experimental conditions. In the literature, most of the biclustering algorithms consider genes or conditions as statistically independent entities. We propose a Bayesian plaid model for biclustering that accounts for the prior dependency between the genes or conditions through a relational graph. Our model assumes that the graph is governed by a Gibbs random field. We developed an algorithm partly based on the Wang-Landau flat-histogram algorithm in order to estimate the posterior distribution of the bicluster membership. We show some experiments with real and simulated data.

Dans le contexte d'expression génétique, un biclustering réfère au groupement simultané de gènes et de conditions expérimentales. Dans la littérature, la plupart des algorithmes de biclustering considèrent les gènes (ou conditions) comme des entités indépendantes. Nous proposons un modèle de plaid bayésien avec la dépendance a priori entre les gènes (ou conditions) à travers un graphe. Notre modèle suppose que le graphe est gouverné par un champ aléatoire de Gibbs. Nous développons un algorithme partiellement basé sur celui de Wang-Landau pour estimer la distribution a posteriori des étiquettes de biclusters. Nous montrons quelques expériences sur des données simulées et réelles.

2C-C4: Studying Gene Expression/Association Étude de l'expression et de l'association des gènes

MAGGIE CHU & WENQING HE (University of Western Ontario)

[Tuesday June 5 / mardi 5 juin, 14:15]

Model-Based Clustering of Time-Course RNA-seq Data

Regroupement modélisé de données ARN-seq sur les temps d'absorption

The next generation sequencing technology (RNA-seq) provides absolute quantification of gene expression using counts of read. Limited work has been done on expression analysis of longitudinal RNA-seq data to account for the time-dependence nature of the count data with overdispersion property. Functional clustering is an important method for examining gene expression patterns and thus discovering co-expressed genes to better understand the biological systems. We propose a model-based clustering method for identifying gene expression patterns using time-course RNA-seq data. The performance of the proposed method is assessed through simulation. A time-course genomic dataset is employed for illustration.

Les dernières technologies de séquençage (ARN-seq) permettent une quantification absolue de l'expression génique par comptage de lectures. Des travaux limités ont étudié l'analyse de l'expression des données longitudinales ARN-seq pour tenir compte de la dépendance temporelle des données de comptage surdispersées. Le regroupement fonctionnel est une méthode d'examen des configurations d'expression génique permettant de découvrir des gènes coexprimés et mieux comprendre les systèmes biologiques. Nous proposons une méthode de groupement modélisée qui permet d'identifier les configurations d'expression génique en utilisant des données ARN-seq sur les temps d'absorption. Nous évaluons sa performance par simulation et l'illustrons sur un ensemble de données génomiques de temps d'absorption.

MARTIN LECLERC & LAJMI LAKHAL-CHAIEB (Université Laval)

[Tuesday June 5 / mardi 5 juin, 14:30]

Estimation of Dependence Between Lifetimes in the Presence of Selection Bias: the Case of Breast Cancer

Estimation de la dépendance entre des durées de vie en présence de biais de sélection : le cas du cancer du sein

Analysis of the association between SNPs and breast cancer in women who carry a mutation in the BRCA1 or BRCA2 genes is characterized by the presence of correlated lifetimes and selection bias. In this presentation, a new approach is developed to estimate the dependence by using a Clayton copula model of varying size and a weighted Nelson-Aalen estimator. The approach was tested through simulation. Application to data from American and European women shows the potential of the method for studying breast cancer risk.

L'analyse de l'association entre des SNPs et le cancer du sein chez les porteuses de mutations sur les gènes BRCA1 ou BRCA2 est caractérisée par la présence de durées de vie corrélées et d'un biais de sélection. Dans cette présentation, une nouvelle approche est développée pour l'estimation de la dépendance au moyen d'une copule de Clayton de dimension variable et d'un estimateur de Nelson-Aalen pondéré. L'approche a été testée au moyen de simulations et une application à des données provenant de femmes américaines et européennes illustre le potentiel de la méthode pour l'étude du risque de cancer du sein.

Science Complex Atrium

Abstracts/Résumés

JUSTIN ANGEVAARE & DANIEL GILLIS (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 13:15]

The Utility of Catch Per Unit Effort Variance

Utilité de la variance des captures par unité d'effort

Population models play a central role in fisheries management. Commercial harvest data, especially catch per unit effort (CPUE), are used to predict population parameters and abundance. A survey of recent literature reveals that CPUE variability is typically ignored. The common practice is to aggregate harvest and effort by year. We present a simulation study to explore traditional models and those that have been modified to include CPUE variability. Daily fish populations and harvest events are simulated over 60 years. Catch and effort are aggregated over several temporal scales. Population level parameter estimates are compared across models and temporal aggregation scales.

Les modèles de population jouent un rôle central dans la gestion des pêches. Les données sur la pêche commerciale, particulièrement les captures par unité d'effort (CPUE), sont utilisées pour prédire les paramètres et l'abondance des populations. Une revue de la littérature récente révèle que la variabilité des CPUE est généralement ignorée. La pratique courante consiste à regrouper les données de capture et d'effort sur une base annuelle. Nous présentons une étude de simulation qui examine les modèles traditionnels et ceux qui ont été modifiés pour y intégrer la variabilité des CPUE. Les populations de poissons et les événements de pêche sont simulés sur une base quotidienne pour une période de 60 ans. Les données de capture et d'effort sont regroupées en fonction de plusieurs échelles temporelles. Les estimés des paramètres de population sont comparés entre les différents modèles et échelles de regroupement temporelles.

SHAHEENA BASHIR (University Health Network, Toronto), **CELIA GREENWOOD** (Lady Davis Institute for Medical Research and McGill University / Lady Davis Institute for Medical Research et McGill University)

[Tuesday June 5 / mardi 5 juin, 13:15]

Joint Tests of Effect Size in Multivariate Meta-Analysis

Tests simultanés de l'ampleur de l'effet dans une méta-analyse multivariée

Multivariate meta-analysis combines estimates of several related parameters over several studies. This work investigates the feasibility of obtaining an F-statistic that can test for differences among more than two groups in a meta-analysis context. The question of obtaining a general F-statistic arose in a meta-analysis of gene expression microarray data with 4 treatment groups and 6 possible pairwise comparisons across 2 studies. We evaluate the performance of our test statistic in comparison to standard random effect meta-analysis approaches in simulations. The simulations show that the quadratic form of the pairwise t statistic from random effect model is distributed as F.

La méta-analyse multivariée combine les estimateurs de plusieurs paramètres reliés dans plusieurs études. Ce travail examine la possibilité d'obtenir une statistique F qui peut tester les différences parmi plus de deux groupes dans un contexte de méta-analyse. Cette question s'est posée dans une méta-analyse des structures d'expression génétique émergeant de données comportant quatre groupes de traitements et six comparaisons possibles par paires dans deux études. Nous évaluons la performance de notre test statistique en le comparant, au moyen de simulations, à des approches standard en matière de méta-analyse à effets aléatoires. Les simulations démontrent que la forme quadratique de la statistique t par paire du modèle à effets aléatoires suit une distribution de Fisher.

2C-P: Poster Session II Séance d'affichage II

VICTORIA BORG DEBONO, SHIYUAN (WILLIAM) ZHANG, CHENGLIN YE & JAMES PAUL (McMaster University),
LINDSAY HURLBURT (University of Toronto), **YAMINI MURTHY & LEHANA THABANE** (McMaster University)

[Tuesday June 5 / mardi 5 juin, 13:15]

The Quality of Reporting of Randomized Controlled Trials (RCTs) in Postoperative Pain Management

Qualité des rapports sur les essais contrôlés randomisés (ECR) dans la gestion de la douleur postopératoire

Good reporting of RCTs is important because it is a proxy for researchers to appraise RCT methodological rigor. The aims of this study were to assess the quality of reporting (QOR) in postoperative pain management RCTs – using the CONSORT criteria – and to determine factors associated with high QOR. Overall, the QOR of anesthesia trials was found to be poor to moderate, results which are consistent with observations from other clinical disciplines. The information gained from this can be used by journals to register the urgency for enhancing the reporting to make RCT findings accessible and comparable.

Les rapports de qualité sur les ECR sont importants puisqu'ils permettent aux chercheurs d'évaluer la rigueur méthodologique des ECR. Cette étude vise à évaluer la qualité des rapports (QR) sur les ECR dans la gestion de la douleur postopératoire à l'aide du critère CONSORT, et à déterminer les facteurs associés à une QR élevée. Globalement, il a été constaté que la QR sur les essais en anesthésie était faible à modérée, et ces résultats concordent avec les observations d'autres disciplines cliniques. Les renseignements ainsi obtenus peuvent être utilisés par des revues pour marquer l'urgence d'améliorer la présentation des résultats dans les rapports en vue de rendre les conclusions des ECR accessibles et comparables.

ZHENGFEI CHEN (University of Toronto)

[Tuesday June 5 / mardi 5 juin, 13:15]

Riemann Manifold HMCMC for Log-Gaussian Cox Processes

MCMC hamiltonien sur des variétés de Riemann pour des processus de Cox log-gaussiens

Log Gaussian Cox Point Processes are a rich class of models for clustered point patterns, and are useful and popular for location data which is distributed inhomogeneously. Riemann Manifold Hamiltonian Monte Carlo method produce highly efficient sampling even in very high dimensions where different scaling may be required for the transient and stationary phases of the Markov chain. Some nice properties LGCP with Matern covariance using HMCMC will be explored, along with an application of RMHMC in LGCP with Matern covariance matrix on a high dimensional data set.

Les processus de Cox log-gaussiens (PCLG) ponctuels forment une riche classe de modèles pour les motifs ponctuels en grappes. Ils sont utiles et populaires pour les données de localisation distribuées de façon inhomogène. La méthode de Monte Carlo hamiltonienne sur des variétés de Riemann (MMCHVR) produit un échantillonnage hautement efficace même dans des cas de très haute dimensionnalité, où différents ajustements d'échelle sont nécessaires lors des phases transitoires et stationnaires de la chaîne de Markov. Nous examinons de belles propriétés des PCLG avec une covariance de Matérn utilisant le MCMC hamiltonien (MCMCH), de même que l'application de la MMCHVR à des PCLG avec une matrice de covariance de Matérn sur un jeu de données de haute dimensionnalité.

JING DING (Acadia University)

[Tuesday June 5 / mardi 5 juin, 13:15]

Population Assessments by Using Age-at-harvest Data and Catch Effort

Évaluations des populations au moyen des données sur l'âge des prises et l'effort de pêche

Age-at-harvest data and catch effort are readily available quantities in wildlife management. They

Les données sur l'âge des prises et l'effort de pêche sont des quantités facilement accessibles dans la gestion de la

2C-P: Poster Session II Séance d'affichage II

provide essential information for modelling populations so as to reveal the situation of animals under exploitation. Three practicable methods are introduced. Downing (1980)'s reconstruction is a commonly used technique which uses backward addition through cohorts to reconstruct populations over time. A linear regression model incorporating catch effort, developed by Paloheimo (1980), is another applicable method. Stochastic catch effort method (Dupont, 1983) is the most advanced method since it requires weaker assumptions about the birth-death-capture process. Hazard models are assumed, and ordinary least-squares can be used for parameter estimation.

faune. Elles fournissent des renseignements essentiels pour la modélisation des populations en vue de révéler la situation des espèces exploitées. Nous introduisons trois méthodes applicables. La reconstruction de Downing (1980) est une technique utilisée couramment qui consiste en des ajouts antéchronologiques aux cohortes afin de procéder à la reconstruction des populations dans le temps. Une autre méthode applicable élaborée par Paloheimo (1980) consiste en un modèle de régression linéaire incorporant l'effort de pêche. La méthode de l'effort de pêche stochastique (Dupont, 1983) constitue la méthode la plus avancée, puisqu'elle requiert des hypothèses plus faibles sur le processus de naissance-mort-capture. Des modèles de risques sont définis et les moindres carrés ordinaires peuvent être utilisés pour en estimer les paramètres.

JAYDE EUSTACE & J. CONCEPCIÓN LOREDO-OSTI (Memorial University of Newfoundland)

[Tuesday June 5 / mardi 5 juin, 13:15]

A Gumbel State Space Model for Extreme-value Data

Modèle à espace d'états de Gumbel pour des valeurs extrêmes

Extreme value theory has many applications in areas such as meteorology, finance and engineering. Recent years have seen a revival of interest in the analysis of series of extreme-value data. For this presentation, some properties of the mixtures of Gumbel and α -stable random variables will be used to construct a state-space model with Gumbel marginals. One feature of α -stable random variables is that their density cannot be expressed in closed form and their expectation does not exist. This poses some inference challenges. The methods proposed will be compared through simulation to currently used models using real data.

La théorie des valeurs extrêmes offre plusieurs applications dans des domaines tels que la météorologie, la finance et l'ingénierie. Au cours des dernières années, l'intérêt a repris concernant l'analyse de séries de valeurs extrêmes. Dans cet exposé, certaines propriétés des mélanges de Gumbel et des variables aléatoires α -stables sont utilisées pour construire un modèle à espace d'états avec des lois marginales de type Gumbel. La densité des variables aléatoires α -stables ne peut pas être exprimée sous forme analytique et leur espérance n'existe pas. Cet aspect pose des défis en matière d'inférence. Les méthodes proposées sont comparées aux modèles actuels par des simulations avec des données réelles.

REMYA M. JOHN & ANGELO CANTY (McMaster University)

[Tuesday June 5 / mardi 5 juin, 13:15]

Statistical Analysis of RNA-Seq Data: A Comparison of Three R Packages

Analyse statistique de données d'ARN-Seq : comparaison de trois bibliothèques R

Next generation sequencing of RNA (RNA-Seq) is a new tool which can be used for analysis of differential expression of genes. A number of R/Bioconductor packages have been developed in recent years to deal with such data. In this project we examine the performance of three widely used packages (DESeq, edgeR and DEGseq) on two

Le séquençage d'ARN de nouvelle génération (ARN-Seq) est un nouvel outil permettant d'analyser l'expression différentielle de gènes. Plusieurs bibliothèques R/BioConductor ont été élaborées au cours des dernières années pour gérer ce type de données. Dans ce projet, nous examinons la performance de trois bibliothèques largement utilisées (DESeq, edgeR et DEGseq) sur deux jeux de données avec répétitions

2C-P: Poster Session II Séance d'affichage II

datasets with biological replication and show that a Poisson model as used in DEGseq is not appropriate in the presence of biological replicates and that there is a lack of consistency between the other two methods even though their overall performance is similar.

biologiques, nous montrons que l'utilisation d'un modèle de Poisson dans DEGseq est inappropriée à la présence de réplicats biologiques et qu'il y a un manque de cohérence entre les deux autres méthodes, même si leur performance générale est similaire.

XIAOQING NIU (University of Alberta), **PENGFEI LI** (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 13:15]

Testing Homogeneity in Normal Scale Mixture Models

Test d'homogénéité pour des modèles gaussiens à dispersion aléatoire

Normal scale mixture models are a special type of the mixture models and have many applications. We propose the retooled EM-test for testing homogeneity in them. Two approaches are carried out. In one way, the estimation of the common mean is initialized by the sample mean. In the other way, to initialize, we maximize the penalized log-likelihood function. We show that the EM-test statistics asymptotically follow the simple distribution $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. Simulation studies show that the EM-test has accurate type I errors, and the second approach enjoys higher power than the first. Two real data examples are analyzed.

Les modèles gaussiens à dispersion aléatoire sont un type particulier de modèle de mélange ayant plusieurs applications. Nous proposons d'évaluer leur homogénéité à l'aide d'un test EM révisé. Deux approches sont mises de l'avant pour l'initialisation des moyennes communes : utiliser les moyennes échantillonnales ou maximiser la log-vraisemblance pénalisée. Nous montrons que la statistique du test EM suit la distribution simple $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. Des études de simulation montrent l'exactitude du seuil du test EM et la puissance supérieure de la seconde approche face à la première. Les données de deux exemples concrets sont analysées.

MATTHEW RUEFFER, JULIE HORROCKS & DANIEL GILLIS (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 13:15]

Spatial and Temporal Modelling of Change in Marine Harvests

Modélisation spatiale et temporelle de la variation des pêches

From 1979 until 2010, catch data on whitefish (*Coregonus clupeaformis*) has been collected for the Canadian portion of Lake Huron. This data, obtained from native fishers and the Ontario Ministry of Natural Resources, contains the geographic location, harvest, relative effort and mesh size. This poster will examine catch per unit effort (CPUE = the quotient of harvest and effort) across regions and years, using exploratory techniques, temporal and spatial models.

Entre 1979 et 2010, des données relatives aux captures du grand corégone (*Coregonus clupeaformis*) ont été recueillies pour la portion canadienne du lac Huron. Ces données, qui proviennent de pêcheurs autochtones et du ministère des ressources naturelles de l'Ontario, comprennent l'emplacement géographique, les captures, l'effort relatif et le maillage. Cette affiche examine les captures par unité d'effort (CPUE = quotient des captures sur l'effort) à travers les régions et les années au moyen de méthodes exploratoires et de modèles temporels et spatiaux.

KATHLEEN RYAN, DANIEL GILLIS & STEPHEN CRAWFORD (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 13:15]

*Depth Distribution of Larval Lake Whitefish (*Coregonus clupeaformis*) in a Great Lakes Embayment*

*Distribution en profondeur des larves de grand corégone (*Coregonus clupeaformis*) dans une baie des Grands Lacs.*

2C-P: Poster Session II Séance d'affichage II

This study investigated the relationship between larval lake whitefish (*Coregonus clupeaformis*) density and water depth in Stokes Bay, Lake Huron. Nearshore waters (1-3 m) of the Great Lakes have been hypothesized to provide favourable habitat for larval whitefish, leading investigators to focus survey work there. The distribution of larvae in Stokes Bay was investigated by surface tows of 500 micron nets between April and May 2011. A novel stratified random sampling design assigned sample sites over depths ranging from 1-11m (n=71). Statistical analyses suggest depth was not significant in predicting larval fish density.

Cette étude examine la relation entre la densité de larves du grand corégone (*Coregonus clupeaformis*) et la profondeur de l'eau à Stokes Bay dans le lac Huron. L'hypothèse a été émise selon laquelle les eaux littorales (1-3 m) des Grands Lacs offriraient un habitat favorable aux larves du grand corégone, ce qui a mené les chercheurs à concentrer leur travail à cet endroit. La distribution des larves à Stokes Bay a été examinée au moyen de filets de 500 microns remorqués en surface entre avril et mai 2011. Un nouveau plan d'échantillonnage aléatoire stratifié a assigné des sites d'échantillonnage à des profondeurs variant de 1 à 11 mètres (n=71). Les analyses statistiques suggèrent que la profondeur n'est pas significative pour prédire la densité de larves.

OLGA SHESTAKOVSKA, JANICE POGUE, JOHN EIKELBOOM & STUART J. CONNOLLY (Population Health Research Institute, McMaster University and Hamilton Health Sciences)

[Tuesday June 5 / mardi 5 juin, 13:15]

Large Hazard Ratios Do Not Always Mean High Predictive Accuracy: A Simulation Study

De grands ratios de risque ne signifient pas toujours une grande exactitude prédictive : étude de simulation

In survival analysis, hazard ratio is a measure of relative risk frequently used to assess the effect of demographics, clinical characteristics, and biomarkers on event rates. We explore the relationship between hazard ratio and measures of predictive accuracy such as Harrell's c-statistic and time-dependent ROC curves in a simulation study. Motivation came from the analysis of a risk score calculated to predict stroke in patients with atrial fibrillation. We observed a strong association between the score and event rates (HR=2.45; 95% CI=1.66-3.75) yet the Harrell's c-statistic was 0.567.

En analyse de survie, le ratio de risque est une mesure du risque relatif fréquemment utilisée pour évaluer les effets de la démographie, de caractéristiques cliniques et de biomarqueurs sur les taux d'événements. Nous explorons la relation entre le ratio de risque et les mesures de l'exactitude prédictive comme la statistique c de Harrell et les courbes ROC dépendantes du temps dans une étude de simulation. La motivation est venue de l'analyse d'un score de risque calculé pour prévoir les accidents vasculaires cérébraux chez les patients atteints de fibrillation auriculaire. Nous observons une forte association entre le score et les taux d'événements (HR=2,45 ; 95% CI=1,66-3,75) alors que la statistique c de Harrell donne 0,567.

MONICA SIRSKI (University of Manitoba)

[Tuesday June 5 / mardi 5 juin, 13:15]

On the Statistical Analysis of Functional Data Arising from Designed Experiments

À propos de l'analyse statistique de données fonctionnelles découlant de plans d'expérience

We investigate various methods for testing whether two groups of curves are statistically significantly different, with the motivation to apply the techniques to the analysis of data arising from designed experiments. We propose a set of tests based on pairwise differences between individual curves. Our objective is to compare the power and

Nous examinons plusieurs méthodes pour vérifier si deux groupes de courbes sont significativement différents d'un point de vue statistique, avec la motivation d'appliquer les méthodes d'analyse des données découlant des plans d'expérience. Nous proposons un ensemble de tests basés sur des différences par paire entre les courbes individuelles. Notre objectif est de comparer la puissance et la robustesse

2C-P: Poster Session II Séance d'affichage II

robustness of a variety of tests, including a collection of permutation tests. We illustrate the application of these tests in the context of a designed 2^4 factorial experiment with a case study using data provided by NASA.

d'une variété de tests, y compris des tests de permutation. Nous illustrons l'application de ces tests dans le contexte d'un plan d'expérience factoriel 2^4 avec une étude de cas utilisant des données fournies par la NASA.

GARY SNEDDON (Mount Saint Vincent University), **MATTHEW KERBY** (Memorial University of Newfoundland)
[Tuesday June 5 / mardi 5 juin, 13:15]

Panel Data Models and Electoral Success of Ex-Cabinet Ministers
Modèles pour données longitudinales et succès électoral d'anciens ministres

Panel data models are of interest in a variety of disciplines where data are collected over time on cross-sections of individuals. We examine these models and estimation methods in studying data on the success of former Canadian federal cabinet ministers in subsequent elections. This will include an examination of some statistical approaches more commonly used in the social science literature.

Les modèles pour données longitudinales sont d'intérêt dans une variété de disciplines où des données transversales sont observées à répétition sur les mêmes individus au fil du temps. Nous examinons ces modèles et leurs méthodes d'estimation en étudiant des données sur le succès électoral d'anciens ministres fédéraux aux élections subséquentes. Nous examinons aussi des approches statistiques plus courantes dans la littérature des sciences sociales.

MARY THIESS (Fisheries and Oceans Canada / Pêches et Océans Canada), **SKIP MCKINNELL** (PICES - North Pacific Marine Organization), **MARC TRUDEL** (Fisheries and Oceans Canada / Pêches et Océans Canada)
[Tuesday June 5 / mardi 5 juin, 13:15]

Time Series Analysis: Assessing the "pink salmon" effect
Analyse de séries chronologiques : évaluation de "l'effet saumon rose"

Fraser River sockeye salmon have received a lot of press recently due to highly variable adult returns in recent years amid a general overall decline. One of many hypotheses to account for some of this variability suggests the potential for competition between sockeye salmon and pink salmon. Isolating the relationship between their respective time series requires the assessment (and quantification) of auto-correlation and non-monotonic trend over time. This poster will outline the challenges of assessing the "pink salmon" effect and highlight some approaches to analyzing these data.

Le saumon sockeye du fleuve Fraser a récemment fait l'objet d'une grande couverture médiatique en raison d'une montaison d'adultes très variable au cours des dernières années dans un contexte de déclin généralisé. L'une des nombreuses hypothèses pouvant expliquer cette variabilité suggère la compétition potentielle entre le saumon sockeye et le saumon rose. L'isolation de la relation entre leur série chronologique respective nécessite l'évaluation (et la quantification) de l'autocorrélation et de la tendance non monotone dans le temps. Cette affiche expose les défis soulevés par l'évaluation de "l'effet saumon rose" et propose certaines approches pour l'analyse de ces données.

MONICA H.T. WONG, PAUL MCNICHOLAS & DAVID M. MUTCH (University of Guelph)
[Tuesday June 5 / mardi 5 juin, 13:15]

Cluster Analyses to Assess Weight Loss Maintenance: An Application of Clustering in Nutrigenomics
Analyse de regroupement pour évaluer le maintien d'une perte de poids : application des analyses de regroupements à la nutriginomique

2C-P: Poster Session II Séance d'affichage II

Within nutrigenomics, data generated by microarray gene expression profiles can be clustered to help identify sub-populations of subjects that respond differently to a dietary intervention. The use of clustering analyses in obesity-related research is burgeoning as personalized nutrition gains popularity. This study focuses on clustering a human adipose tissue gene expression dataset obtained during a low-calorie diet intervention to identify differential responders. The aims of the study were to identify the best performing clustering method for samples, pick out differential responders to the low-calorie diet, and help find the biological pathways affected during the low-calorie diet by the differential responders.

En nutrigénomique, une analyse de regroupements des profils d'expression génique provenant de biopuces peut aider à identifier des sous-populations dont la réponse aux interventions diététiques est différente. Alors que les plans de nutrition personnalisés gagnent en popularité, l'utilisation d'analyses de regroupements foisonne dans la recherche sur l'obésité. Cette étude vise à identifier les individus dont la réponse diffère à partir d'une analyse de regroupements de profils d'expression génique obtenus durant une diète faible en calories, ainsi qu'à identifier les voies biologiques affectées chez ces individus durant la diète.

HAIYAN YANG & J. CONCEPCIÓN LOREDO-OSTI (Memorial University of Newfoundland)

[Tuesday June 5 / mardi 5 juin, 13:15]

Joint Modelling of Genetic Linkage and Association

Modélisation conjointe de la liaison et de l'association génétique

A joint modelling of genetic linkage and association within families having a remote common ancestor is presented. This joint modelling uses a likelihood approach that allows the inclusion of other covariates into the model for quantitative traits; for binary traits, a score test is proposed. The simulation study show that the joint test is more powerful than linkage or association test alone where both sources of variation are present. Furthermore, the proposed method also allows to test against specific alternatives. For example, against the significance of linkage given that there is no association and so on.

Un modèle conjoint est présenté pour la liaison et l'association génétique à l'intérieur des familles ayant un ancêtre commun éloigné. Ce modèle conjoint est basé sur la vraisemblance, ce qui permet l'inclusion de covariables quantitatives. Pour les traits binaires, un test de score est proposé. L'étude de simulation montre que le test conjoint est plus puissant qu'un test sur la liaison ou l'association seulement où les deux sources de variations sont présentes. De plus, la méthode proposée permet des contre-hypothèses spécifiques, par exemple la significativité de la liaison sachant qu'il n'y a pas d'association.

LI HUA YUE & WENQING HE (University of Western Ontario)

[Tuesday June 5 / mardi 5 juin, 13:15]

Logistic Prediction on Prostate Cancer Status Using Imaging Data

Prévisions logistiques de l'état du cancer de la prostate à l'aide de données d'imagerie

New imaging techniques, such as CT and MRI, have become commonly available for diagnosis and management of prostate cancer. Accurately identifying prostate cancer is essential to guide the treatment therapy. With the pathological cancer status available in our study, we can build prediction models using imaging characteristics as pre-

De nouvelles techniques d'imagerie médicale, telles que la tomodensitométrie et la résonance magnétique, sont devenues couramment disponibles pour le diagnostic et la gestion du cancer de la prostate. Il est essentiel d'identifier précisément la nature du cancer pour l'élaboration de traitements. Les données sur l'état pathologique du cancer de la prostate disponibles pour cette étude nous permettent de

2C-P: Poster Session II

Séance d'affichage II

dictors to predict prostate cancer status, and the model is expected to be used for future diagnosis of cancer using imaging tools. We show some preliminary results on the prediction of prostate cancer status using CT, T2WMR, and DCEMR features. The prediction accuracy is evaluated using AUC with a 10-fold cross-validation.

construire un modèle prédictif basé sur les caractéristiques des données d'imagerie. Nous pouvons ainsi prédire le statut du cancer de la prostate, et nous nous attendons à ce que ce modèle soit utilisé dans le futur pour le diagnostic du cancer de la prostate basé sur l'imagerie médicale. Nous présentons des résultats préliminaires de prévision du statut du cancer basée sur des propriétés d'images obtenues par tomographie, T2WMR et DCEMR. L'exactitude des prévisions est évaluée par l'aire sous la courbe ROC, déterminée par une validation croisée séparant les données en dix morceaux.

WEI YUHONG & PAUL MCNICHOLAS (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 13:15]

Mixture-model Averaging

Moyennage de modèles de mélange

Criteria commonly used for choosing the number of components in a finite mixture model include the Akaike information criterion, Bayesian information criterion, approximate weight of evidence, and the Integrated complete likelihood. We argue that it is not reasonable when the difference between the biggest value of such a criterion and some other values is small. It is, therefore, worth considering a model-averaging approach. We consider an averaging of the top M mixture models with applications in clustering and classification.

Les critères couramment utilisés pour choisir le nombre de composantes d'un modèle de mélange fini comprennent l'AIC, le BIC, la pondération approximative des informations probantes et l'intégration de la vraisemblance complète. Nous avançons que ces approches ne sont pas raisonnables lorsque la différence entre la valeur optimale du critère et les autres valeurs est petite. Par conséquent, il est préférable d'envisager une approche de moyennage des modèles. Nous considérons le moyennage des M meilleurs modèles de mélange et présentons des applications pour la classification et les analyses de regroupements.

2D-I1: Nonparametric Bayes Methods Méthodes bayésiennes non paramétriques

Organizer and Chair / Responsable et président: Saman Muthukumarana (University of Manitoba)

Rozanski Hall, room/local 103

Abstracts/Résumés

ALEXANDRE BOUCHARD-CÔTÉ, ARDAVAN SAEEDI & JOHN PETKAU (University of British Columbia), **ARNAUD DOUCET** (University of Oxford, U.K.)

[Tuesday June 5 / mardi 5 juin, 15:15]

The Gamma-Exponential Process
Le processus gamma exponentiel

We introduce the Gamma-Exponential Process (GEP), a prior over a large family of continuous time stochastic processes. A hierarchical version of this prior yields a useful model for analyzing complex time series. Models based on GEPs display many attractive properties: conjugacy, exchangeability and closed-form predictive distribution for the waiting times, and exact Gibbs updates for the time scale parameters. After establishing these properties, we show how posterior inference can be carried efficiently using Particle MCMC methods (Andrieu et al., 2010).

Nous présenterons le Processus Gamma Exponentiel (PGE), une loi a priori applicable à une grande famille de processus stochastiques à temps continu. Une version hiérarchique du processus sert de fondation à un modèle utile pour la modélisation de séries temporelles complexes. Les modèles basés sur les PGE possèdent plusieurs propriétés intéressantes : conjugaison, échangeabilité, une forme analytique pour la distribution prédictive des temps d'attente et un noyau de Gibbs pour l'échantillonnage des paramètres d'échelle du temps. De plus, l'inférence postérieure peut être effectuée efficacement grâce à un algorithme PMCMC (Andrieu et coll., 2010).

SIMON GUILLOTTE (Université du Québec à Montréal), **FRANÇOIS PERRON** (Université de Montréal)

[Tuesday June 5 / mardi 5 juin, 15:45]

A Sieves Model for Copulas
Un modèle de tamis pour copules

We consider the problem of estimating a copula function, and adopt a Bayesian approach. On the space of copula functions, we construct a finite-dimensional approximation subspace which is parametrized by a doubly stochastic matrix. A major problem here is the selection of a prior distribution on the space of doubly stochastic matrices also known as the Birkhoff polytope. In an objective Bayes perspective, we will discuss the Jeffreys' prior on the polytope. The Bayes estimator resulting from the Jeffreys' prior is evaluated numerically via MCMC methodology. The construction of the Markov chain poses some difficulties which will be highlighted.

Nous considérons le problème d'estimation d'une copule et adoptons une approche bayésienne. Sur l'espace des copules, nous construisons un sous-espace d'approximation à dimension finie qui est paramétré par une matrice doublement stochastique. Dans cet article, le principal problème est la sélection d'une distribution a priori sur les matrices doublement stochastiques, également connue sous le nom de polytope de Birkhoff. Dans une perspective bayésienne objective, nous discuterons des lois a priori de Jeffreys sur le polytope. L'estimateur bayésien qui découle de la loi a priori de Jeffreys est évalué numériquement par la méthodologie Monte Carlo par chaîne de Markov. La construction de la chaîne de Markov pose des difficultés que nous soulignerons.

2D-I1: Nonparametric Bayes Methods Méthodes bayésiennes non paramétriques

TIM SWARTZ (Simon Fraser University), **PARAMJIT GILL** (University of British Columbia, Okanagan)

[Tuesday June 5 / mardi 5 juin, 16:15]

Stylometric Analyses using Dirichlet Process Mixture Models

Analyses stylométriques par l'utilisation de modèles de mélanges de processus de Dirichlet

Stylometry refers to the statistical analysis of literary style of authors based on the characteristics of expression in their writings. We propose an approach to stylometry based on a Bayesian Dirichlet process mixture model using multinomial word frequency data. The parameters of the multinomial distribution of word frequency data are the “word prints” of the author. Our approach is based on model-based clustering of the vectors of probability values of the multinomial distribution. The resultant clusters identify different writing styles that assist in author attribution for disputed works in a corpus.

La stylométrie réfère à l'analyse statistique du style littéraire d'auteurs en fonction des caractéristiques de leur expression écrite. Nous proposons une approche de stylométrie basée sur un modèle bayésien de mélanges de processus de Dirichlet pour des données multinomiales sur la fréquence d'utilisation des mots. Les paramètres de la distribution multinomiale sont l'« empreinte littéraire » de l'auteur. Notre approche consiste en une analyse de regroupement basée sur un modèle pour les vecteurs de probabilité de la distribution multinominale. Les grappes qui en découlent identifient les différents styles de rédaction, ce qui contribue à déterminer l'attribution à un auteur pour les travaux contestés d'un corpus.

2D-I2: Recent Advances in Analysis of Survival Data from Cross Sectional Sampling Progrès récents en analyse de survie à partir de données transversales

Organizer and Chair / Responsable et président: Masoud Asgharian (McGill University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Rozanski Hall, room/local 101

Abstracts/Résumés

PIERRE-JÉRÔME BERGERON (University of Ottawa), **MASOUD ASGHARIAN** (McGill University), **EWA MAKVANDI-NEJAD** (University of Ottawa)

[Tuesday June 5 / mardi 5 juin, 15:15]

Nonparametric Estimation of the Survival Function for Prevalent and Incident Cases under Stationary Incidence

Estimation nonparamétrique de la fonction de survie pour des cas prévalents et incidents lorsque l'incidence est stationnaire

In epidemiological studies, subjects with disease (prevalent cases) differ from newly diseased (incident) cases. They tend to survive longer due to sampling bias. The usual estimator of the survival function for incident cases with right-censoring is the Kaplan-Meier (KM) estimator. Under stationary incidence, prevalent cases are length-biased. The nonparametric maximum likelihood estimator of the survival function correcting length-bias is obtained through a completely different approach than the KM estimator. In large longitudinal studies, both incident and prevalent cases may occur. We present a product-limit method that combines both estimators and uses all the data, and obtain some of its asymptotic properties. We finally provide an illustration on CSHA data.

En épidémiologie, les sujets déjà malades (cas prévalents) diffèrent des nouveaux malades (cas incidents) : ils sont caractérisés par une durée de vie plus longue. Lorsque l'incidence de la maladie est stationnaire, il y a biais de longueur des cas prévalents. L'estimateur de la fonction de survie avec censure à droite est donc différent pour les deux types de sujets. Dans de grandes études longitudinales, les deux types de sujets peuvent être observés. On présente une méthode nonparamétrique combinant les deux estimateurs et qui utilise toutes les données, avec ses propriétés asymptotiques et une illustration sur les données de l'ÉSCV.

MARCO CARONE (University of California, Berkeley), **DANIEL SCHARFSTEIN** (Johns Hopkins Bloomberg School of Public Health), **MARK VAN DER LAAN** (University of California, Berkeley), **MASOUD ASGHARIAN** (McGill University)

[Tuesday June 5 / mardi 5 juin, 15:35]

Quantifying the Impact of Disease on Lifetime Under Cross-sectional Sampling

Quantifier l'impact d'une maladie sur la durée de vie à l'aide d'un échantillon transversal

Common regression models are inadequate to characterize the relationship between age at disease onset and time from onset to death because of the intricate dependence structure between the two. We propose a novel semiparametric model to study the impact of disease on lifetime, and discuss

Puisque la relation de dépendance entre l'âge de début d'une maladie et la durée de vie résiduelle est complexe, les modèles de régression conventionnels sont peu utiles à son étude. Nous proposons un nouveau modèle semi-paramétrique afin d'étudier l'impact d'une maladie sur la durée de vie. Nous décrivons ensuite l'estimation de ses para-

2D-I2: Recent Advances in Analysis of Survival Data from Cross Sectional Sampling Progrès récents en analyse de survie à partir de données transversales

estimation and inference using data from a cross-sectional survey with follow-up. Such data are particularly challenging to analyze because they are subject to systematic biases and informative censoring. We conclude with a brief discussion of the impact of dementia on the lifetime of elderly Canadians using data from the Canadian Study of Health and Aging.

mètres à l'aide de données de survie échantillonnées transversalement ; cette étape est délicate car ces données comportent un biais structurel et une censure instructive. Finalement, nous discutons brièvement de l'impact de la démence sur la durée de vie des Canadiens âgés à l'aide de données provenant de l'Étude sur la santé et le vieillissement au Canada.

ASHKAN ERTEFAEI (University of Michigan), **MASOUD ASGHARIAN & DAVID STEPHENS** (McGill University)

[Tuesday June 5 / mardi 5 juin, 15:55]

Double Bias

Double biais

We consider estimating the grouping (exposure) effect on survival time from observational data when, in addition to the lack of randomization, the data constitute a length-biased sample; we hence term this a double-bias problem. We introduce two approaches based on weighted and double robust estimating equations for estimating grouping effect. We apply the proposed methods to a set of length-biased survival data collected as part of the Canadian Study of Health and Aging (CSHA) to compare survival of subjects with dementia among institutionalized patients versus those recruited from the community and depict their adjusted survival curves.

Nous estimons l'effet de regroupement (ou d'exposition) sur le temps de survie de données d'observation lorsque, en plus de l'absence de randomisation, les données constituent un échantillon biaisé en longueur. Nous appelons ce phénomène un problème de double biais. Nous proposons deux approches basées sur des équations d'estimations pondérées et doublement robustes pour estimer l'effet de regroupement. Nous appliquons les méthodes proposées à un ensemble de données de survie biaisées en longueur dans le cadre de l'Étude sur la santé et le vieillissement au Canada (ESVC) en vue de comparer la survie de sujets atteints de démence parmi des patients placés en établissement par rapport à d'autres recrutés dans la communauté, et de représenter leurs courbes de survie ajustées.

TAKI R. SHINOHARA & MEI-CHENG WANG (Johns Hopkins Bloomberg School of Public Health)

[Tuesday June 5 / mardi 5 juin, 16:15]

Alternating Event Processes during Lifetimes: Population Dynamics and Statistical Inference

Processus d'événements alternant au cours de la durée de vie : dynamique des populations et inférence statistique

There are many biomedical applications in which patients experience nontrivial durations associated with each event. This results in a process where the disease status of a patient alternates between exacerbations and remissions. In this work, we consider the dynamics of a chronic disease and its associated exacerbation-remission process over two time scales: calendar time and time-since-onset. In particular, over calendar time, we explore population dynamics and the relationship between incidence, prevalence and duration for such alternating event processes. We provide nonparametric estimation techniques for characteristic quantities

Il existe plusieurs applications biomédicales dans le cadre desquelles les patients traversent des événements à durée non triviale. Il découle de ce fait un processus où le statut de la maladie du patient alterne entre l'exacerbation et la rémission. Dans le cadre de ce travail, nous considérons la dynamique d'une maladie chronique et de ses processus associés d'exacerbation-rémission sur deux échelles de temps : le calendrier civil et le temps écoulé depuis l'apparition. Nous explorons notamment la dynamique des populations et la relation entre l'incidence, la prévalence et la durée de tels processus d'événements alternatifs sur un calendrier civil. Nous proposons des techniques d'estimation non-paramétriques des quantités caractéristiques du pro-

2D-I2: Recent Advances in Analysis of Survival Data from Cross Sectional Sampling **Progrès récents en analyse de survie à partir de données transversales**

of the process, and we develop techniques for estimating semiparametric models of prevalence.

cessus et nous développons des techniques pour estimer les modèles semi-paramétriques de prévalence.

2D-I3: Teaching and Learning Statistics with Technology Enseignement et apprentissage de la statistique à l'aide des nouvelles technologies

Chair/Président: Saumen Mandal (University of Manitoba)

Organizers/Responsables: Saumen Mandal and/et Zeneida Mateo (University of Manitoba)

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

Alexander Hall, room/local 200

Abstracts/Résumés

BRUCE DUNHAM & EUGENIA YU (University of British Columbia), **JOANNE QUAN** (University of Toronto)

[Tuesday June 5 / mardi 5 juin, 15:15]

Using On-Line Simulation Tools to Promote Student Understanding of Sampling Distributions.

L'utilisation d'outils de simulation en ligne pour favoriser la compréhension par les étudiants des distributions d'échantillonnage

It is common for learners on an introductory statistics course to have difficulty grasping the concept of sampling distributions. On-line simulation-based applets can help students discover and explore aspects of sampling distribution theory. Several on-line tools are reviewed, and suggestions are made as to how best to use one such tool (namely the sampling distribution simulator in the Rice Virtual Lab in Statistics) in an introductory course.

Il est courant que les étudiants d'un cours d'introduction à la statistique aient de la difficulté à comprendre le concept de distribution d'échantillonnage. Les applications de simulation en ligne peuvent aider les étudiants à découvrir et à explorer les aspects de la théorie de la distribution d'échantillonnage. Nous examinons plusieurs outils en ligne et formulons des suggestions sur la meilleure manière d'utiliser l'un de ces outils (à savoir le simulateur de distribution d'échantillonnage du Rice Virtual Lab in Statistics) dans un cours d'introduction.

BETHANY WHITE (University of Western Ontario)

[Tuesday June 5 / mardi 5 juin, 15:45]

Harnessing the Power of Technology to Get Students Talking about Statistics

Exploiter la puissance de la technologie pour inciter les étudiants à parler de statistique

Technology is omnipresent in today's world. Our students have grown up with the Internet and social media is second nature to them. How can we take advantage of this in our classes to promote learning? One possibility is to include online discussions in our courses. Online discussions can work for any class format ranging from face-to-face to online classes and can take place in learning management systems or even social media websites like Facebook. In this talk, I will present my experiences with structured online discussions, describe discussion activities used in an introductory statis-

La technologie est omniprésente dans le monde d'aujourd'hui. Nos étudiants ont grandi avec l'Internet et les médias sociaux sont une seconde nature pour eux. Comment pouvons-nous en tirer profit dans nos classes pour promouvoir l'apprentissage? Une possibilité consiste à inclure des discussions en ligne dans nos cours. Les discussions en ligne peuvent fonctionner pour tout format de cours, allant des classes en personne aux cours en ligne et peuvent s'intégrer dans les systèmes de gestion de l'apprentissage ou même dans les sites Web de médias sociaux comme Facebook. Dans cet exposé, je présente mes expériences en matière de discussions structurées en ligne, je décris des

2D-I3: Teaching and Learning Statistics with Technology

Enseignement et apprentissage de la statistique à l'aide des nouvelles technologies

tics course and share some student feedback.

activités de discussion utilisées dans un cours d'introduction à la statistique et je partage quelques commentaires d'étudiants.

ROGER WOODARD (North Carolina State University)

[Tuesday June 5 / mardi 5 juin, 16:15]

A Large Course Redesign of Introductory Statistics: You Don't Need 65 Other People with You to Learn a Definition
Un grand remaniement d'un cours d'introduction à la statistique : vous n'avez pas besoin de 65 personnes à vos côtés pour apprendre une définition

We present the results of a redesign project in a large introductory Statistics course. This project converted two of the three hours of student contact to an online format. Students still come together one day a week to participate in discussions where an instructor teaches them the important ideas and concepts. During this time students also participate in group activities that illustrate key ideas. The online component of the course incorporates narrated presentations of slides and animations along with interactive content. This course capitalizes on technology to provide a more personalized and pedagogically advantageous course.

Nous présentons les résultats d'un projet de remaniement d'un grand cours d'introduction à la statistique. Dans le cadre de ce projet, nous avons converti deux des trois heures de contact avec les étudiants vers un format en ligne. Les étudiants se réunissent quand même une fois par semaine pour participer à des discussions où un enseignant leur montre les idées et concepts importants. Les étudiants participent alors à des activités de groupe qui illustrent les idées principales. La composante en ligne du cours intègre des présentations narrées des acétates et des animations en plus de contenu interactif. Ce cours tire profit de la technologie pour fournir une expérience plus personnalisée et avantageuse du point de vue pédagogique.

2D-14: The Census of Population in Canada: Looking Back at 2011 and Into the Future Le Recensement de la population au Canada : Retour sur 2011 et regard vers l'avenir

Chair/Président: Don McLeish (University of Waterloo)

Organizer/Responsable: Julie Trépanier (Statistics Canada / Statistique Canada)

Rozanski Hall, room/local 104

Abstracts/Résumés

PATRICE MATHIEU (Statistics Canada / Statistique Canada)

[Tuesday June 5 / mardi 5 juin, 15:15]

Data Collection for the 2011 Census and National Household Survey

La collecte du Recensement de la population et de l'Enquête nationale auprès des ménages de 2011

The most recent Census of population in Canada took place in 2011. Important changes have been made, in part due to the Canadian government's decision to restrict the Census to the short form and to replace the long form by a voluntary survey, the National Household Survey (NHS). Furthermore, a new wave collection methodology was introduced in order to stimulate self-response, in particular Internet response. This presentation will give an overview of the collection methodology, focussing on the aspects described above.

Le plus récent Recensement de la population au Canada s'est déroulé en 2011. Des changements importants ont été apportés, notamment suite à la décision gouvernementale de restreindre le Recensement au formulaire court et de remplacer le formulaire long par une enquête volontaire, l'Enquête nationale auprès des ménages (ENM). De plus, une nouvelle méthode de collecte par vague a été introduite afin de stimuler des réponses, avec en priorité des réponses par Internet. Cette présentation décrira les grandes lignes de la méthodologie de collecte, puis abordera plus spécifiquement quelques aspects reliés aux changements ci-haut mentionnés.

NORMAND LANIEL (Statistics Canada / Statistique Canada)

[Tuesday June 5 / mardi 5 juin, 15:45]

The Estimation Methodology for the 2011 Census of Population and the National Household Survey

La méthodologie d'estimation du Recensement de la population et de l'Enquête nationale auprès des ménages de 2011

Following the decision to replace the long form sample of the 2011 Census of Population by a voluntary survey, the National Household Survey (NHS), the estimation methodology used in previous censuses has been redesigned. This redesign includes a more elaborated adjustment for total nonresponse which uses the information collected with the short form and other data sources. In addition, weighting via calibration to census counts will benefit from the addition of questions on languages on the short form. This presentation will outline the estimation methodology for the Census and the NHS. Some results will be presented for the Census.

Suite à la décision de remplacer le formulaire long du Recensement de la population de 2011 par une enquête volontaire, l'Enquête nationale auprès des ménages (ENM), la méthodologie d'estimation utilisée dans les recensements précédents a été remaniée. Ce remaniement inclut un ajustement pour la non-réponse totale plus élaboré qui, entre autres, utilise les informations recueillies à l'aide du formulaire court. La pondération par calage aux chiffres du recensement bénéficiera de l'addition de questions sur les langues au formulaire court. Cette présentation donnera un aperçu de la méthodologie d'estimation du Recensement et l'ENM. Quelques résultats pour le Recensement seront présentés.

2D-I4: The Census of Population in Canada: Looking Back at 2011 and Into the Future Le Recensement de la population au Canada : Retour sur 2011 et regard vers l'avenir

JULIE TRÉPANIÉ (Statistics Canada / Statistique Canada)

[Tuesday June 5 / mardi 5 juin, 16:15]

The 2016 Census Strategy Project

Le projet de la stratégie du Recensement de 2016

In December 2010, Statistics Canada launched the 2016 Census Strategy Project. Its objective was to study options and ultimately deliver a recommendation on the methodology of the 2016 Census to the federal government in early 2012. The project reviewed the approaches for population censuses that exist around the world and evaluated their applicability to the Canadian context. It comprised a review of the constitutional and statutory requirements and the provision of a content determination framework for the future. This presentation will provide a summary to the statistical community of the work performed as part this project.

Statistique Canada a lancé en décembre 2010 le Projet de la stratégie du Recensement de 2016 afin d'étudier des options quant à la méthodologie du Recensement de 2016 et d'émettre ultimement une recommandation au gouvernement fédéral au début 2012. Le projet a examiné les approches qui existent dans le monde pour tenir des recensements de la population et évalué leur applicabilité au contexte canadien. Il comprenait une revue des exigences constitutionnelles et législatives et l'élaboration d'un cadre de détermination du contenu pour le futur. Cette présentation, destinée à la communauté statistique, dressera un sommaire des activités menées dans ce projet.

2D-I5: Theory and Applications of Self-Normalized Processes Théorie et application des processus auto-normalisés

Chair/Président: Hao Yu (University of Western Ontario)
Organizer/Responsable: Yuliya Martsynyuk (University of Manitoba)

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

Alexander Hall, room/local 100

Abstracts/Résumés

YULIYA MARTSYNYUK (University of Manitoba)

[Tuesday June 5 / mardi 5 juin, 15:15]

Invariance Principles for a Multivariate Student Process in the Generalized Domain of Attraction of the Multivariate Normal Law

Principes d'invariance d'un processus multivarié de Student dans le domaine d'attraction généralisé d'une loi normale multivariée

For a d -variate Student process based on independent copies of a random vector X , with trajectories in the space of R^d -valued cadlag functions on $[0, 1]$, our main result establishes a uniform Euclidean norm approximation in probability with a sequence of appropriate d -variate processes with Wiener process components, assuming that X is in the generalized domain of attraction of the multivariate normal law (GDAN) and some additional conditions. As a consequence, a functional central limit theorem is also concluded for the Student process. The condition $X \in \text{GDAN}$ for these invariance principles is shown to be not only sufficient, but also necessary.

Pour un processus de Student d -varié basé sur des copies indépendantes d'un vecteur aléatoire X , avec des trajectoires dans l'espace des fonctions càdlàg sur $[0, 1]$ prenant des valeurs dans R^d , notre résultat principal établit une approximation uniforme de la norme euclidienne à l'aide d'une suite de processus d -varié avec des composantes d'un processus de Wiener, supposant que X se trouve dans le domaine généralisé d'attraction d'une loi normale multivariée et quelques autres conditions. Par conséquent, nous pouvons en tirer un théorème limite central fonctionnel pour le processus de Student. Nous montrons que la condition $X \in \text{GDAN}$ pour ces principes d'invariance n'est pas seulement suffisante, mais aussi nécessaire.

MASOUD NASARI & MIKLÓS CSÖRGŐ (Carleton University)

[Tuesday June 5 / mardi 5 juin, 15:45]

Another Look at Bootstrapping the Student t -Statistic

Un autre regard sur le bootstrap de la statistique t de Student

Let $\{X, X_i, i \geq 1\}$ be i.i.d. non-degenerate random variables. For random samples $\{X_i, 1 \leq i \leq n\}_{n \geq 1}$ let $\{v_i^{(n)}, 1 \leq i \leq n\}_{n \geq 1}$ be the bootstrap weights, consider a triangular array of random variables that are to be determined by the scheme of bootstrap in hand so that $\sum_{i=1}^n v_i^{(n)} = m_n$, the bootstrap size,

Soient $\{X, X_i, i \geq 1\}$ des variables aléatoires i.i.d. non dégénérées. Pour des échantillons aléatoires $\{X_i, 1 \leq i \leq n\}_{n \geq 1}$, soient $\{v_i^{(n)}, 1 \leq i \leq n\}_{n \geq 1}$ les poids bootstrap, considérons une matrice triangulaire de variables aléatoires déterminées par le schéma du bootstrap de telle sorte que $\sum_{i=1}^n v_i^{(n)} = m_n$, la taille du bootstrap et $E(v_i^{(n)}/m_n) = 1/n$. Posons $\bar{X}_{m_n}^* =$

2D-I5: Theory and Applications of Self-Normalized Processes Théorie et application des processus auto-normalisés

and $E(v_i^{(n)}/m_n) = 1/n$. Put $\bar{X}_{m_n}^* = \sum_{i=1}^n v_i^{(n)} X_i/m_n$, $\bar{X}_n = \sum_{i=1}^n X_i/n$. t -statistic versions of the bootstrap process $\{\bar{X}_{m_n}^* - \bar{X}_n\}_{n \geq 1}$ will be studied via conditioning on the bootstrap weights and also by the traditional method of conditioning on the samples.

$\sum_{i=1}^n v_i^{(n)} X_i/m_n$, $\bar{X}_n = \sum_{i=1}^n X_i/n$. Nous étudions les versions de processus bootstrap $\{\bar{X}_{m_n}^* - \bar{X}_n\}_{n \geq 1}$ basés sur la statistique t en conditionnant sur les poids bootstrap et également par la méthode traditionnelle de conditionnement sur les échantillons.

MIKLÓS CSÖRGŐ (Carleton University)

[Tuesday June 5 / mardi 5 juin, 16:15]

Weak Invariance Principles and Strong Limit Laws for Self-Normalized Processes

Principes d'invariance faible et lois limites fortes pour des processus autonormalisés

Let $\{X, X_n, n \geq 1\}$ be i.i.d. non-degenerate random variables. Put $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$ and $V_n^2 = \sum_{i=1}^n X_i^2, n \geq 1$. We will attempt a glimpse at Donsker-type weak invariance principles that have been recently established for the self-normalized partial sums processes $\{S_{[nt]}/V_n, 0 \leq t \leq 1\}_{n \geq 1}$ when X belongs to the domain of attraction of a stable law with index $\alpha \in (0, 1]$. Special attention will be given to the case of $\alpha = 2$, i.e., when $X \in \text{DAN}$, and to applications to change-point analysis in this domain with X possibly having an infinite variance. A Strassen-type strong invariance principle will also be mentioned.

Soient $\{X, X_n, n \geq 1\}$ des variables aléatoires i.i.d. non dégénérées. Posons $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$ et $V_n^2 = \sum_{i=1}^n X_i^2, n \geq 1$. Nous tentons d'examiner des principes d'invariance faible de type Donsker qui ont été établis récemment pour les processus autonormalisés de sommes partielles $\{S_{[nt]}/V_n, 0 \leq t \leq 1\}_{n \geq 1}$ lorsque X appartient au domaine d'attraction d'une loi stable avec indice $\alpha \in (0, 1]$. Une attention particulière sera portée au cas où $\alpha = 2$, c'est-à-dire lorsque $X \in \text{DAN}$, et à des applications d'analyse du point de rupture pour ce domaine lorsque X a possiblement une variance infinie. Nous abordons également un principe d'invariance forte de type Strassen.

2D-C1: Business and Industrial Statistics and Quality Control Statistique industrielle et de gestion et contrôle de la qualité

Chair/Président: Roman Viveros-Aguilera (McMaster University)

Rozanski Hall, room/local 105

Abstracts/Résumés

DAVTALAB OLYAIE MOSTAFA & GHOLAMREZA JAHANSHAHLOO (Tarbiat Moallem University, Iran), **MASOUD ASGHARIAN** (McGill University), **VAHID PARTOVI NIA** (École Polytechnique de Montréal)

[Tuesday June 5 / mardi 5 juin, 15:15]

Efficiency in Stochastic Data Envelopment Analysis

Efficacité dans l'analyse d'enveloppement de données stochastiques

Data envelopment analysis is a linear programming method to evaluate the relative efficiency of a set of decision making units, with a lot of applications in many domains, e.g. management, industrial engineering, and economics. Decision making units produce a multi-output values using several inputs. In some cases decision making units' data are measured with uncertainty. However, common data envelopment analysis handles accurate data, and therefore produce a single value for the relative efficiency. We study the relative efficiency when the available data is distributed randomly around the measured values.

L'analyse d'enveloppement des données est une méthode de programmation linéaire pour évaluer l'efficacité relative pour un ensemble d'unités de décision, avec plusieurs applications dans certains domaines comme la gestion, l'ingénierie industrielle et l'économie. Les unités de décision produisent des valeurs multi-sorties à l'aide de plusieurs entrées. Dans certains cas, les valeurs des unités sont mesurées avec une incertitude, mais l'analyse d'enveloppement des données traite les données précises. Nous étudions l'efficacité relative lorsque les données disponibles sont distribuées de façon aléatoire autour des valeurs mesurées.

GEMAI CHEN (University of Calgary)

[Tuesday June 5 / mardi 5 juin, 15:30]

Discovery Models — A Least Squares Approach With Applications

Modèles de découverte — méthode des moindres carrés et applications

Discovery models have been used in industry to assess natural resource reserves. In this talk, I will introduce a discovery model in two ways, one discrete and one continuous, and discuss nonparametric and parametric fitting of this model. Application of this discovery model is illustrated with two real data sets.

Les modèles de découverte sont utilisés en industrie pour évaluer les réserves de ressources naturelles. Je présente un modèle de découverte sous deux versions, l'une discrète et l'autre continue, et je discute de l'ajustement non paramétrique et paramétrique de ce modèle. J'en illustre enfin l'application à l'aide de deux ensembles de données réelles.

AMIR HOSEIN HOMAIE SHANDIZI & VAHID PARTOVI NIA (École Polytechnique de Montréal)

[Tuesday June 5 / mardi 5 juin, 15:45]

A Test for Attribute Gauge R&R

Un test pour le gage R&R

2D-C1: Business and Industrial Statistics and Quality Control Statistique industrielle et de gestion et contrôle de la qualité

Measurement systems studies are an integral part of most quality improvement processes. A desirable measurement system must produce repeatable and reproducible values; such measurement systems are called R&R. The attribute measurement systems are rarely studied. We propose a significance test of R&R for an attribute gauge, and demonstrate its desirable properties. The efficiency of the methodology is inspected in a simulation study, and applied to an example.

Les systèmes de mesure sont une partie intégrante de la plupart des processus d'amélioration de la qualité. Un système de mesure souhaitable doit produire des valeurs répétables et reproductibles ; un tel système de mesure est appelé R&R. Les systèmes de mesure par attributs sont rarement étudiés. Nous proposons un test de R&R pour un gage discret, et démontrons ses propriétés souhaitables. L'efficacité de la méthodologie est inspectée dans une étude de simulation, et appliqué dans un exemple.

JAYASANKAR VATATHOOR & ASOKAN MULAYATH VARIYATH (Memorial University of Newfoundland)

[Tuesday June 5 / mardi 5 juin, 16:00]

Robust Control Charts for Multivariate Individual Observations

Cartes de contrôle robustes pour les observations individuelles multivariées

Control charts are available to effectively monitor changes in the variances and correlations of a multivariate normal process when individual observations are collected. The outliers present in the data can have an adverse effect on the results on the control chart and hence it is important to identify and eliminate the outliers. We propose modified control charts using high breakdown robust estimation methods based on the re-weighted minimum covariance determinant or re-weighted minimum volume ellipsoid and study its performance using simulations by considering different data scenarios.

Les cartes de contrôle permettent de surveiller les changements dans les variances et corrélations d'un processus normal multivarié lors de la collecte d'observations individuelles. Les valeurs aberrantes parmi les données peuvent avoir un effet nuisible sur les résultats obtenus de la carte de contrôle, d'où l'importance de les identifier et de les éliminer. Nous proposons des cartes de contrôle modifiées utilisant des méthodes robustes d'estimation à point de rupture élevé basées sur un déterminant de covariance minimum repondéré ou sur l'ellipsoïde du volume minimum repondéré, et nous examinons leur performance au moyen de simulations selon différents scénarios de données.

ICK HUH, ROMAN VIVEROS-AGUILERA & NARAYANASWAMY BALAKRISHNAN (McMaster University)

[Tuesday June 5 / mardi 5 juin, 16:15]

The MEWMA with Two Unequal Smoothing Parameters

Le MEWMA avec deux paramètres de lissage inégaux

A new bivariate exponentially weighted moving average (MEWMA) control statistic with two smoothing parameters is modelled as a Markov chain to investigate the possibilities of improving the familiar MEWMA control chart. An extensive comparison reveals that the new Markov chain approximation approach enhances the efficiency and the flexibility of the control chart in certain directions. Using the unequal smoothing parameters, the ARL depends on the direction of the process mean shift, not on the value of the non-centrality parameter. It is shown that the control chart can show smaller

Une nouvelle statistique de contrôle à deux paramètres de lissage pour une moyenne mobile à poids exponentiels (MEWMA) bivariée est modélisée comme une chaîne de Markov pour investiguer les possibilités d'amélioration de la carte de contrôle habituelle du MEWMA. Une comparaison complète montre que la nouvelle approximation par chaîne de Markov améliore l'efficacité et la flexibilité de la carte de contrôle dans certaines directions. En utilisant des paramètres de lissage indépendants, l'ARL dépend de la direction du décalage moyen du processus, et non de la valeur du paramètre de décentralité. Il est démontré que la carte de contrôle donne de plus petits ARL optimaux lorsque le

2D-C1: Business and Industrial Statistics and Quality Control Statistique industrielle et de gestion et contrôle de la qualité

optimal ARLs when the mean shift takes place in preferred directions. décalage moyen est dans la direction sélectionnée.

2D-C2: Financial Applications Applications en finance

Chair/Président: Michael Evans (University of Toronto)

Rozanski Hall, room/local 102

Abstracts/Résumés

RENÉ FERLAND & FRANÇOIS WATIER (Université du Québec à Montréal)

[Tuesday June 5 / mardi 5 juin, 15:15]

Mean-Variance Efficiency in a Market with Regime-Switching Zero-Coupons
Efficacité moyenne-variance dans un marché à changement de régime

We study a mean-variance investment problem in a continuous-time framework where the interest rate follows a regime-switching Vasicek model. Using BSDEs (backward stochastic differential equations) we construct an optimal portfolio and show that the stochastic optimization problem is reduced to finding solutions to systems of ODEs (ordinary differential equations).

Nous étudions un problème d'optimisation de type moyenne-variance en temps continu où le taux d'intérêt suit un modèle de Vasicek à changement de régime. Nous utilisons des équations différentielles stochastiques reculant dans le temps pour construire un portefeuille optimal et nous montrons que le problème d'optimisation stochastique se réduit à la résolution de systèmes d'équations différentielles ordinaires.

FRANÇOIS WATIER (Université du Québec à Montréal), **CHANTAL LABBÉ** (HEC Montréal)

[Tuesday June 5 / mardi 5 juin, 15:30]

Goal Achieving Probabilities of Optimal Cone-Constrained Mean-variance Portfolios
Probabilités d'atteinte des objectifs de portefeuilles moyenne-variance optimaux à contrainte conique

Li and Zhou (2006) established that an investor, following an unconstrained mean-variance strategy, will achieve its discounted targeted wealth with a probability greater than 80%. We will show that in the case of optimal cone-constrained portfolios this lower bound probability still holds. Furthermore, we will give an expression for the probability that the investor reaches his target before a possible bankruptcy.

Li et Zhou (2006) ont établi qu'un investisseur atteint sa richesse actualisée visée avec une probabilité supérieure à 80% en suivant une stratégie moyenne-variance sans contrainte. Nous montrerons que cette probabilité de borne inférieure est également valide dans le cas de portefeuilles optimaux à contrainte conique. De plus, nous proposerons une expression de la probabilité que l'investisseur atteigne sa cible avant une faillite possible.

TAEHAN BAE (University of Regina), **IAN ISCOE** (Algorithmics Inc.)

[Tuesday June 5 / mardi 5 juin, 15:45]

Sum of Bernoulli Mixtures: Beyond Conditional Independence
Somme de mélanges de Bernoulli : au-delà de l'indépendance conditionnelle

Continuous Bernoulli mixtures are widely used to model the dependence between binary random variables such as the dependent defaults in a credit portfolio. We consider the distribution of the sum of exchangeable Bernoulli mixtures under a general

Les mélanges continus de loi Bernoulli permettent de modéliser la dépendance entre des variables aléatoires binaires, comme les défaillances dépendantes d'un portefeuille de crédit. Nous étudions la distribution de la somme des mélanges échangeables de Bernoulli sous une structure de dé-

2D-C2: Financial Applications

Applications en finance

dependence structure. The level of dependence is measured in terms of the limit, with respect to a threshold tending to 1, of the conditional correlation between two Bernoulli mixtures, given that the mixing random variable is larger than the threshold. The large sample distribution of the empirical frequency and its use in approximating the risk measures, Value-at-Risk and Conditional Tail Expectation, are presented.

pendance générale. Le niveau de dépendance se mesure comme la limite, par rapport à un seuil tendant vers 1, de la corrélation conditionnelle entre deux mélanges de Bernoulli, la variable aléatoire de mélange étant supérieure au seuil. Nous présentons la distribution de la fréquence empirique dans un grand échantillon et l'utilisons pour approximer la valeur à risque et l'espérance conditionnelle unilatérale.

ZHENYU CUI, CAROLE BERNARD & DON MCLEISH (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 16:00]

Nearly Exact Option Price Simulation using Characteristic Functions

Simulation presque exacte de prix d'options à partir de fonctions caractéristiques

This paper presents a new approach to perform a nearly unbiased simulation using inversion of the characteristic function. As an application, we are able to give unbiased estimates of the price of forward starting options in the Heston model and of continuously monitored Parisian options in the Black-Scholes framework. This method of simulation can be applied to problems for which the characteristic functions are known but the corresponding probability density functions are complicated.

Dans cet article, nous développons une nouvelle approche pour simuler par une méthode Monte Carlo directement à partir de la fonction caractéristique. Cette technique de simulation est sans biais pour des distributions à support fini, et sinon presque sans biais. Comme applications, nous estimons des prix d'options dans le modèle de volatilité stochastique de Heston, ainsi que des prix d'options parisiennes dans le cadre du modèle Black et Scholes. Cette méthode de simulation peut être appliquée à tous problèmes pour lesquels les fonctions caractéristiques sont parfaitement connues, mais où les densité correspondantes ont une forme compliquée.

BRIAN M. HARTMAN (University of Connecticut), **CHRIS GROENDYKE** (Robert Morris University)

[Tuesday June 5 / mardi 5 juin, 16:15]

Model Selection and Averaging in Regime-Switching Models of Various Types

Sélection et moyennage de modèles à changements de régimes de divers types

Simulated asset returns are used in many areas of actuarial science. For example, life insurers use them to price annuities, life insurance, and investment guarantees. The quality of those simulations has come under increased scrutiny during the current financial crisis. When simulating the asset process, properly choosing which model or models to use, and accounting for your uncertainty in that choice is essential. We investigate how to best choose a model from a flexible set of regime-switching models where the individual regimes are not constrained to be from the same distributional family.

Les retours sur actifs simulés sont utilisés dans de nombreux domaines de l'actuariat. Dans un cadre d'assurance-vie, par exemple, les assureurs les utilisent pour fixer les prix des rentes, de l'assurance-vie et des investissements garantis. La qualité de ces simulations est étudiée encore plus minutieusement depuis la crise financière actuelle. Lors de la simulation du processus d'actifs, choisir correctement les modèles à utiliser et tenir compte de votre incertitude sur ce choix est essentiel. Nous enquêtons sur la meilleure façon de choisir un modèle à partir d'un ensemble flexible de modèles à changements de régimes où les régimes individuels ne sont pas contraints de provenir de la même famille de distribution.

2D-C2: Financial Applications

Applications en finance

CHENGGUO WENG (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 16:30]

Constant Proportion Portfolio Insurance under Regime Switching Lévy Exponential Process

Assurance portefeuille à proportions constantes dans le cadre d'un processus exponentiel de Lévy avec changements de régime

Constant Proportion Portfolio Insurance (CPPI) is one of the most popular portfolio managing techniques, designed to participate in potential gains on an upward market moving while limiting the downside risk. One main concern on CPPI is the gap risk, defined as the risk for the portfolio value to fall below a given protection floor. In this talk, I will introduce my findings on the gap risk when the asset price follows a regime switching Lévy exponential process. Analytical forms of its shortfall probability, expected shortfall and characteristic function will be presented.

L'une des techniques de gestion de portefeuilles les plus populaires, à savoir la méthode de l'assurance portefeuille à proportions constantes (CPPI), est conçue pour participer aux gains potentiels du changement de régime d'un marché à la hausse tout en limitant les risques de baisse. La principale préoccupation de la CPPI est le risque d'écart, défini comme le risque que la valeur du portefeuille baisse sous un seuil de protection donné. Dans cet exposé, je présente mes résultats sur le risque d'écart lorsque le prix des actifs suit un processus exponentiel de Lévy avec changements de régime. La probabilité de déficit, le déficit prévu et la fonction caractéristique seront présentés sous forme analytique.

2D-C3: Mixture Models, Clustering and Classification Modèles de mélanges, partitionnement et classification

Chair/Président: Abbas Khalili (McGill University)

MacNaughton Building, room/local 105

Abstracts/Résumés

LIAM CALLAGHAN & AYESHA ALI (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 15:15]

Modelling Plant-Pollinator Interactions with Mixtures of Linkage Rules

Modélisation des interactions plante-pollinisateur avec diverses règles de liaison

In a given ecosystem, there may be several mechanisms facilitating an observed interaction between a plant and pollinator. Hence pollination networks should be modelled as a mixture of linkage rules. We propose Latent Dirichlet Allocation (LDA) from artificial intelligence to model the observed interactions in an ecosystem as a finite mixture of (latent) interaction groups in which plant and pollinator pairs with common linkage rule(s) are placed in an interaction group. LDA is studied through a simulation study using the BIC for model selection. The results suggest that LDA works well for compartmental webs, but loses accuracy as nest-ness increases.

Dans tout écosystème, divers mécanismes facilitent l'interaction observée entre plantes et pollinisateurs. Les réseaux de pollinisation peuvent donc se modéliser par un mélange de règles de liaison. Nous proposons l'Allocation de Dirichlet latente (LDA) tirée de l'intelligence artificielle pour modéliser ces interactions comme mélange fini de groupes d'interaction (latente), où les paires de plantes et de pollinisateurs partageant des règles de liaison sont regroupées. La LDA est étudiée via une étude de simulation utilisant le critère BIC pour la sélection de modèle. Les résultats indiquent que la LDA fonctionne bien pour les réseaux compartimentés, mais perd en précision quand l'emboîtement augmente.

ARMIN HATEFI & MOHAMMAD JAFARI JOZANI (University of Manitoba)

[Tuesday June 5 / mardi 5 juin, 15:30]

Fisher Information in Ranked Set Samples from Finite Mixture Models

Information de Fisher et échantillonnage d'ensembles ordonnés appliqués aux modèles de mélanges finis

We study the efficiency of ranked set sampling (RSS) relative to simple random sampling (SRS) for making inference about the parameters of a finite mixture model using the Fisher information matrix. We propose different variations of RSS (for both perfect and imperfect ranking mechanisms) and show how to obtain the Fisher information matrix for each variation. Our findings show that statistical inference about different features of finite mixture models based on RSS techniques is more efficient than the corresponding ones using SRS. Theoretical results are augmented by a simulation study based on a mixture of two univariate normal distributions.

Nous étudions l'efficacité des procédures d'échantillonnage d'ensembles ordonnés (EEO) relativement à l'échantillonnage aléatoire simple (EAS) dans un contexte d'inférence à propos des paramètres d'un modèle de mélange fini basée sur la matrice d'information de Fisher. Nous proposons quelques variantes de l'EEO (avec ou sans erreurs de classement) et obtenons la matrice d'information pour chacune d'elles. Nos investigations suggèrent que les techniques d'inférence basées sur l'EEO ont une plus grande efficacité que leurs équivalents basés sur l'EAS. Nous présentons une étude de simulation concernant le mélange de deux populations normales univariées.

2D-C3: Mixture Models, Clustering and Classification Modèles de mélanges, partitionnement et classification

POOYAN KHAJEHPOUR TADAVANI & ALI GHODSI (University of Waterloo)

[Tuesday June 5 / mardi 5 juin, 15:45]

Partitioning High Dimensional Data to Low Dimensional and Localized Subspaces

Partitionnement de données de grande dimension en sous-espaces de petite dimension localisés

It is generally conceivable to assume that high dimensional data points are on submanifolds of the space. These submanifolds can be modeled by a number of linear subspaces. This is the main intuition behind a majority of subspace clustering algorithms. However, subspaces computed by these algorithms, consist of disconnected subsets of the submanifolds and therefore, they do not form localized clusters. We propose Low Dimensional Localized Clustering (LDLC), a new method for subspace clustering. LDLC, unlike existing methods, respects the topology of the submanifolds and assigns the data points to localized clusters such that the total reconstruction error is minimized.

On pense généralement que les points de données de grande dimension se trouvent sur des sous-variétés de l'espace qui peuvent se modéliser par plusieurs sous-espaces linéaires. C'est là l'intuition principale sur laquelle se fondent la plupart des algorithmes de classification des sous-espaces. Toutefois, les sous-espaces calculés par ces algorithmes sont des sous-ensembles de sous-variétés déconnectés et ne forment donc pas de groupes localisés. Nous proposons le Groupement localisé de petite dimension, une nouvelle méthode de regroupement des sous-espaces. Contrairement aux méthodes existantes, elle respecte la topologie des sous-variétés et affecte les points de données à des groupes localisés, minimisant ainsi l'erreur de reconstruction totale.

IRENE VRBIK & PAUL MCNICHOLAS (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 16:00]

Clustering and Classification with Parsimonious Skew-t Mixture Models

Regroupement et classification à l'aide de modèles de mélange parcimonieux de distributions t asymétriques

With the advancement of computer technology, mixture model-based approaches to clustering have become increasingly popular. In recent work, a robust, flexible mixture modelling approach using the skew-t distribution has been explored. We propose a skew-t analogue of the popular MCLUST models that impose an eigenvalue decomposition of the covariance structure. An "exact" EM algorithm is outlined and our approach is applied to some benchmark clustering datasets.

Avec l'avancement des technologies informatiques, les méthodes de regroupement basées sur les modèles de mélange sont devenues de plus en plus populaires. Des travaux récents ont exploré un modèle de mélange robuste et flexible utilisant une distribution t asymétrique. Nous proposons une distribution t asymétrique analogue à celle des modèles populaires MCLUST qui imposent une décomposition en valeurs propres de la structure de covariance. Un algorithme EM "exact" est présenté et notre approche est appliquée au regroupement de certains jeux de données de référence.

BRIAN FRANCAZAK, PAUL MCNICHOLAS & RYAN BROWNE (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 16:15]

The ParSAL Family of Mixture Models

Modèles de mélanges de distributions d'une famille parcimonieuse de lois Laplace asymétriques décentrées

We introduce a family of shifted asymmetric Laplace (SAL) distributions for model-based clustering and classification. These models will arise through eigen-decomposition of the component covariance

Nous présentons une famille de distributions de Laplace asymétriques décentrées pour les modèles d'analyse de regroupement et de classification. Ces modèles proviendront de la décomposition spectrale des matrices de covariance

2D-C3: Mixture Models, Clustering and Classification Modèles de mélanges, partitionnement et classification

matrices. An expectation-maximization (EM) algorithm is used for parameter estimation and the integrated completed likelihood (ICL) for model selection. The capabilities of our parsimonious SAL family are demonstrated on both real and simulated data and the results compared to the popular Gaussian alternative.

des composantes. Un algorithme d'espérance-maximisation (EM) est utilisé pour l'estimation des paramètres et la vraisemblance complétée intégrée sert à la sélection de modèle. Nous illustrons ce que notre famille parcimonieuse de distributions Laplace asymétriques décentrées peut faire sur des données réelles et simulées, et nous la comparons à la populaire alternative gaussienne.

RACHEL O'REILLY & PAUL MCNICHOLAS (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 16:30]

Cross-Validation for Mixture Model Selection

Validation croisée pour sélection de modèles de mélange

In recent years, the use of mixture models for classification and clustering has gained popularity. Although there are existing methods for model selection, such as the BIC, AIC, and ICL, no one approach has proven particularly effective. We investigate the use of cross-validation to assess the consistency of classification under a potential model, and thus assess the certainty of the model in revealing underlying data structure. The procedure will be illustrated using both real and simulated data.

Dans les dernières années, l'utilisation de modèles de mélange pour la classification et le classement a gagné en popularité. Bien que plusieurs méthodes soient disponibles pour la sélection de modèles, tel que le BIC, l'AIC et l'ICL, aucune ne se montre particulièrement efficace. Nous étudions l'utilisation de la validation croisée pour évaluer la performance de la classification pour un modèle potentiel, et évaluons ainsi la fiabilité du modèle pour mettre à jour la structure des données. Notre procédure est illustrée à l'aide de jeux de données réels et simulés.

2D-C4: Modelling the Spread of Infectious Diseases Modélisation de la propagation des maladies infectieuses

Chair/Président: Patrick Brown (Cancer Care Ontario)

MacNaughton Building, room/local 113

Abstracts/Résumés

LIN ZHANG & ROB DEARDON (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 15:15]

Comparing Bayesian Statistical Models of Infectious Disease Outbreaks via Probability Scoring

Comparaison de modèles bayésiens d'épidémies de maladies infectieuses au moyen d'une pondération probabiliste

Model comparison and assessment are key components in statistical modeling. Here we consider these components for infectious disease models. The approach taken will be to develop the tool of proper probability scoring rules to assess calibration and sharpness of probabilistic forecasts for spatio-temporal epidemic models. Scoring rules work by rewarding a numerical value to a model based on the predictive distribution and on the infection/non-infection event observed. We show how such methods can be used and compare the performance of different probability score-based methods.

La comparaison et l'évaluation de modèles sont des composantes clés de la modélisation statistique. Nous considérons ces composantes pour les modèles de maladies infectieuses. Notre approche consiste à développer des règles de pondération basées sur des probabilités appropriées afin d'évaluer la calibration et la précision des prévisions probabilistes pour des modèles épidémiques spatio-temporels. Les règles de pondération attribuent une valeur numérique au modèle en fonction de la distribution prédictive et des événements infectieux et non infectieux observés. Nous montrons comment utiliser ces méthodes et nous comparons la performance de différentes méthodes de pondération probabiliste.

LILIA LETICIA RAMIREZ RAMIREZ (Instituto Tecnológico Autónomo de México (ITAM)), **MARY THOMPSON & YULIA GEL** (University of Waterloo), **EILEEN DE VILLA** (Peel Public Health), **MATT MCPHERSON** (Infonaut)

[Tuesday June 5 / mardi 5 juin, 15:30]

Simulation of Infectious Disease Spread as a Tool to Evaluate Control Measures During an Epidemic Outbreak

Simulation de la propagation d'une maladie infectieuse comme outil pour évaluer les mesures de lutte en cas de flambée épidémique

SIMID (SIMulation of Infectious Diseases) is a program developed with the Region of Peel, Ontario, and Infonaut Inc. to picture general scenarios for outbreaks of directly transmissible diseases. Its main goal is to facilitate for health officers and decision makers the evaluation of diverse control measures, such as prophylactic vaccination, school closures and isolation. The program was implemented for the Region of Peel, using its social and demographic characteristics to model an individual contact network with emphasis on school contacts. In this presentation we talk about the program's theoretical bases, its algorithms, its main characteristics and its possible extensions.

SIMID (SIMulation of Infectious Diseases) est un programme développé avec la région de Peel (Ontario) et Infonaut Inc. pour imaginer des scénarios de flambée de maladies directement transmissibles. Il vise à aider les agents de santé et les décideurs à évaluer diverses mesures de lutte telles la vaccination prophylactique, la fermeture des écoles et l'isolement. Il a été mis en œuvre dans la région de Peel en utilisant ses caractéristiques sociales et démographiques pour modéliser un réseau de contact individuel, notamment dans les écoles. Nous en discutons les bases théoriques, les algorithmes, les caractéristiques principales et les éventuelles extensions.

2D-C4: Modelling the Spread of Infectious Diseases Modélisation de la propagation des maladies infectieuses

NADIA BIFOLCHI, ROB DEARDON & ZENY FENG (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 15:45]

Spatial Approximations of Network-based Individual Level Infectious Disease Models

Approximations spatiales des modèles de maladie infectieuse individualisés en réseau

Often, when modeling infectious disease spread, the complex network through which the disease propagates is approximated by simpler spatial information. Here, we simulate epidemic spread through various contact networks and fit spatial-based models in a Bayesian framework using Markov chain Monte Carlo methods. These spatial models are individual-level models which account for the spatio-temporal dynamics of infectious disease. The focus here is on choosing a spatial model which best predicts the true probabilities of infection, as well as determining under which conditions such spatial models fail.

Les modèles de propagation des maladies infectieuses approximent souvent les réseaux complexes de propagation par des informations spatiales simplifiées. Ici, nous simulons la propagation via divers réseaux de contact et ajustons des modèles spatiaux dans un cadre bayésien par des méthodes Monte Carlo par chaîne de Markov. Ces modèles spatiaux sont des modèles individualisés qui tiennent compte de la dynamique spatio-temporelle de la maladie infectieuse. Il s'agit de sélectionner le modèle qui prédit le mieux les probabilités d'infection véritables et de déterminer dans quelles conditions ces modèles échouent.

LORNA DEETH, ROB DEARDON & DANIEL GILLIS (University of Guelph)

[Tuesday June 5 / mardi 5 juin, 16:00]

Variations of the Deviance Information Criterion for Infectious Disease Models

Variations sur le critère d'information de déviance pour les modèles de maladies infectieuses

The deviance information criterion (DIC) is a model comparison tool suited for complex models fitted within a Bayesian framework. Since its development, variations of the DIC have been proposed for application to missing data models. The DIC as a method of model selection is investigated when applied to latent conditional individual-level models (LC-ILMs), a class of models that can be used for infectious disease modeling, and which display mixture model-like characteristics due to their dependence on a latent grouping variable. The effectiveness of the traditionally defined DIC is compared to alternative definitions, to assess which is most applicable for LC-ILMs.

Le critère d'information de déviance (CID) est un outil de comparaison adapté aux modèles complexes ajustés dans un cadre bayésien. Depuis son apparition, des variantes du CID ont été proposées pour l'application aux modèles de données manquantes. L'utilisation du DIC en tant que méthode de sélection de modèles est étudiée pour l'application aux modèles conditionnels à variable latente au niveau individuel (LC-ILM), une classe de modèles qui peuvent être utilisés pour la modélisation de maladies infectieuses et qui offrent des caractéristiques semblables aux modèles de mélange à cause de la présence de la variable de groupe latente. L'efficacité du CID traditionnel est comparée à celle de définitions alternatives afin d'identifier la définition la plus adaptée aux modèles LC-ILM.

CHEN ZHANG & ALWELL OYET (Memorial University of Newfoundland)

[Tuesday June 5 / mardi 5 juin, 16:15]

Estimation and Forecasting of a Lag 2 Dynamic Model for Infectious Diseases

Estimation et prévision d'un modèle dynamique à décalage de deux périodes pour les maladies infectieuses

We discuss estimation and forecasting of an extension to lag 1 longitudinal dynamic model for cor-

Nous discutons l'estimation et la prévision d'une généralisation du modèle dynamique longitudinal à décalage d'une

2D-C4: Modelling the Spread of Infectious Diseases Modélisation de la propagation des maladies infectieuses

related data used by Oyet & Sutradhar (2011) for modelling the spread of infectious disease. The lag 1 model only allow individuals with infectious at time point $t - 1$ to cause new infectious at time point t . Clearly, if at time point $t - 2$, there is an individual who is still infected by the disease, it is possible for this individual to infect others at time point t . The present model discussed in this paper allows for such a possibility.

période pour des données corrélées utilisées par Oyet & Sutradhar (2011) pour modéliser la propagation d'une maladie infectieuse. Le modèle à décalage d'une période permet aux individus infectés à la période $t - 1$ de propager la maladie uniquement à la période t . Il est cependant évident qu'un individu infecté à la période $t - 2$ qui est toujours porteur de la maladie à la période t peut infecter d'autres personnes. Le modèle présenté prévoit cette éventualité.

RAJAT MALIK, ROB DEARDON & GRACE P.S. KWONG (University of Guelph), **BEN COWLING** (University of Hong Kong)

[Tuesday June 5 / mardi 5 juin, 16:30]

Modeling the Spread of Influenza within Households in Hong Kong

Modéliser la propagation de l'influenza dans les ménages de Hong Kong

Individual-level models (ILMs) can be used to model the spread of infectious diseases in discrete time. The key feature of these ILMs is that they can take into account covariate information on susceptible and infectious individuals as well as shared covariate information such as geography or contact measures. Here, such ILMs are fitted in a Bayesian framework using Markov Chain Monte Carlo techniques to data from a study on influenza within households in Hong Kong during 2009. The focus here is to estimate the effect of vaccination on infection risk and choose a model that best fits the disease spread.

Les modèles de niveau individuel (ILM) peuvent être utilisés pour modéliser l'étendue de maladies infectieuses en temps discret. Une caractéristique clé des ILM est qu'ils peuvent prendre en compte l'information sur des covariables propres aux individus infectés ou susceptibles de l'être, ainsi que des covariables communes telles que la géographie ou des mesures de contact. Dans notre exposé, les ILM sont ajustés dans un contexte bayésien à l'aide de techniques de chaînes de Markov Monte Carlo pour étudier l'influenza dans les ménages de Hong Kong en 2009. L'objectif principal est d'estimer l'effet de la vaccination sur le risque d'infection et de choisir un modèle qui décrit le mieux possible la propagation de la maladie

3A-A1: CJS Award Address
Allocution du récipiendaire du Prix de la RCS

Chair/Président: Paul McNicholas (University of Guelph)
Organizer/Responsable: Denis Larocque (HEC Montréal)

Rozanski Hall, room/local 101

Abstract/Résumé

CHONG GU AND PING MA / CHONG GU ET PING MA (Purdue University; University of Illinois, Urbana-Champaign)

[Wednesday June 6 / mercredi 6 juin, 8:45]

Nonparametric regression with cross-classified responses

Régression non-paramétrique avec variables réponses provenant de classifications croisées

We develop nonparametric regression models with cross-classified responses, effectively “unwinding” contingency tables along an “x-axis” and estimating table probabilities as functions of covariates. Conditional independence structures can be explored/exploited via selective term eliminations in functional ANOVA decompositions, and confidence intervals are available for “y-contrasts” such as odds ratios among table margins. Random effects can also be incorporated for the handling of correlated data. The modeling tools are implemented in a suit of R functions, and will be illustrated in a few examples concerning the elucidation of the transcriptional regulatory networks using gene expression and transcription factor binding motifs.

Nous développons des modèles de régression non-paramétriques avec variables réponses provenant de classification croisées. Notre méthode revient à « dérouler » des tableaux de contingence selon un « axe des x » et à estimer les probabilités des cellules en fonction de covariables. Les structures d’indépendance conditionnelle peuvent être explorées/exploitées via une élimination sélective des termes dans une décomposition de type anova fonctionnelle. Des intervalles de confiance peuvent être calculés pour des « contrastes y » tels que des rapports de cote d’effectifs marginaux. Des effets aléatoires peuvent aussi être inclus pour modéliser des données corrélées. Les méthodes de modélisations sont implémentées dans une librairie R, et seront illustrées à l’aide de quelques exemples concernant l’élucidation de réseaux de régulation de la transcription à l’aide de données sur l’expression de gènes et des matrices (motifs) de sites de fixation de facteur de transcription.

3A-A2: CRM-SSC Prize in Statistics Address Allocution du récipiendaire du Prix CRM-SSC

Organizer and Chair / Responsable et président: David Stephens (McGill University)

Rozanski Hall, room/local 104

Abstract/Résumé

CHANGBAO WU (University of Waterloo), **JIAHUA CHEN** (University of British Columbia), **JAE-KWANG KIM** (Iowa State University)

[Wednesday June 6 / mercredi 6 juin, 8:45]

Semiparametric Fractional Imputation and Sparse and Efficient Replication Variance Estimation for Complex Surveys
Imputation fractionnaire semi-paramétrique et estimation parcimonieuse et efficace de la variance par réplication pour des enquêtes complexes

This talk covers two seemingly different but related topics. We first propose a semiparametric fractional imputation method for handling item nonresponses, assuming certain baseline auxiliary variables are observed for all units in the sample. The proposed strategy combines the strengths of conventional single imputation and multiple imputation methods, and is easy to implement. We then discuss how to achieve sparsity and efficiency on replication variance estimation in the sense that a small number sets of replication weights can produce valid and efficient variance estimators for various finite population parameters. Simulation results and some related issues will be presented and discussed.

Cette présentation aborde deux sujets qui semblent différents, mais qui sont en fait reliés. Nous proposons d'abord une méthode semi-paramétrique d'imputation fractionnaire pour la non-réponse partielle qui suppose que certaines variables auxiliaires sont mesurées pour toutes les unités de l'échantillon. Facile à mettre en œuvre, la stratégie proposée combine les avantages de l'imputation unique et de l'imputation multiple conventionnelles. Nous discutons ensuite de la façon d'obtenir parcimonie et efficacité pour l'estimation de la variance par réplication, dans le sens où un petit nombre de poids de réplication peuvent produire des estimateurs efficaces de la variance pour plusieurs paramètres de population finie. Des résultats de simulation seront présentés, ainsi que certains problèmes connexes.

3A-A3: Pierre Robillard Award Address
Allocution du récipiendaire du Prix Pierre-Robillard

Organizer and Chair / Responsable et présidente: Keumhee Carriere Chough (University of Alberta)

Rozanski Hall, room/local 103

Abstract/Résumé

BEI CHEN (McMaster University), **BOVAS ABRAHAM** & **YULIA GEL** (University of Waterloo)

[Wednesday June 6 / mercredi 6 juin, 8:45]

Linearization Methods in Time Series Analysis

Méthodes de linéarisation pour l'analyse de séries chronologiques

Many real life time series exhibit patterns that can be well-described using nonlinear processes. A drawback of nonlinear models is that parameter estimation can be computationally expensive and sensitive to initial values. In this talk, I propose a class of computationally efficient techniques which approximate nonlinear processes by linear ones, so-called linearization methods. First, I present a regularized autoregressive method for estimating multiple frequencies in sinusoidal processes. Second, I introduce a bootstrap scheme based on a linear representation of GARCH models for forecasting and diagnostic testing.

Plusieurs séries chronologiques tirées de données réelles ont des comportements qui peuvent être décrits à l'aide de processus non-linéaires. L'estimation des paramètres d'un modèle non linéaire peut demander une grande puissance de calcul et être sensible aux valeurs initiales, un désavantage pour ces modèles. Dans cet exposé, je propose une classe de techniques efficaces en termes de temps de calcul qui approximent les processus non-linéaires à l'aide de processus linéaires (méthodes de linéarisation). Je présente d'abord une méthode autorégressive linéaire pour estimer des fréquences multiples dans un processus sinusoidal. J'introduis ensuite une méthode bootstrap basée sur une représentation linéaire des modèles GARCH pour la prévision et les tests diagnostiques.

3B-I1: Confidentiality Confidentialité

Chair/Président: Mike Hidioglou (Statistics Canada / Statistique Canada)

Organizer/Responsable: Jean-Louis Tambay (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Alexander Hall, room/local 100

Abstracts/Résumés

ANNE-SOPHIE CHAREST (Université Laval)

[Wednesday June 6 / mercredi 6 juin, 10:15]

Differential Privacy for Synthetic Datasets

Confidentialité différentielle pour jeux de données synthétiques

Any organization wishing to publish statistical data must respect the confidentiality agreement they have with their respondents. This task is complicated by the increase in personal information available in recent years, and it thus becomes essential to define clearly the desired protection. We present here the criterion of differential privacy, a strict criterion which protects respondents against an adversary with arbitrary available resources. We explain how to generate synthetic datasets satisfying differential privacy, and discuss the quality of inference from such data.

Toute organisation qui souhaite publier des données statistiques doit s'assurer de respecter son engagement de confidentialité envers ses répondants. Avec la multiplication de l'information disponible sur chaque individu, cette tâche est de plus en plus ardue et il devient important de définir clairement la protection désirée. Nous présentons ici le critère de confidentialité différentielle, un critère très strict qui n'assume rien à propos des ressources disponibles à l'adversaire cherchant à révéler les données confidentielles. Nous expliquons comment générer des données synthétiques satisfaisant la confidentialité différentielle et discutons de la qualité de l'inférence avec ces données.

ÉRIC GAGNON & VALÉRIE ROY (Institut de la statistique du Québec)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Statistical Disclosure Control of Remote Access Microdata Files

Contrôle de divulgation statistique pour fichiers de microdonnées accessibles à distance

For several years now, the Institut de la statistique du Québec (ISQ) has been offering remote access to its survey microdata files. This secure way of accessing data is increasingly popular among external researchers. The terms and conditions and procedure of remote access will be discussed at the beginning of the presentation. Then we will look at the statistical disclosure control (SDC) of microdata accessed remotely. SDC causes a slight decrease in the analytical potential of microdata. Certain results regarding the information loss in those files will also be presented.

Depuis quelques années, l'Institut de la statistique du Québec (ISQ) offre l'accès à distance à des fichiers de microdonnées provenant de ses enquêtes. Cet accès sécuritaire est de plus en plus populaire auprès des chercheurs externes. Les différentes modalités encadrant cet accès ainsi que son fonctionnement seront tout d'abord abordés lors de cette présentation. Ensuite, il sera question du contrôle de divulgation statistique (CDS) appliqué aux microdonnées accessibles à distance. Une faible diminution du potentiel analytique des microdonnées résulte du CDS. À ce sujet, certains résultats concernant la perte d'information sur ces fichiers seront également présentés.

3B-I1: Confidentiality Confidentialité

JULIA HOENINGER (Federal Statistical Office, Germany), **JÖRG HÖHNE** (State Statistical Institute Berlin-Brandenburg, Germany)

[Wednesday June 6 / mercredi 6 juin, 11:15]

Morpheus – Remote Access to Micro Data

Morpheus – Accès à distance aux microdonnées

Providing remote micro data access and automatically checking statistical output are the goals of a novel approach called Morpheus. In this system, researchers work remotely on anonymous micro data files with common statistical software packages and get their results back in real time. Thus, while developing program syntax, helpful data browsing is possible. Additionally, a quality measure is provided for every single result such that researchers can see themselves if the anonymous results are close to the results with original data. Work in progress of this approach to remote access to data of official statistics in Germany will be presented.

Une nouvelle approche, intitulée Morpheus, vise à fournir un accès à distance aux microdonnées et à vérifier automatiquement les résultats statistiques. Ce système permet aux chercheurs de travailler à distance sur des fichiers de microdonnées anonymes à l'aide de logiciels statistiques courants, et d'obtenir les résultats en temps réel. Ainsi, tout en développant la syntaxe du programme, une consultation utile des données est possible. De plus, une mesure de qualité est associée à chaque résultat, permettant aux chercheurs de vérifier si les résultats anonymes sont comparables aux résultats obtenus à l'aide des données originales. Nous présentons un projet en cours qui utilise cette approche pour accéder à distance à des données de statistiques officielles en Allemagne.

3B-I2: Current Issues in the Design and Analysis of Cluster Randomization Trials Problèmes actuels dans la conception et l'analyse d'essais randomisés par grappes

Organizer and Chair / Responsable et président: Yutaka Yasui (University of Alberta)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Rozanski Hall, room/local 104

Abstracts/Résumés

ALLAN DONNER (University of Western Ontario)

[Wednesday June 6 / mercredi 6 juin, 10:15]

Current Issues in the Design and Analysis of Cluster Randomization Trials

Problèmes actuels dans la conception et l'analyse d'essais randomisés par grappes

Cluster randomization trials are those that randomize intact social units or clusters of individuals to different intervention groups. Such trials are particularly widespread in the evaluation of educational programs and innovations in the provision of health care. This talk will discuss issues in the design and analysis of such trials that have proved controversial among practitioners. They include the choice of unit of inference, potential operational threats to trial validity, factors influencing the selection of a design, consideration of when non-randomized allocation might be preferred, and recent work focusing on ethical issues.

Les essais randomisés par grappes répartissent des unités sociales intactes ou des grappes d'individus dans différents groupes d'intervention. Ces essais sont particulièrement répandus dans l'évaluation de programmes d'éducation et d'innovations dans la prestation de soins de santé. Dans cet exposé, nous discutons de problèmes dans la conception et l'analyse d'essais qui sont controversés au sein des praticiens. Les thèmes abordés comprennent le choix de l'unité d'inférence, les menaces opérationnelles potentielles quant à la validité des essais, les facteurs ayant une influence sur le choix du plan, les cas où la répartition non-aléatoire peut être préférable et les travaux récents axés sur des problèmes éthiques.

3B-I3: Spatial and Environmental Statistics Statistique spatiale et environnementale

Organizer and Chair / Responsable et président: Theodoro Koulis (University of Manitoba)

Rozanski Hall, room/local 103

Abstracts/Résumés

ROB DEARDON (University of Guelph)

[Wednesday June 6 / mercredi 6 juin, 10:15]

Computationally Efficient Forms of Individual-Level Models for Large-Scale Spatial Infectious Disease Systems
Formes de modèles à temps de calcul efficace pour systèmes spatiaux de maladies infectieuses à grande échelle au niveau des individus

Individual-level models (ILMs) for infectious disease, fitted in a Bayesian MCMC framework, are an intuitive and flexible class of models that can be used to take into account population heterogeneity via various individual-level covariates. ILMs containing a geometric distance kernel to account for geographic heterogeneity provide a natural way to model the spatial spread of disease. However, in even moderately large populations, the likelihood calculations required can be prohibitively time consuming. It is possible to speed up the computation via techniques that make use of a linearized distance kernel. Here we examine some methods of carrying out such analyses.

Les modèles pour les maladies infectieuses ajustés dans un cadre bayésien de Monte Carlo par chaînes de Markov constituent une classe intuitive et flexible de modèles qui peuvent être utilisés pour prendre en compte l'hétérogénéité de la population à l'aide de diverses covariables au niveau individuel. Les modèles ayant un noyau de distance géométrique pour tenir compte de l'hétérogénéité géographique sont un choix naturel pour modéliser la propagation spatiale de la maladie d'un individu à l'autre. Cependant, même dans le cas de populations de taille modérée, les calculs par la vraisemblance peuvent demander un temps de calcul exorbitant. Il est possible d'accélérer ces calculs en linéarisant le noyaux de distance. Dans cet exposé, nous examinons quelques méthodes pour effectuer ce genre d'analyse.

SUBASH LELE, PETER SOLYMOS & ERIN BAYNE (University of Alberta)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Zero Inflated Models, Detection Error and Conditional Likelihood Inference with Applications in Ecology
Modèles à surabondance de zéros, erreur de détection et inférence fondée sur la vraisemblance conditionnelle avec applications en écologie

Occupancy and abundance surveys are fraught with detection error where a species that is present may not be detected. In addition, locations may have zero abundance because there are not enough individuals to occupy all the suitable sites or because the species has not dispersed to the location as yet. We use zero inflated count and ordinal data models for analyzing these data. We show that the parameters of the detection model, zero inflation and the intercept of the abundance model are confounded. We use conditional likelihood to separate these highly confounded parameters and conduct appropriate scientific inference.

Les enquêtes d'abondance et d'occupation sont parsemées d'erreurs de détection où une espèce présente demeure parfois non détectée. De plus, certains endroits peuvent avoir une abondance nulle car il n'y a pas assez d'individus pour occuper tous les endroits appropriés ou parce que l'espèce ne s'y est pas encore dispersée. Nous utilisons des modèles de données discrètes et ordinales à surabondance de zéros pour analyser ces données. Nous montrons que les paramètres du modèle de détection, la surabondance de zéro et l'ordonnée à l'origine du modèle d'abondance sont indiscernables. Nous utilisons la vraisemblance conditionnelle pour distinguer ces paramètres confondus et étayons une conclusion scientifique appropriée.

3B-I3: Spatial and Environmental Statistics Statistique spatiale et environnementale

JEFFREY PICKA (University of New Brunswick)

[Wednesday June 6 / mercredi 6 juin, 11:15]

Assessing the Fit of Models for Disordered Spatial Patterns

Évaluation de l'adéquation de modèles pour les structures spatiales désordonnées

Disordered spatial patterns can be modeled both by classical probabilistic spatial processes and by models which attempt to represent the dynamics of the processes from which the patterns emerge. In the latter case, these models can only be said to represent nature if they are stochastic and if their fit to data is assessed by statistical inference. The sources of uncertainty in these models will be discussed, together with conditions for when model assessment is possible. Methods for model assessment based on multivariate and spatial statistics will be proposed. Examples will be taken from forest fire evolution and powder flow.

Les structures spatiales désordonnées peuvent être modélisées par des modèles probabilistes spatiaux classiques ou par des modèles qui tentent de représenter la dynamique des processus à partir desquels les structures émergent. Dans ce dernier cas, on suppose que ces modèles représentent la nature seulement s'ils sont stochastiques et si leur ajustement aux données est confirmée par l'inférence statistique. Nous discuterons des sources d'incertitude dans ces modèles, ainsi que des conditions rendant leur analyse possible. On proposera des méthodes d'analyse des modèles basées sur des statistiques multivariées et spatiales. Des exemples seront tirés de l'évolution des feux de forêt et de l'écoulement des poudres.

3B-I4: Stochastic Models and Applications Modèles stochastiques et applications

Organizer and Chair / Responsable et président: Shui Feng (MacMaster University)

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

Alexander Hall, room/local 200

Abstracts/Résumés

DELI LI (Lakehead University)

[Wednesday June 6 / mercredi 6 juin, 10:15]

On Jiang's Asymptotic Distribution of the Largest Entry of a Sample Correlation Matrix

À propos de la distribution asymptotique de Jiang du plus grand élément d'une matrice de corrélation échantillonnale

Consider a double array of i.i.d. random variables and its sample correlation matrix. In this talk we present some recent results on the weak asymptotic behaviour, especially Jiang's asymptotic distribution of the largest entry of this correlation matrix which was introduced by Jiang (2004) as a test statistic in a statistical hypothesis testing problem.

Nous considérons une matrice infinie de variables aléatoires i.i.d. et sa matrice de corrélation échantillonnale. Dans cet exposé, nous présentons quelques résultats récents sur le comportement asymptotique faible. Plus particulièrement, nous étudions la distribution asymptotique du plus grand élément de la matrice de corrélation, présentée par Jiang (2004) comme une statistique de test dans un problème de test d'hypothèses.

MAHMOUD ZAREPOUR & LUAI AL LABADI (University of Ottawa)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Some Asymptotic Theory in Nonparametric Bayesian Inference

Résultats asymptotiques pour l'inférence bayésienne non-paramétrique

The Dirichlet process and the Poisson-Dirichlet process play an important role in nonparametric Bayesian inference. In this talk, we introduce a consistent process facilitating easy and rapid sampling from a Dirichlet process. Then we let $P_{a,\theta,H}$ be a Poisson-Dirichlet process on \mathbb{R} with a base probability measure H and parameters $a > 0$ (concentration) and θ . We show that $\sqrt{a}(P_{a,\theta,H}((-\infty, t]) - H((-\infty, t]))$ converges to a Brownian bridge as $a \rightarrow \infty$. This shows a striking similarity with the empirical process. We also derive a Glivenko-Cantelli theorem for the Poisson-Dirichlet process. The weak convergence of the quantile process is also obtained.

Les processus de Dirichlet et de Poisson-Dirichlet jouent un rôle important dans l'inférence bayésienne non-paramétrique. Dans cet exposé, nous présentons un processus convergent facilitant l'échantillonnage simple et rapide d'un processus de Dirichlet. Supposons ensuite que $P_{a,\theta,H}$ soit un processus de Poisson-Dirichlet sur \mathbb{R} avec une mesure de probabilité de base H et des paramètres $a > 0$ (concentration) et θ . Nous montrons que $\sqrt{a}(P_{a,\theta,H}((-\infty, t]) - H((-\infty, t]))$ converge vers un pont brownien lorsque $a \rightarrow \infty$. La similitude par rapport à un processus empirique est frappante. Nous dérivons également un théorème Glivenko-Cantelli pour le processus de Poisson-Dirichlet. Nous obtenons aussi la convergence faible du processus des quantiles.

3B-I4: Stochastic Models and Applications Modèles stochastiques et applications

FANG XU & SHUI FENG (McMaster University)

[Wednesday June 6 / mercredi 6 juin, 11:15]

The Gamma-Dirichlet Structure and the Quasi Invariance Property
La structure de Gamma-Dirichlet et la propriété de quasi-invariance

The Gamma-Dirichlet structure corresponds to the decomposition of the gamma process into the independent product of a gamma random variable and a Dirichlet process. This structure enables us to obtain the quasi invariance of the two-parameter Poisson-Dirichlet distribution. The dynamic version of the Gamma-Dirichlet structure can be seen as the relationship between the measure-valued branching diffusion with immigration (MBI) and the Fleming-Viot process with parent independent mutation (FV). From the dynamical perspective, we derived the transition function of the FV process from that of the MBI process, and established the reversibility of the latter.

La structure de Gamma-Dirichlet correspond à la décomposition d'un processus gamma en un produit indépendant d'une variable aléatoire gamma et d'un processus de Dirichlet. Cette structure permet d'obtenir la quasi-invariance de la distribution Poisson-Dirichlet à deux paramètres. Une version dynamique de la structure Gamma-Dirichlet peut être vue comme la relation entre la diffusion de branchement avec immigration à valeurs dans un espace de mesures (MBI) et le processus Fleming-Viot avec mutation indépendante du parent (FV). Du point de vue dynamique nous dérivons la fonction de transition du processus FV à partir de celle du processus MBI, et prouvons la réversibilité de ce second processus.

3B-15: Teaching Statistical Consulting with a Mixture of Classroom and Practicum Enseignement de la consultation statistique: un mélange d'exposés magistraux et d'expériences pratiques

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

Session sponsored by the Statistical Education Committee

Séance parrainée par le Comité d'éducation en statistique

Rozanski Hall, room/local 101

Abstracts/Résumés

TONY PANZARELLA (University of Toronto)

[Wednesday June 6 / mercredi 6 juin, 10:15]

Lab Course in Biostatistics – Providing Practical Experience for Future Consulting Biostatisticians

Cours de laboratoire en biostatistique - Fournir une expérience pratique de consultation aux futurs biostatisticiens

The Lab Course in Statistical Design and Analysis is a mandatory course for graduate students within the Division of Biostatistics at the Dalla Lana School of Public Health, University of Toronto. In addition to classroom lectures it provides students with hands-on experience, in the form of a practicum. Each student is paired with a biostatistician-mentor, and works at the mentor's worksite several hours per week. This format has been followed for several years now and is popular with students, especially the practicum component. This presentation provides a summary of the course from the perspective of its director and lead instructor.

Le cours de laboratoire en planification et en analyse statistique est obligatoire pour tous les étudiants des cycles supérieurs de la division biostatistique de l'École de santé publique Dalla Lana à l'Université de Toronto. En plus de présentations en classe, ce cours fournit une expérience pratique sous la forme d'un stage. Chaque étudiant est jumelé à un mentor biostatisticien et travaille sur le lieu de travail de celui-ci plusieurs heures par semaine. Ce format, utilisé depuis plusieurs années, est populaire auprès des étudiants, surtout pour sa composante pratique. Cette présentation résume le cours du point de vue de son directeur et de son instructeur principal.

TAMARA ARENOVICH (University of Toronto)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Biostatistics Practicum – A Supervisor's Perspective

Stage en biostatistique - La perspective d'une superviseure

The Division of Biostatistics at the Dalla Lana School of Public Health, University of Toronto, offers a laboratory course in statistical design and analysis to all graduate students. This mandatory course provides students with the opportunity to gain practical experience in data analysis in a professional setting. This presentation provides a summary of the program from the perspective of a practicum site coordinator. Tamara Arenovich is the Manager of the Biostatistical Consulting Service at the Centre for Addiction and Mental Health

La division biostatistique de l'École de santé publique Dalla Lana à l'Université de Toronto offre un cours de travaux pratiques en planification et en analyse statistique à tous ses étudiants des cycles supérieurs. Ce cours obligatoire permet aux étudiants d'obtenir de l'expérience pratique en analyse de données dans un contexte professionnel. Cette présentation résume le programme du point de vue d'une coordonnatrice des stages. Tamara Arenovich, gestionnaire du service de consultation en biostatistique du *Centre for Addiction and Mental Health* à Toronto, a coordonné les stages au cours des six dernières années.

3B-I5: Teaching Statistical Consulting with a Mixture of Classroom and Practicum Enseignement de la consultation statistique: un mélange d'exposés magistraux et d'expériences pratiques

in Toronto and has served as a practicum site coordinator for the past six years.

DAVID SOAVE (University of Toronto)

[Wednesday June 6 / mercredi 6 juin, 11:15]

Statistical Consulting with a Mixture of Classroom and Practicum – A Student Perspective

Consultation statistique avec un mélange de cours et de pratique – Une perspective étudiante

This presentation provides a summary of a student experience in the Lab Course in Statistical Design and Analysis for biostatistics graduate students at the Dalla Lana School of Public Health, University of Toronto. Both the in-class lecture and the practicum placement components are reviewed with specific focus on the placement. In addition to a general retrospective of the valuable experience, details regarding a priori expectations, the placement interview process and the practicum project are all discussed.

Cette présentation donne un aperçu de l'expérience d'un étudiant pour le cours de travaux pratiques en planification et en analyse statistique s'adressant aux étudiants des cycles supérieurs en biostatistiques à l'École de santé publique Dalla Lana à l'Université de Toronto. Les cours et l'expérience pratique seront tous deux passés en revue, en mettant l'accent sur la pratique. En plus d'une rétrospective générale de cette expérience utile, des détails concernant les attentes a priori, le processus d'interview pour le placement des étudiants et le projet en tant que tel seront présentés.

3B-C1: Clustering and Classification Méthodes de partitionnement et de classification

Chair/Président: Matías Salibián-Barrera (University of British Columbia)

MacNaughton Building, room/local 105

Abstracts/Résumés

MALIKA CHARRAD, NADIA GHAZZALI, AZAM NIKNAFS, VÉRONIQUE BOITEAU & MARC-OLIVIER ROBERT LAMBERT (Université Laval)

[Wednesday June 6 / mercredi 6 juin, 10:15]

An Examination of Indices for Determining the Number of Clusters : the Nb.Clusters Package

Analyse des indices de détermination du nombre de classes : la librairie Nb.Clusters

The Nb.Clusters package is developed to test and compare the performance and the ability of a wide range of indices to detect the relevant number of clusters in a data set. It provides thirty indices to determine the number of clusters and some hierarchical and non hierarchical clustering methods and offers the user the best clustering scheme from different results. In this presentation, we will provide a brief overview of implemented clustering algorithms and indices and we will demonstrate how this package can be used to find the optimal partitioning of real and artificial data sets.

La librairie Nb.Clusters est développée pour tester et comparer la performance et l'aptitude d'un ensemble d'indices pour détecter le bon nombre de classes dans un jeu de données. Il permet d'appliquer trente indices de détermination du nombre de classes ainsi que des méthodes hiérarchiques et non hiérarchiques de classification et de proposer à l'utilisateur la meilleure partition. Dans cette présentation, nous allons donner un aperçu sur les algorithmes de classification et les indices implémentés et nous allons démontrer comment cette librairie package peut être utilisée pour identifier la meilleure partition pour des jeux de données réelles et artificielles.

THUVARAHA VANNIYASINGAM & LEHANA THABANE (McMaster University), **REITZE N. RODSETH** (University of KwaZulu-Natal, South Africa), **GIOVANA A. LURATI BUSE & DANIEL BOLLIGER** (University Hospital Basel, Switzerland), **BRUCE M. BICCARD** (University of KwaZulu-Natal, South Africa), **CHRISTOPH S. BURKHART** (University Hospital Basel, Switzerland)

[Wednesday June 6 / mercredi 6 juin, 10:30]

Using the Minimum P-Value Method to Predict the Occurrence of Cardiovascular Complications Within 30 Days of Vascular Surgery

Utilisation de la méthode de la valeur p minimale pour prévoir l'apparition de complications cardiovasculaires dans les 30 jours suivant une chirurgie vasculaire

The minimum p-value approach is often used to determine a single threshold value that stratifies patients into high-risk and low-risk categories. The cutpoint is based on the maximum chi-squared statistic or, equivalently, the minimum p-value. In vascular surgery patients hormones secreted from the heart (B-type natriuretic peptide and N-terminal fragment) have been independently associated with cardiovascular complications. In this individual patient data meta-analysis, we aim to deter-

La méthode de la valeur p minimale est souvent utilisée pour déterminer une valeur de seuil unique permettant de séparer les patients en catégories de risque de niveau élevé et faible. La valeur seuil est basée sur la variable du khi-carré maximale ou, de façon équivalente, sur la valeur p minimale. Dans le cadre des chirurgies vasculaires, les hormones sécrétées par le cœur des patients (peptide natriurétique de type B et fragment N-terminal) ont été associées de façon indépendante aux complications cardiovasculaires. Dans cette méta-analyse de données individuelles, nous

3B-C1: Clustering and Classification

Méthodes de partitionnement et de classification

mine the optimal threshold to predict cardiac death and nonfatal myocardial infarction within 30 days of vascular surgery using the minimum p-value method.

voulons déterminer, à l'aide de la méthode de la valeur p minimale, le seuil optimal pour tenter de prévoir les morts cardiaques et les infarctus du myocarde non mortels survenant dans les 30 jours suivant une chirurgie vasculaire.

TOLUPE SAJOBI, LISA LIX & BOLANLE DANSU (University of Saskatchewan)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Robust Measures of Variable Importance for Two Independent Groups

Mesures robustes de l'importance des variables pour deux groupes indépendants

Robust variable importance measures based on descriptive discriminant analysis are developed using trimmed estimators. The effects of non-normality, mean configuration, number of variables, and correlation structure on the proportion of correctly ranked variables for each measure were investigated using Monte Carlo techniques. Measures of performance include percentages of average any-variable and per-variable correct ranking. The average any-variable and per-variable correct ranking for robust variable importance measures were about 12% and 6% higher than those based on least squares estimators when the data were non-normal. The proposed measures are useful for identifying the variables that discriminate between groups in non-normal data.

Nous développons des mesures robustes de l'importance des variables fondées sur l'analyse discriminante descriptive à l'aide d'estimateurs tronqués. Nous étudions les effets de la non-normalité, de la configuration moyenne, du nombre de variables et de la structure de corrélation sur la proportion de variables correctement classées pour chaque mesure par des techniques de Monte Carlo. Les mesures de la performance incluent le pourcentage de classification correcte moyen de toutes les variables, et variable par variable. Pour les mesures robustes de l'importance des variables, ce pourcentage est d'environ 12% et 6% supérieur à celui trouvé par les estimateurs des moindres carrés ordinaires lorsque les données sont non normales. Les mesures proposées sont utiles pour identifier les variables qui discriminent entre les groupes pour des données non normales.

JABED HOSSAIN TOMAL, WILLIAM WELCH & RUBEN H. ZAMAR (University of British Columbia)

[Wednesday June 6 / mercredi 6 juin, 11:00]

Ensembling Descriptor Sets to Rank Activity of Compounds in QSAR Studies

Assemblage d'ensembles de descripteurs pour classer l'activité des substances dans les études RQSA

In QSAR studies, molecular descriptors are used to model biological activity of compounds. The statistical model aims to rank rare actives early in a list of compounds. There exist several descriptor sets, and a particular set might do well in ranking activity of compounds for some assays but fail to do well for other assays. The classifier "random forest" has been found highly accurate in QSAR studies. To enhance its performance, we propose an ensemble method which not only gives good predicted activity rankings but also confirms to be computationally efficient. We demonstrate our method in four assay datasets.

Dans les études RQSA, des descripteurs moléculaires sont utilisés pour modéliser l'activité biologique des substances. Le modèle statistique vise à classer les actifs rares au début d'une liste de substances. Il existe plusieurs ensembles de descripteurs, dont l'un peut bien fonctionner pour certains essais, mais pas pour d'autres. Le classificateur « forêt aléatoire » s'avère très efficace dans les études RQSA. Pour en améliorer la performance, nous proposons une méthode d'ensemble qui non seulement donne une bonne classification d'activité prévue, mais s'avère aussi efficace d'un point de vue informatique. Nous l'illustrons par quatre ensembles de données d'essais.

3B-C1: Clustering and Classification

Méthodes de partitionnement et de classification

DANIEL GILLIS (University of Guelph)

[Wednesday June 6 / mercredi 6 juin, 11:15]

Signal Processing for Species Identification

Traitement des signaux pour l'identification d'espèces

Industrial processes often require water intake from the Great Lakes for cooling purposes. This process has the side effect of entraining fish and fish larvae from various species; the list of which is not fully known. A method for analyzing an unlabelled biological mash sampled from intake reservoirs is presented. The method uses Bayesian signal processing to classify biological markers within the mash against a known marker database. A simulation study is presented to assess the robustness of the model.

Les processus industriels nécessitent souvent des prises d'eau provenant des Grands Lacs à des fins de refroidissement. Malheureusement, des poissons et des larves de poisson de diverses espèces (la liste n'est pas entièrement connue) sont aspirés dans les systèmes industriels. Nous présentons une méthode d'analyse de la purée de biomasse (dont la composition est inconnue) provenant de réservoirs de prise d'eau. La méthode basée sur le traitement de signal bayésien permet de classifier les marqueurs biologiques dans la biomasse par rapport à une base de données de marqueurs connus. Nous présentons une étude de simulation pour analyser la robustesse de ce modèle.

3B-C2: Mixed Effects Models Modèles à effets mixtes

Chair/Président: Georges Monette (York University)

MacNaughton Building, room/local 113

Abstracts/Résumés

VAHID PARTOVI NIA (École Polytechnique de Montréal), **REZA DRIKVANDI** (Shahid Beheshti University, Iran), **GEERT VERBEKE** (Katholieke Universiteit Leuven, Belgium), **AHMAD KHODADADI** (Shahid Beheshti University, Iran)

[Wednesday June 6 / mercredi 6 juin, 10:15]

Testing Variance Components in Linear Mixed Effects Models

Le test de composantes de la variance dans les modèles linéaires avec effets mixtes

Testing variance components with zero is a well-known and non-standard hypothesis testing problem in linear mixed effects models with many applications. We introduce a test statistic using the variance least square estimator and propose to approximate its finite sample distribution using a permutation procedure. The procedure covers test of multiple variance components and any subset of them which is barely feasible using the existing methodologies. An application of the proposed method is demonstrated on real data.

Le test de composantes de la variance avec zéro est un problème bien connu et non-standard dans les modèles linéaires avec effets mixtes. Nous introduisons une statistique de test en utilisant l'estimateur des moindres carrés de la variance et proposons d'approximer sa loi en utilisant une procédure de permutation. La méthodologie permet de tester plusieurs composantes de la variance, de même que tous les sous-ensembles. La méthode est illustrée sur des données réelles.

YAN YAN WU (Samuel Lunenfeld Research Institute), **GEORGES MONETTE & AUGUSTINE WONG** (York University), **LAURENT BRIOLLAIS** (Samuel Lunenfeld Research Institute)

[Wednesday June 6 / mercredi 6 juin, 10:30]

Third-Order Method For the Test of Variance Components in Linear Mixed Model

Méthode de troisième ordre pour le test des composantes de la variance dans le modèle linéaire mixte

We consider the problem of testing for variance components in linear mixed models. The likelihood based on third-order asymptotic method is proposed for approximating the p-values for testing a scalar parameter, such as a random intercept or a random slope. The third-order p-values use a simple modification of the familiar first-order method, likelihood ratio test and Wald statistics. Extensive simulations show that the proposed approximations outperform the existing likelihood based methods with both maximum likelihood and restricted maximum likelihood.

Nous considérons le problème du test des composantes de la variance dans les modèles linéaires mixtes. La vraisemblance, basée sur une méthode asymptotique de troisième ordre, est proposée pour l'approximation des valeurs p pour tester un paramètre de type scalaire, comme une ordonnée à l'origine aléatoire ou une pente aléatoire. Les valeurs p de troisième ordre utilisent une simple modification de la méthode connue de premier ordre, le test de rapport de vraisemblances et la statistique de Wald. Des simulations intensives montrent que les approximations proposées surpassent les méthodes de vraisemblance existantes.

3B-C2: Mixed Effects Models Modèles à effets mixtes

ELEANOR M. PULLENAYEGUM & FENG XIE (McMaster University)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Methodology for Scoring the EQ-5D

Méthodologie de notation du EQ-5D

The EQ-5D yields a quality weight used to calculate quality adjusted life years. The new EQ-5D consists of five questions, each offering five possible responses, yielding 3125 health states. These health states must be converted into quality weights through a scoring algorithm. Methodology for creating the algorithm is lacking. The input data to the algorithm will consist of discrete choice experiments, with some time-trade-off tasks. This suggests using Generalised Linear Mixed Models to construct latent utilities, and regression to map onto utilities. The performance of these methods under varying inter-rater agreement and sample sizes is explored through simulation.

EQ-5D génère des coefficients de qualité utilisés pour calculer des années de vie pondérées par la qualité. Le nouveau EQ-5D inclut cinq questions, avec cinq réponses chacune, générant 3 125 états de santé. Ceux-ci doivent être convertis en coefficients de qualité via un algorithme de notation. Or, il n'existe aucune méthodologie pour créer ce dernier. Les intrants de l'algorithme sont des expériences de choix discret, avec quelques questions d'arbitrage temporel. On pourrait donc utiliser des modèles mixtes linéaires généralisés pour construire les utilités latentes et la régression pour effectuer la correspondance avec les utilités. Nous explorons par simulation la performance de ces méthodes pour plusieurs niveaux de concordance interévaluateurs et diverses tailles d'échantillon.

WEIQIANG WANG & ZENY FENG (University of Guelph)

[Wednesday June 6 / mercredi 6 juin, 11:00]

Modeling the Uncertainty of Susceptible and Non-susceptible Multiple Outcomes

Modélisation de l'incertitude de résultats multiples sensibles et non sensibles

Many diseases progress with multiple outcomes after the onset of the disease. With some potential factors, patients may or may not be susceptible to the event of interest. The objectives of the study are: (1) to identify and measure the effect of risk factors on the distribution of event time and the odds of susceptibility; (2) to model and predict the disease progression. We propose a mixed effect model nested within a mixture model to account for the susceptible and non-susceptible uncertainty and the correlations among the events. Simulation studies are conducted to assess the performance of our method.

Après leur apparition, plusieurs maladies progressent en présentant des résultats multiples. Selon certains facteurs potentiels, les patients peuvent être sensibles ou non à l'événement d'intérêt. Les objectifs de l'étude sont (1) d'identifier et mesurer les effets des facteurs de risque sur la distribution du moment de l'événement et les chances de sensibilité et (2) de modéliser et prévoir la progression de la maladie. Nous proposons un modèle à effets mixtes emboîté dans un modèle de mélange qui prend en considération les incertitudes sensible et non sensible, ainsi que les corrélations parmi les événements. Nous procédons à des études de simulation pour évaluer la performance de notre méthode.

HAOCHENG LI & GRACE Y. YI (University of Waterloo)

[Wednesday June 6 / mercredi 6 juin, 11:15]

A Variable Selection Method for Random Effects Models

Une méthode de sélection de variables pour les modèles à effets aléatoires

Random effects models are commonly used for the analysis of longitudinal data. Including irrelevant

Les modèles à effets aléatoires sont fréquemment utilisés pour l'analyse de données longitudinales. L'inclusion de co-

3B-C2: Mixed Effects Models Modèles à effets mixtes

covariates and random effects in the model may considerably degrade the quality of the results. To address this problem, we explore a variable selection method that can simultaneously select both fixed and random effects. Our method is developed based on penalizing a composite likelihood function. Empirical studies will be presented to assess the performance of the proposed method.

variables inutiles et des effets aléatoires dans le modèle peut diminuer la qualité des résultats de façon considérable. Pour résoudre ce problème, nous explorons une méthode de sélection de variables qui permet de sélectionner simultanément des effets fixes et aléatoires. Notre méthode est basée sur une pénalisation de la fonction de vraisemblance composite. Des études empiriques seront présentées pour analyser la performance de la méthode proposée.

JESSE D. RAFFA & JOEL A. DUBIN (University of Waterloo)

[Wednesday June 6 / mercredi 6 juin, 11:30]

Multivariate Longitudinal Data Analysis Using Hidden Markov Models

Analyse de données longitudinales multivariées par modèles de Markov cachés

Extending hidden Markov models (HMMs) to include random effects to describe subject-specific differences between HMM processes have allowed for effective modeling of longitudinal data in several disease areas. We propose further extensions to such models to accommodate multivariate longitudinal data of mixed data types. This approach was motivated by a smoking cessation clinical trial where subjects are monitored longitudinally using several distinct measures of smoking status. Under such models, the inclusion of hidden states describes heterogeneity due to changes in disease state (e.g. abstinence, smoking). The advantages and consequences of using such models will also be discussed.

L'extension des modèles de Markov cachés (MMC) pour inclure des effets aléatoires décrivant les différences spécifiques au sujet entre processus MMC a permis une modélisation efficace des données longitudinales pour diverses maladies. Nous proposons d'étendre ces modèles pour traiter de données longitudinales multivariées mixtes. Cette approche est motivée par un essai clinique sur l'abandon du tabac où les sujets sont surveillés de manière longitudinale à l'aide de différentes mesures de l'usage de tabac. Dans ces modèles, l'inclusion d'états cachés décrit l'hétérogénéité attribuable aux changements de pathologie (par exemple l'abstinence et le tabagisme). Nous discutons des avantages et des conséquences du recours à de tels modèles.

3B-C3: Models with Binary Responses Modèles à variables indépendantes binaires

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

Rozanski Hall, room/local 102

Abstracts/Résumés

ALAN KER (University of Guelph), **TOLGA ERGUN** (Suffolk University), **ABDOUL SAM** (Ohio State University)

[Wednesday June 6 / mercredi 6 juin, 10:15]

Semiparametric Estimation of the Link Function in Binary-Choice Single-Index Models

Estimation semi-paramétrique de la fonction lien dans les modèles de choix binaire à indice simple

In this manuscript we propose a new semiparametric estimator for binary-choice single-index models which uses parametric information in the form of a known link function and nonparametrically corrects it. Asymptotic properties are derived and the finite sample performance of the proposed estimator is compared to those of the parametric probit and semiparametric single-index model estimator of Klein and Spady.

Dans cet article, nous proposons un nouvel estimateur semi-paramétrique pour les modèles de choix binaire à un seul indice qui utilise l'information paramétrique sous forme d'une fonction lien connue et qui la corrige de façon non-paramétrique. Les propriétés asymptotiques de l'estimateur proposé sont développées et sa performance sur des échantillons de taille finie est comparée à celles du probit paramétrique et de l'estimateur semi-paramétrique du modèle à un seul indice de Klein et Spady.

FRANÇOIS LAROCHELLE & GOU TINGTING (Research and Development Canada, Department of National Defence)

[Wednesday June 6 / mercredi 6 juin, 10:30]

A Complementary Perspective on the Use of ROC Curve to Assess Logistic Regression Models

Une perspective complémentaire sur l'utilisation de la courbe ROC pour évaluer un modèle de régression logistique

The Receiver Operating Characteristics (ROC) curve is commonly used to assess the power of a logistic regression model for predicting a binary outcome. In this presentation, we consider questions such as “how do the estimated model parameters affect the Area under the ROC Curve (AUC)” and “what is the maximum AUC that can empirically be observed”. Based on the answers, we discuss the benefits of assessing the goodness-of-fit of a fitted model by comparing the model's AUC with the AUC for the saturated model. Examples from the analysis of non-response for the Canadian Forces “Your-Say” Survey are discussed.

La courbe Receiver Operating Characteristic (ROC) est couramment utilisée pour évaluer la puissance d'un modèle de régression logistique à prédire un résultat binaire. Lors de cette présentation, nous considérerons des questions telles que “comment les valeurs estimées des paramètres d'un modèle influencent-elles l'aire sous la courbe ROC (ASC)” et “quelle est l'ASC empirique maximale qui puisse être observée”. Nous considérerons aussi les avantages d'évaluer la qualité de l'ajustement d'un modèle en comparant son ASC avec celle du modèle saturé. Des exemples tirés de l'analyse de la non-réponse pour l'enquête des Forces canadiennes “À vous la parole” seront présentés.

3B-C3: Models with Binary Responses Modèles à variables indépendantes binaires

JINHUI MA, PARMINDER RAINA, JOSEPH BEYENE & LEHANA THABANE (McMaster University)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Comparison of Marginal and Cluster-Specific Models in Analysis of Cluster Randomized Trials with Missing Binary Outcomes

Comparaison de modèles marginaux et spécifiques aux grappes dans l'analyse d'essais randomisés avec des réponses binaires manquantes

Marginal and cluster-specific models are commonly used to analyze binary outcomes in cluster randomized trials to estimate the population-averaged and the cluster-specific treatment effect, respectively. However, limited attention has been paid to their performance when there are missing outcomes and some multiple imputation strategies are used to handle the missing data. Under the assumption that the probability of having missing outcomes is covariate dependent, we compare the performance of the generalized estimating equations and random-effects logistic models using a simulation study. The methods are compared in terms of standardized bias, coverage, root mean squared error and average standard error.

Les modèles marginaux et spécifiques aux grappes sont couramment utilisés pour analyser des réponses binaires dans des essais randomisés avec grappes pour estimer la moyenne de l'effet du traitement pour la population et dans chaque grappe. Toutefois, peu d'intérêt a été accordé à la performance dans le cas où des données de réponse sont manquantes et des stratégies d'imputation multiple sont utilisées. En supposant que la probabilité qu'une réponse soit manquante dépend des covariables, nous comparons la performance des équations d'estimation généralisées et des modèles logistiques avec effets aléatoires à l'aide d'une étude de simulation. Le biais standardisé, la probabilité de couverture, la racine carrée de l'erreur moyenne quadratique et l'erreur moyenne standardisée pour les deux méthodes sont comparés.

BINGRUI SUN & BRAJENDRA SUTRADHAR (Memorial University of Newfoundland)

[Wednesday June 6 / mercredi 6 juin, 11:00]

Correlations Versus Odds Ratios Based Inferences in Bivariate Binary Models

Inférences fondées sur les corrélations ou les rapports de cotes dans les modèles binaires bivariés

When bivariate binary responses are collected from a large number of independent individuals in a cross-sectional setup, it is standard to use a multinomial distribution with a cell probability modeled as a function of two marginal parameters and an association parameter. This approach makes the interpretation of marginal probabilities in terms of the model parameters difficult. In this talk, we model the marginal probabilities as functions of corresponding marginal parameters only and a correlation based conditional probability function is modeled to obtain the joint probabilities. Likelihood and quasi-likelihood estimation for inference about this model are discussed.

Lorsque des réponses binaires bivariées sont collectées auprès d'un grand nombre d'individus indépendants dans un dispositif transversal, l'approche standard consiste à utiliser une distribution multinomiale avec une probabilité de cellule modélisée comme fonction de deux paramètres marginaux et un paramètre d'association. Cela rend difficile toute interprétation des probabilités marginales en termes des paramètres du modèle. Nous modélisons les probabilités marginales comme fonctions uniquement des paramètres marginaux correspondants, puis modélisons une fonction de probabilité conditionnelle fondée sur la corrélation afin d'obtenir les probabilités communes. Nous discutons de l'estimation de la vraisemblance et de la quasi-vraisemblance pour l'inférence dans ce modèle.

3B-C3: Models with Binary Responses Modèles à variables indépendantes binaires

RONG CHU, STEPHEN D. WALTER, GORDON GUYATT, P.J. DEVEREAUX, MICHAEL WALSH, KRISTIAN THORLUND & LEHANA THABANE (McMaster University)

[Wednesday June 6 / mercredi 6 juin, 11:15]

Methodological Issues on Prognostic Imbalance in Randomized Controlled Trials (RCTs) Evaluating A Binary Outcome
Questions méthodologiques concernant le déséquilibre pronostique dans les essais contrôlés randomisés (ECR) évaluant un résultat binaire

Chance imbalance in baseline prognostic factors (PF) can lead to over or underestimation of treatment effects in RCTs. We simulated data from RCTs by varying risk of the outcome, effect of the treatment, power and prevalence of the PF, and sample size. Logistic regression models with and without adjustment for the PF were compared in terms of bias, standard error, coverage of confidence interval and statistical power. Covariate adjustment improves estimation accuracy and efficiency. The probability of prognostic imbalance in small trials can be substantial. Two thousand patients may be needed to minimize the chance and impact of prognostic imbalance.

Un déséquilibre fortuit des facteurs pronostiques (FP) de référence peut résulter en une surestimation ou sous-estimation des effets de traitement dans les essais contrôlés randomisés (ECR). Nous simulons des données d'ECR en variant le risque du résultat, l'effet du traitement, la puissance et la prévalence des FP, et la taille d'échantillon. Nous comparons les biais, l'erreur-type, la couverture de l'intervalle de confiance et l'efficacité statistique de modèles de régression logistique ajustés ou non pour les FP. L'ajustement de covariance améliore la précision et l'efficacité des estimations. Dans les petits essais, la probabilité de déséquilibre pronostique est élevée ; 2000 patients peuvent être nécessaire pour en minimiser le risque et l'impact.

RAFIQUL CHOWDHURY & YUN-HEE CHOI (University of Western Ontario)

[Wednesday June 6 / mercredi 6 juin, 11:30]

A Prediction Model to Identify Elderly Who Get Hospitalized – An Application Using Longitudinal Data

Modèle de prévision pour identifier les aînés à hospitaliser – une application fondée sur des données longitudinales

One of the most expensive services used by older persons is hospitalization or nursing home use in developed countries. It is a great challenge of the policy-makers for the cost-effective allocation of health care resources for aging populations. It is important to identify that elderly people living in the community are at risk for a variety of adverse outcomes which might lead to hospitalization. In this paper we presented a predictive model based on a regressive model approach to predict future hospitalizations among the elderly. The application is based on repeated measure data from Health and Retirement Study (HRS) in USA.

Dans les pays développés, l'un des services les plus coûteux qu'utilisent les aînés est l'hospitalisation ou l'accueil en maison de soins. Les décideurs ont bien du mal à allouer les ressources efficacement pour une population vieillissante. Il est important d'identifier quels aînés vivant dans la communauté sont exposés à divers risques de santé pouvant mener à l'hospitalisation. Nous présentons un modèle de prévision fondé sur une approche de modèle régressif pour prédire l'hospitalisation des aînés. L'application s'appuie sur des données de mesures répétées tirées du Health and Retirement Study (HRS) américain.

3B-C4: Spectral Methods in Time series Analyse spectrale de séries chronologiques

Chair/Président: Rafal Kulik (University of Ottawa)

Rozanski Hall, room/local 105

Abstracts/Résumés

CHARLOTTE L. HALEY & DAVID J. THOMSON (Queen's University)

[Wednesday June 6 / mercredi 6 juin, 10:15]

Jackknifed Error for Multitaper Estimates of Autocovariance in Time Series

Estimation de l'erreur par la méthode du jackknife pour les estimateurs à fenêtres de pondération multiples de l'autocovariance dans des séries chronologiques

In this study, we aim to quantify bias and variance in estimates of the autocovariance function obtained from a multitaper spectrum estimate. Because the multitaper spectrum is formed using a combination of spectrum estimates obtained by tapering with a set of orthogonal optimally-bandlimited tapers, individual spectrum estimates are reasonably independent. In simulated data from a simple AR(2) process, a jackknifed estimate of the bias in the multitaper estimate was computed. The multitaper covariances had less bias than the Bartlett (conventional) estimate, especially for large lags. Jackknifed variance estimates, with the distribution of the Pearson correlation coefficient, give approximate confidence intervals.

L'objectif de cette étude consiste à quantifier le biais et la variance des estimateurs de la fonction d'autocovariance obtenue à partir d'une estimation spectrale à fenêtres de pondération multiples. Les estimations spectrales individuelles sont raisonnablement indépendantes car le spectre à fenêtres de pondération multiples provient d'une combinaison d'estimations spectrales obtenue par la pondération d'une série de fenêtres de pondération orthogonales à largeur de bande limitée. Dans les données simulées obtenues à partir d'un processus simple AR(2), une estimation par jackknife du biais de l'estimateur basé sur les fenêtres de pondération multiples a été calculée. Les covariances issues des fenêtres de pondération sont moins biaisées que l'estimation (conventionnelle) de Bartlett, particulièrement en ce qui concerne les grands décalages. Les estimations de variance par la méthode du jackknife, ainsi que la distribution du coefficient de corrélation de Pearson donnent des intervalles de confiance approximatifs.

AARON SPRINGFORD (Queen's University)

[Wednesday June 6 / mercredi 6 juin, 10:30]

Identification of Time Series Outliers Applied to Schumann Resonance Data

Identification des valeurs aberrantes dans des séries chronologiques avec application aux données sur la résonance de Schumann

One application of time series analysis is the identification of outliers. Schumann resonances (SR) are extremely low frequency electromagnetic resonances that propagate between the Earth's surface and ionosphere. The geometry of the ionosphere is affected by the solar wind, and in particular may be sensitive to X-ray flare events. Thus, X-ray flares should be correlated with outliers in

L'une des applications de l'analyse des séries chronologiques est l'identification de valeurs aberrantes. Les résonances de Schumann (RS) sont des résonances électromagnétiques de très faible fréquence qui se propagent entre la surface de la Terre et l'ionosphère. Or, la géométrie de l'ionosphère est affectée par les vents solaires et est notamment sensible aux éruptions de rayons X. Ces dernières devraient donc être corrélées avec des valeurs aberrantes

3B-C4: Spectral Methods in Time series Analyse spectrale de séries chronologiques

SR frequency. I examined two time series of SR frequency in order to identify outliers. There appears to be no relationship between SR frequency outliers and powerful X-type flares, suggesting that X-ray flares affect ionosphere geometry in a more subtle way.

dans la fréquence de RS. J'ai étudié deux séries chronologiques de fréquences de RS pour identifier des valeurs aberrantes. Aucune relation ne semble exister entre ces valeurs aberrantes et les éruptions solaires de type X, ce qui suggère que ces éruptions affectent la géométrie de l'ionosphère plus subtilement.

DAVID J. THOMSON (Queen's University)

[Wednesday June 6 / mercredi 6 juin, 10:45]

Incautious Use of Robust Spectrum Estimation Procedures

Utilisation imprudente des procédures robustes d'estimation du spectre

Studies of the low-frequency seismic noise spectrum are difficult because, for noise studies, earthquakes are outliers and there are about 100 magnitude 6 earthquakes per year. A common approach is to divide the data into many short (about one day) data segments, compute a spectrum on each, and take the median spectrum at each frequency. This effectively eliminates the effects of earthquakes but is biased and inefficient. Worse, in some cases it also has "thrown out the baby with the bath water" and missed important effects. We examine this procedure using exploratory data analysis and suggest some improvements.

L'étude du spectre de basse fréquence du bruit sismique est difficile car les séismes sont des valeurs aberrantes dans le cadre d'études du bruit et il se produit environ 100 séismes de magnitude 6 par année. L'approche courante consiste à séparer les données en plusieurs segments courts (d'une journée environ), à calculer un spectre pour chacun, et d'évaluer le spectre médian de chaque fréquence. Bien qu'elle élimine effectivement les effets des séismes, cette méthode est biaisée et inefficace. Pire encore, elle a dans certains cas "jeté le bébé avec l'eau du bain" en passant à côté d'effets importants.

DAVID RIEGERT, AARON SPRINGFORD, WESLEY BURR & DAVID J. THOMSON (Queen's University)

[Wednesday June 6 / mercredi 6 juin, 11:00]

Is There Structure in High-Frequency Variation of Great Lakes Water Levels?

Y a-t-il une structure dans la variation des hautes fréquences du niveau de l'eau dans les Grands Lacs?

Many authors have suggested that time series of water level in the Great Lakes basin are well-described by a power-law in the frequency domain. This suggestion implies that high-frequency variation in water level is negligible. One expects high frequency structure to be present in water level data because it is seen regularly in other climate data. However, previous authors have employed less than state-of-the-art spectrum estimation techniques. We re-examine several time series of water level in the Great Lakes basin in order to determine if in fact there is appreciable structure present at high frequencies as expected.

De nombreux auteurs suggèrent que la série chronologique du niveau de l'eau des Grands Lacs peut se décrire par une loi de puissance dans le domaine fréquentiel. Cela implique que les variations des hautes fréquences sont négligeables. On s'attend toutefois à de hautes fréquences dans de telles données hydrologiques car cela se retrouve fréquemment dans les données climatiques. Certains auteurs ont employé des techniques d'estimation par spectre quelque peu désuètes. Nous réexaminons plusieurs séries chronologiques du niveau de l'eau dans ce bassin hydrologique afin de déterminer si une telle structure existe réellement dans les hautes fréquences.

3C-A: Survey Methods Section Presidential Invited Address: Empirical Likelihood in Survey Sampling
Allocution de l'invité du Président du Groupe des méthodes d'enquêtes: Vraisemblance empirique et échantillonnage

Organizer and Chair / Responsable et président: David Haziza (Université de Montréal)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Rozanski Hall, room/local 104

Abstract/Résumé

CHANGBAO WU (University of Waterloo)

[Wednesday June 6 / mercredi 6 juin, 13:15]

Empirical Likelihood in Survey Sampling

La vraisemblance empirique dans les enquêtes par sondage

We provide an overview of the recent developments of empirical likelihood (EL) methods in survey sampling. We start with the basic formulation of EL functions for non-stratified and stratified sampling designs, and then discuss a few extensions, including EL for multiple frame surveys and EL-based Bayesian approach for complex survey data. Computational algorithms will be described, and some related topics, such as EL-based calibration and ranking methods, bootstrap procedures, EL and estimating equations, EL-based model selection using survey data, and EL for small area estimation, will also be briefly discussed.

Nous présentons un aperçu des récents développements des méthodes basées sur la vraisemblance empirique (VE) dans les enquêtes par sondage. Nous commençons par une formulation simple des fonctions de VE pour des plans d'échantillonnage non-stratifiés et stratifiés, puis nous discutons de quelques extensions, y compris la VE pour les enquêtes à bases multiples et l'approche bayésienne basée sur la VE pour les données d'enquêtes complexes. Nous décrivons aussi des algorithmes de calcul et nous abordons brièvement quelques autres sujets connexes, notamment les méthodes de calibration et de classement, les procédures bootstrap, les équations de VE et d'estimation, la sélection de modèle basée sur la VE à l'aide de données d'enquêtes, ainsi que la VE pour petits domaines.

3C-I1: Causality and Graphical Markov Models in High Dimensions Causalité et modèles graphiques markoviens en plusieurs dimensions

Organizer and Chair / Responsable et présidente: Ayesha Ali (University of Guelph)

Rozanski Hall, room/local 101

Abstracts/Résumés

FREDERICK EBERHARDT (Washington University, St. Louis), **ANTTI HYTTINEN & PATRIK HOYER** (University of Helsinki, Finland)

[Wednesday June 6 / mercredi 6 juin, 13:15]

Causal Discovery from Experiments

Découverte causale à partir d'expériences

Randomized controlled trials have in many ways become the gold standard for causal discovery. However, when a large number of variables are involved, as for example when identifying a gene regulatory network, a single experiment will generally not be sufficient to identify the full causal structure. In this presentation I will discuss optimal strategies for the selection of experiments given a variety of background assumptions about the model space. The strategies rely on necessary and sufficient conditions that characterize a set of experiments adequate for causal discovery.

Les essais randomisés sont devenus la norme pour les découvertes causales. Cependant, lorsqu'un grand nombre de variables entre en jeu, lors de l'identification d'un réseau de gènes régulateurs par exemple, une seule expérience ne sera généralement pas suffisante pour identifier toute la structure causale. Dans cette présentation, nous allons discuter des stratégies optimales de sélection d'expériences en fonction de nombreuses hypothèses de départ sur l'espace de modèles. Ces stratégies reposent sur des conditions nécessaires et suffisantes qui caractérisent un ensemble d'expériences adéquates pour la découverte causale.

HELENE MASSAM (York University), **GERARD LETAC** (Université Paul Sabatier, France)

[Wednesday June 6 / mercredi 6 juin, 13:45]

Bayes Factors and the Geometry of Discrete Loglinear Models

Facteurs de Bayes et géométrie du modèle discret log-linéaire

We consider the class of hierarchical loglinear models for discrete data given under the form of a contingency table with multinomial sampling. When the prior distribution on the loglinear parameters is conjugate, the Bayes factor between two models is a function of the hyperparameters (m, α) of the conjugate prior. When the data is on the boundary of C and $\alpha \rightarrow 0+$, we need to study the asymptotic behaviour of the Bayes factor, which, as we will see, is dictated by the dimension of the face of C to which the data belongs.

Nous considérons la classe des modèles hiérarchiques log-linéaires pour un tableau de contingence sous échantillonnage multinomial. Lorsque la distribution a priori des paramètres log-linéaires est conjuguée, le facteur de Bayes pour la comparaison des deux modèles est une fonction des hyperparamètres (m, α) de cette distribution. Si les données se retrouvent sur la frontière de C , nous devons étudier le comportement asymptotique du facteur de Bayes lorsque $\alpha \rightarrow 0+$. Nous verrons que celui-ci dépend de la dimension de la face de C à laquelle les données appartiennent.

ERICA MOODIE (McGill University), **BIBHAS CHAKRABORTY** (Columbia University)

[Wednesday June 6 / mercredi 6 juin, 14:15]

Q-learning for Estimating Optimal Dynamic Treatment Rules from Observational Data

Q-learning pour l'estimation des règles optimales de traitement dynamiques de données d'observation

3C-I1: Causality and Graphical Markov Models in High Dimensions Causalité et modèles graphiques markoviens en plusieurs dimensions

Q-learning, a popular reinforcement learning approach, has recently been applied to estimate dynamic treatment regimes. While, in principle, Q-learning can be used for both randomized and observational data, literature thus far has focused exclusively on the randomized treatment setting. We extend an existing implementation of Q-learning to incorporate confounding covariates. We provide results of an extensive simulation study to compare different adjustment methods and show how some forms of model mis-specification may be diagnosed. Methods are illustrated using the PROBIT data to study the effect of breastfeeding on IQ.

Q-learning, une approche d'apprentissage de renforcement populaire, a récemment été utilisée pour estimer les régimes de traitement dynamiques. Alors que, en principe, l'approche Q-learning peut être utilisée pour des données randomisées ou d'observation, la littérature à ce jour a porté exclusivement sur les traitements randomisés. Je discuterai d'une extension de l'approche existante de Q-learning en intégrant des covariables confondantes. Je fournirai ensuite quelques résultats d'une vaste étude de simulation visant à comparer les différentes méthodes d'ajustement, et montrerai comment certaines formes de mauvaise spécification du modèle peuvent être diagnostiquées. Les méthodes seront illustrées à l'aide des données de PROBIT pour étudier l'effet de l'allaitement maternel sur le QI.

3C-I2: Copula Models Modèles de copules

Chair/Président: Jean-François Plante (HEC Montréal)
Organizer/Responsable: Johanna Nešlehová (McGill University)

Alexander Hall, room/local 200

Abstracts/Résumés

LOUIS-PAUL RIVEST (Université Laval)

[Wednesday June 6 / mercredi 6 juin, 13:15]

Applications of Exchangeable Copula Models

Quelques applications des modèles de copule échangeables

Exchangeable multivariate copulas, especially those in the Archimedean family, are used to model a within cluster correlation. New copula based estimators of the intra-cluster correlation for both discrete and continuous data are discussed. Copulas provide alternatives to the inclusion of a random cluster effect in a model to account for a within cluster dependency; some properties of copula based predictions of the cluster mean value will be given. This will allow the construction of unit level small area predictions using copulas.

On utilise des copules multidimensionnelles échangeables pour modéliser la corrélation intra-grappe. De nouveaux estimateurs, basés sur les copules, pour le coefficient de corrélation intra-grappe de données discrètes et continues sont construits. L'utilisation des copules constitue une alternative à l'inclusion d'un effet grappe aléatoire pour représenter une dépendance intra-grappe. Quelques propriétés de prédictions de la moyenne d'une grappe construites à l'aide de copules sont données. Ceci permet de construire des nouveaux prédicteurs, basés sur les copules, pour des caractéristiques de petits domaines.

JOHANNA NEŠLEHOVÁ & CHRISTIAN GENEST (McGill University), **BRUNO RÉMILLARD** (HEC Montréal)

[Wednesday June 6 / mercredi 6 juin, 13:45]

On the Multi-linear Empirical Copula Process

À propos du processus de copule empirique multilinéaire

A critical review of copula-based modelling approaches for discrete variables will be presented. Conceived for the analysis of continuous data, copula inference techniques are based on relations that may no longer hold when the marginal distributions have atoms. An indiscriminating transposition of these techniques for the treatment of discrete data may thus affect the analyses and lead to misleading or invalid conclusions. The multi-linear empirical copula provides a promising alternative avenue. The limiting behaviour of the process associated with this copula will be described along with its use for the construction of new copula inference tools for discrete data.

On portera un regard critique sur les pratiques de modélisation de variables discrètes à l'aide de copules. Conçues pour l'analyse de données continues, les techniques d'inférence de copules s'appuient sur des faits parfois non avérés lorsque les marges ont des atomes. Une transposition aveugle de ces techniques au traitement de données discrètes peut donc fausser les analyses et conduire à des conclusions erronées ou trompeuses. La copule empirique multilinéaire ouvre toutefois des perspectives prometteuses. On décrira le comportement limite du processus associé à cette copule et son emploi pour l'élaboration de nouveaux outils d'inférence de copules adaptés aux données discrètes.

3C-I2: Copula Models Modèles de copules

BRUNO RÉMILLARD (HEC Montréal)

[Wednesday June 6 / mercredi 6 juin, 14:15]

Testing Hypotheses for the Copula of Dynamic Models

Tests d'hypothèses pour la copule de modèles dynamiques

The asymptotic behaviour of the empirical copula constructed from residuals of stochastic volatility models is studied. It is shown that if the stochastic volatility matrix is diagonal, then the empirical copula process behaves like if the parameters were known, a remarkable property. However, this is not true in general. Applications for goodness-of-fit and detection of structural change in the copula of the innovations are discussed.

On étudie le comportement asymptotique de la copula empirique construite à partir des résidus de modèles à volatilité stochastique. On montre que si la matrice de volatilité est diagonale, alors le processus de copule empirique a le même comportement que si les paramètres étaient connus, ce qui est remarquable. Cependant, ce n'est pas vrai dans le cas général. On présente des applications pour la détection de changements structuraux et pour des tests d'adéquations pour la copule des innovations.

3C-I3: Design and Analysis of Experiments in Modern Agriculture Conception et analyse d'expériences en agriculture moderne

Organizer and Chair / Responsable et président: Tom Loughin (Simon Fraser University)

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Rozanski Hall, room/local 103

Abstracts/Résumés

CHRIS BRIEN (University of South Australia), **BRONWYN D. HARCH & RAY L. CORRELL** (CSIRO Mathematical and Information Sciences, Australia), **ROSEMARY A. BAILEY** (Queen Mary University of London, U.K.)

[Wednesday June 6 / mercredi 6 juin, 13:15]

Principles in the Design of Multiphase Experiments with a Later Laboratory Phase: Orthogonal Designs

Principes généraux dans la conception de plans d'expérience à plusieurs phases comprenant une phase ultérieure en laboratoire : plans orthogonaux

It is common for the material produced from field and other experiments to be processed in a laboratory. Such experiments are multiphase, although this has often not been realized. We provide general principles for designing multiphase experiments that employ orthogonal designs. In particular, we discuss the randomization in the laboratory phase of material produced from the first phase, taking into account the first-phase randomization. Factor-allocation diagrams are employed to depict the randomizations in a design and skeleton ANOVA tables to evaluate their properties. The techniques are illustrated using several scenarios for an agricultural plant experiment.

Il est courant que du matériel obtenu sur le terrain et via d'autres expériences soit traité en laboratoire. De telles expériences comportent plusieurs phases, bien que ce fait soit souvent oublié. Nous proposons des principes généraux pour la conception d'expériences à plusieurs phases utilisant des plans orthogonaux. En particulier, nous discutons de la randomisation, pendant la phase en laboratoire, du matériel recueilli au cours de la première phase en tenant compte de la randomisation de celle-ci. Des diagrammes de répartition factoriels sont utilisés pour illustrer les randomisations dans un plan et le squelette des tableaux d'ANOVA est utilisé pour évaluer leurs propriétés. Les techniques sont illustrées à l'aide de plusieurs scénarios d'une expérience agricole portant sur des plantes.

GEORGE MILLIKEN (Kansas State University)

[Wednesday June 6 / mercredi 6 juin, 13:45]

Basic Designs of Experiments for Comparing Precision Agriculture Treatments on Research and Commercial Fields
Plans d'expérience simples pour comparer des traitements agricoles de précision dans des champs de recherche et commerciaux

Designing experiments to evaluate precision agricultural treatments on research and commercial fields is a new frontier in the world of design of experiments. Most precision agriculture studies involve defining management zones, determining an appropriate prescription and then applying the prescribed rates of the treatments to the management zones. This process does not allow one to answer

La conception d'expériences pour évaluer les traitements agricoles de précision dans les parcelles de recherche et les champs commerciaux constitue une nouvelle avenue à explorer en matière de planification d'expériences. La plupart des études sur l'agriculture de précision nécessitent la définition de zones de gestion, la détermination de la prescription appropriée, puis l'application des taux de traitements prescrits aux zones de gestion. Ce processus ne permet

3C-I3: Design and Analysis of Experiments in Modern Agriculture Conception et analyse d'expériences en agriculture moderne

the question “Are the prescribed rates the appropriate rates for each management zone?” A strategy is presented that enables one to determine if the prescribed rates are appropriate.

pas de répondre à la question suivante : « Les taux prescrits sont-ils appropriés à chaque zone de gestion ? » Nous présentons une stratégie permettant de déterminer si les taux prescrits sont appropriés.

3C-14: Robustness for Complex Data Robustesse et données complexes

Chair/Président: Sanjoy Sinha (Carleton University)
Organizer/Responsable: Joanna Mills Flemming (Dalhousie University)

Alexander Hall, room/local 100

Abstracts/Résumés

CHRIS FIELD (Dalhousie University)

[Wednesday June 6 / mercredi 6 juin, 13:15]

Robustness Issues in Molecular Evolution

Questions de robustesse en évolution moléculaire

In this talk, I will outline my perspectives on robust statistics indicating some of the key ideas. The focus will be on developing robust statistical procedures for modelling biological processes with particular interest in molecular evolution and estimation of the phylogeny. In this setting, the standard models are approximations to reality and the data itself can be quite noisy and subject to outliers. Both factors indicate a strong need for procedures which work well for neighbourhoods of the putative model and are not unduly influenced by outlying observations, namely robust procedures.

Dans cet exposé, je vais décrire mes perspectives sur la statistique robuste en identifiant certaines idées clés. Je vais me concentrer sur la conception de procédures statistiques robustes pour la modélisation de processus biologiques avec un intérêt particulier sur l'évolution moléculaire et l'estimation de la phylogénie. Dans ce contexte, les modèles standards sont des approximations de la réalité et les données mêmes peuvent être assez bruitées et sujettes à des valeurs aberrantes. Les deux facteurs indiquent un fort besoin pour des procédures robustes qui fonctionnent bien pour des voisinages du modèle supposé et qui ne sont pas influencées excessivement par des observations aberrantes.

MATÍAS SALIBIÁN-BARRERA (University of British Columbia), **GRACIELA BOENTE** (Universidad de Buenos Aires, Argentina)

[Wednesday June 6 / mercredi 6 juin, 13:45]

On Robust Functional Principal Components

Composantes principales robustes pour données fonctionnelles

In this talk we study robust estimates of principal components for functional data. This work was motivated by a problem of outlier detection in atmospheric data collected by weather balloons launched into the atmosphere and stratosphere. Since principal components give best (in the L_2 sense) finite-dimensional approximation to random elements over separable Hilbert spaces, we consider the problem of robustly estimating these finite-dimensional approximating subspaces. We obtain some general optimality results and discuss a new family of estimators that we call S-estimators and compare their properties with those of previ-

Dans cette présentation, nous étudions des estimations robustes des composantes principales pour des données fonctionnelles. Ce travail a été motivé par un problème de détection de valeurs aberrantes dans des données atmosphériques recueillies par des ballons météorologiques lancés dans l'atmosphère et la stratosphère. Puisque les composantes principales fournissent la meilleure approximation (au sens L_2) de dimension finie d'éléments aléatoires dans les espaces de Hilbert séparables, nous considérons des estimations robustes pour ces sous-espaces. Nous introduisons une nouvelle famille d'estimateurs que nous appelons S-estimateurs. Nous comparons leurs propriétés à celles des estimateurs proposés antérieurement qui peuvent être

3C-14: Robustness for Complex Data Robustesse et données complexes

ously proposed estimators that can be used in this setting. utilisés dans ce contexte.

3C-C1: Financial Applications and Heavy-Tailed Distributions Applications en finance et distributions à ailes relevées

Chair/Président: François Watier (Université du Québec à Montréal)

Rozanski Hall, room/local 102

Abstracts/Résumés

VINCENZO COIA & MEI LING HUANG (Brock University)

[Wednesday June 6 / mercredi 6 juin, 13:15]

On Estimation of Heavy Tailed Distributions

De l'estimation des distributions à ailes relevées

The heavy tailed distributions have many important applications in many fields. There are theoretical difficulties in studying the heavy tailed distributions since they often have infinite moments. There also are bias issues with the existing estimation methods. This paper explores new estimation methods to avoid these difficulties and to improve the efficiencies and reduce bias. Comparisons of efficiencies and biases of the new estimators relative to existing estimators are studied. The results of simulations confirm the theoretical results. We also investigate a real-world example of applications of heavy tailed distributions. Goodness-of-Fit tests are performed for this example.

Les distributions à ailes relevées connaissent des applications importantes dans divers domaines. Cependant, leur étude présente des difficultés théoriques car elles incluent souvent des moments infinis et les méthodes d'estimation existantes sont biaisées. Nous explorons de nouvelles méthodes d'estimation qui évitent ces difficultés, améliorent l'efficacité et réduisent le biais. Nous comparons les efficacités et les biais des nouveaux estimateurs à ceux des estimateurs existants. Les résultats des simulations confirment les résultats théoriques. Nous étudions enfin un exemple réel d'application des distributions à ailes relevées et effectuons des tests d'adéquation sur celui-ci.

DAVID EDWARD WILSON (University of Waterloo)

[Wednesday June 6 / mercredi 6 juin, 13:30]

Estimating the Meixner Process

Estimation du processus de Meixner

Brownian motion is a poor model for log asset prices in part because the tails are too light to fit empirical returns. Merton jump-diffusion is a popular alternative where the log price includes compound Poisson jumps. Unfortunately, it is difficult to accurately estimate the jump component. Among many other choices the Meixner process stands out. It has no Gaussian component and consists purely of random jumps with a deterministic drift. The log returns follow a Meixner distribution which has four parameters and orthogonal support. The tails are semi-heavy and the model is shown to be well estimated by maximum likelihood.

Le mouvement brownien constitue un modèle inadéquat pour le logarithme des prix des actifs notamment parce que les queues sont trop légères pour s'ajuster aux rendements empiriques. Le processus de diffusion par saut de Merton est une solution de rechange populaire dans laquelle le logarithme des prix est sujet à des sauts de Poisson composés. Malheureusement, il est difficile d'estimer précisément la composante des sauts. Le processus de Meixner se démarque des nombreuses autres options. Il ne possède aucune composante gaussienne et est composé exclusivement de sauts aléatoires avec dérive déterministe. Le logarithme des rendements suit une distribution de Meixner à quatre paramètres définie sur un support orthogonal. Les queues sont semi-lourdes et le modèle est bien estimé par la méthode du maximum de vraisemblance.

3C-C1: Financial Applications and Heavy-Tailed Distributions Applications en finance et distributions à ailes relevées

ZHAOXIA (JEANA) REN & ADAM KOLKIEWICZ (University of Waterloo)

[Wednesday June 6 / mercredi 6 juin, 13:45]

Estimation of the Jump Frequency in Jump-Diffusion Models

Estimation de la fréquence de chocs dans les modèles de distribution de chocs

We consider the problem of estimation of jump frequency in the context of general jump-diffusion models used in Finance. The method uses the characteristic function and is semi-parametric since it does not depend on any specific form of the distribution of jump size. We investigate properties of the estimator and also illustrate the method with some examples.

Nous étudions le problème de l'estimation de la fréquence des chocs dans le contexte des modèles de distribution des chocs généraux utilisés en finance. La méthode utilise la fonction caractéristique. La méthodologie est semi-paramétrique car elle ne dépend d'aucune forme spécifique quand à la distribution de la taille des chocs. Nous explorons les propriétés de l'estimateur et illustrons la méthode à l'aide de plusieurs exemples.

NADER NEMATOLLAHI (Allameh Tabataba'i University, Iran and University of Manitoba / Allameh Tabataba'i University, Iran et University of Manitoba), **AZADEH KIAPOUR** (Islamic Azad University, Iran)

[Wednesday June 6 / mercredi 6 juin, 14:00]

Estimation of Premium and Prediction of Claim Size under Scale Invariant Squared Error Loss Function

Estimation des primes et prédiction de la taille des réclamations sous une fonction de perte quadratique invariante sous un changement d'échelle

Calculation of premium in a Bayes setup analysis requires complete specification of the prior and it is difficult to justify any particular choice. In robust Bayesian methodology, uncertainty in the prior can be modeled by specifying a class Γ of priors instead of a single prior. In this article, we study the robust Bayesian premium estimation and prediction under the scale invariant squared error loss function. We derive the conditional Γ -minimax, most stable, posterior regret Γ -minimax and least sensitive premium estimation and prediction in a gamma model. Theoretical results are augmented by a simulation study.

Le calcul des primes dans un cadre bayésien requiert une spécification complète de la distribution a priori pour laquelle il est difficile de justifier un choix spécifique. Dans la méthodologie bayésienne robuste, l'incertitude sur la distribution a priori peut être modélisée en spécifiant une classe de distributions a priori Γ , au lieu d'une seule. Dans cet article, nous étudions l'estimation des primes dans un modèle bayésien robuste et la prédiction sous une fonction de perte quadratique invariante sous un changement d'échelle. Nous dérivons le minimax conditionnel Γ , très stable, le Γ -minimax du regret a posteriori et l'estimateur et le prédictor des réclamations le moins sensible dans un modèle gamma. Une étude de simulations vient étayer les résultats théoriques.

FRANCK ADEKAMBI (University of the Witwatersrand, South Africa)

[Wednesday June 6 / mercredi 6 juin, 14:15]

Ruin Probability in Health Care Insurance Pricing with Constant Force of Interest

Probabilité de ruine dans la fixation des prix d'assurance avec intensité d'intérêt constante

Ramsay (1984) in paper published in *Insurance: Mathematics and Economics*, entitled "The asymptotic ruin problem when the healthy and sick periods form an alternating renewal process" found expressions for the probabilities of sickness and health, and an approximation of the probability of

Dans le journal *Insurance : Mathematics and Economics*, Ramsay (1984) a publié un article intitulé « le problème de ruine asymptotique quand des périodes de bonne santé et de maladie forment un processus de renouvellement alternant » dans lequel il donne des expressions de probabilité de maladie et de santé, ainsi qu'une approximation de la

3C-C1: Financial Applications and Heavy-Tailed Distributions **Applications en finance et distributions à ailes relevées**

ruin but ignoring the effect of force of interest. This paper aims to consider the Ramsay model modified by the inclusion on interest rate. Upper bounds for the ultimate ruin probability are derived by martingale and recursive techniques.

probabilité de ruine. Cependant, il ignore l'effet de l'intensité d'intérêt. Cet article vise à considérer le modèle de Ramsay modifié par l'inclusion d'un taux d'intérêt. Les limites supérieures de probabilité de ruine finale sont obtenues à partir des techniques récursives et de martingales.

3C-C2: Functional Data, Principal Components and Clustering

Données fonctionnelles, composantes principales et méthodes de partitionnement

Chair/Président: Giles Hooker (Cornell University)

MacNaughton Building, room/local 113

Abstracts/Résumés

JAMES RAMSAY (McGill University)

[Wednesday June 6 / mercredi 6 juin, 13:15]

Multivariate and Functional Principal Components without Eigenanalysis

Composantes principales multivariées et fonctionnelles sans analyse de valeurs propres

Principal components analysis is an invaluable tool in multivariate or functional data analysis, but it suffers from several defects. The new approach is based on a parameter cascade that defines factor scores as smooth functions of principal component vectors/functions. In the functional case, registration of time warping parameters are in turn defined as functions of factor scores and hence of principal components. In this approach, any suitable loss function may be employed for any variable in the multivariate case, or vary over time in the functional case. The method is also adaptable to non-flat manifold estimation.

L'analyse en composantes principales est un outil précieux dans l'analyse des données multivariées ou fonctionnelles, mais elle souffre de plusieurs défauts. Cette nouvelle approche se fonde sur une cascade de paramètres qui définit les poids factoriels comme des fonctions lisses de vecteurs/fonctions des composantes principales. Dans le cas fonctionnel, l'enregistrement des paramètres d'adaptation temporelle se définit comme fonctions de poids factoriels, et donc de composantes principales. Dans cette approche, toute fonction de perte appropriée peut être utilisée pour toute variable dans le cas multivarié, ou varier dans le temps dans le cas fonctionnel. La méthode s'adapte aussi à l'estimation des variétés non planes.

FOLLY ADJOGOU & ALEJANDRO MURUA (Université de Montréal)

[Wednesday June 6 / mercredi 6 juin, 13:30]

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 and model-based clustering. The functional modelling is based on splines. The main data groups are modelled as arising from clusters in the space of spline coefficients (the factors). The clusters are modelled by a mixture of Student's t -distributions whose degrees of freedom are unknown. The model is embedded into a Bayesian framework. The MAP estimators are found with the EM algorithm. Our main applications are to gene expression and clinical data.

Nous développons un modèle flexible pour l'analyse et la classification de données longitudinales balancées ou non balancées. Le modèle combine l'analyse fonctionnelle en composantes principales et le regroupement en classes. La modélisation fonctionnelle est basée sur les splines et les groupes principaux sont issus de l'espace des coefficients dans la base des splines. Le regroupement en classes repose sur un modèle de mélange de distributions t de Student de degrés de liberté inconnus. Nous utilisons une approche bayésienne où les estimateurs sont obtenus par l'algorithme EM. Nos applications principales concernent les données d'expression génétique et les essais cliniques.

3C-C2: Functional Data, Principal Components and Clustering

Données fonctionnelles, composantes principales et méthodes de partitionnement

YUNQI JI & ANTONIO CIAMPI (McGill University)

[Wednesday June 6 / mercredi 6 juin, 13:45]

Model Based Clustering of Longitudinal Data: Application to Modeling Disease Course

Classification basée sur un modèle pour données longitudinales : application à la modélisation de l'évolution d'une maladie

In epidemiological studies, the identification of disease trajectories over time is of scientific interest. We consider the problem of clustering for time dependent data. The model is a mixture of regressions, within variance-covariance matrices that are allowed to vary within the extended linear mixed model family. The associated maximum likelihood estimates for the parameters are obtained using expectation-maximization (EM) algorithms. We also develop two accelerated EM algorithms and further conduct parallel computations for the algorithms. We discuss the application to biomedical data set and analyze the longitudinal delirium data.

Dans les études épidémiologiques, on s'intéresse à l'identification de la trajectoire des maladies. Nous considérons le problème de la classification pour données dépendantes du temps. Notre modèle est un mélange de régressions avec des matrices de variance-covariance libres de varier à l'intérieur de la famille étendue de modèles linéaires mixtes. Les estimateurs à maximum de vraisemblance associés sont obtenus à l'aide d'algorithmes de maximisation de l'espérance (algorithmes EM). Nous développons aussi deux algorithmes EM accélérés et parallélisons les calculs des algorithmes. Nous discutons de l'application de cette méthode à un jeu de données biomédicales et analysons des données longitudinales sur le délirium.

SANJEENA DANG & PAUL MCNICHOLAS (University of Guelph)

[Wednesday June 6 / mercredi 6 juin, 14:00]

Variational Approximations for Flexible Mixture Model-Based Clustering

Approximations variationnelles pour classification par mélange de modèles

Model-based clustering approaches are reviewed, with particular emphasis on an approach that utilizes a Gaussian mixture model with eigen-decomposition of the covariance structure. A Bayesian framework for parameter estimation is introduced as an alternative to the expectation-maximization framework, and we deviate from the traditional Bayesian information criterion approach to model selection. More specifically, variational approximations are used for the estimation of the parameters as well as the number of components and deviance information criterion is used for model selection. Our approach is illustrated on simulated data sets as well as an Italian wine data set.

Les méthodes de classification basées sur un modèle sont passées en revue, en mettant l'accent sur une approche qui utilise un mélange de modèles gaussiens avec une décomposition de la covariance en valeurs propres. Une approche bayésienne est suggérée comme alternative à l'approche de maximisation de l'espérance, et nous dévions de l'approche bayésienne traditionnelle de sélection de modèle. Plus précisément, des approximations variationnelles sont utilisées pour l'estimation des paramètres et du nombre de composantes, et le critère d'information de déviance est utilisé pour la sélection d'un modèle. Nous illustrons notre approche sur des données simulées ainsi que sur un jeu de données concernant les vins italiens.

3C-C3: Methods for Genetic Association Méthodes d'association génétique

Chair/Président: Angelo Canty (McMaster University)

MacNaughton Building, room/local 105

Abstracts/Résumés

ANTONIO CIAMPI (McGill University), **KATHLEEN OROS KLEIN, J. BRENT RICHARDS & CELIA GREENWOOD** (Lady Davis Institute for Medical Research, Jewish General Hospital)

[Wednesday June 6 / mercredi 6 juin, 13:15]

The Role of Statistical Learning Methods in Genetic Association Studies

Rôle des méthodes d'apprentissage statistique dans les études d'association génétique

Genome-wide association studies (GWAS) often use univariate tests to identify associations between SNPs and traits, although biology suggests that there must be interactions between genes. There exist powerful but rarely-used methods originating from the statistical learning literature that may assist in identifying multi-SNP relationships and gaining understanding of SNP-trait associations. We demonstrate the performance of such methods on sets of SNPs within 100kb of approximately 20 candidate genes associated with bone density measures.

Les études d'association pangénomiques (GWAS) emploient souvent des tests univariés pour identifier les associations entre SNP et caractères, même si la biologie suggère une interaction entre les gènes. Or, certaines méthodes puissantes, mais rarement utilisées, tirées de la littérature sur l'apprentissage statistique peuvent aider à identifier des relations multi-SNP et mieux comprendre les associations SNP-caractères. Nous démontrons la performance de telles méthodes sur des ensembles de SNP à environ 100ko d'une vingtaine de gènes candidats associés aux mesures de densité osseuse.

BINOD NEUPANE, JEMILA HAMID & JOSEPH BEYENE (McMaster University)

[Wednesday June 6 / mercredi 6 juin, 13:30]

Multivariate Meta-Analysis in Genetic Association Studies

Méta-analyse multivariée dans les études d'association génétique

In meta-analysis, multivariate approaches might be preferable over univariate methods for estimating parameters of interest based on correlated phenotypes. In genetic association studies, missing data are less likely while associations are generally very small to modest and could be heterogeneous, so power to detect such associations as well as their unbiased and precise estimations are the concerns. We carried out simulations to investigate performance of multivariate meta-analytic methods. Our preliminary results suggest that multivariate techniques perform better in terms of power, coverage and bias in some scenarios. We will present detailed results of our comparative analysis for different realistic scenarios.

Pour les méta-analyses, des approches multivariées peuvent être préférables aux approches univariées pour estimer des paramètres fondés sur des phénotypes corrélés. Dans les études d'association génétique, les données manquantes sont rares et les associations généralement très faible ou modestes, et peuvent être hétérogènes. Donc, on s'inquiète de détecter de telles associations et de les estimer de façon précise et sans biais. Nous utilisons des simulations pour étudier la performance de méthodes de méta-analyse multivariées. Nos résultats préliminaires suggèrent que sous certains scénarios, les techniques multivariées performant mieux en terme de puissance, de couverture et de biais. Nous présenterons des résultats détaillés de notre analyse comparative pour différents scénarios réalistes.

3C-C3: Methods for Genetic Association Méthodes d'association génétique

JULIA TALEBAN, LAURA L. FAYE, LEI SUN, APOSTOLOS DIMITROMANOLAKIS & SHELLEY BULL (Samuel Lunenfeld Research Institute)

[Wednesday June 6 / mercredi 6 juin, 13:45]

Bootstrap Bias-reduction in Genetic Association Analysis of Time-to-event Outcomes

Réduction du biais par bootstrap dans une analyse d'association génétique du temps avant événement

While it is necessary to control the false-positive error rate in genome-wide studies, selection of genetic markers (SNPs) with small association p-values introduces optimistic bias into parameter estimates. Motivated by an investigation of complications in diabetes, we extend a bootstrap-resampling method for quantitative/binary traits to analyse time to nephropathy under a proportional hazards model. We evaluate relative and absolute bias via simulations. For SNPs with low to moderate power, bootstrap estimates are closer to the truth than uncorrected estimates, but the method tends to over-correct when power is high. Among false positives, however, bias-reduction shrinks estimates appropriately toward the null.

Bien qu'il soit nécessaire de contrôler le taux de faux négatifs dans les études pangénomiques, la sélection de marqueurs génétiques (SNP) avec de petites valeurs-p d'association introduit un biais optimiste dans les valeurs estimées des paramètres. Motivés par une étude sur les complications du diabète, nous élargissons une méthode de rééchantillonnage bootstrap pour les caractères quantitatifs/binaires pour analyser le temps avant la néphropathie dans un modèle de risque proportionnel. Nous évaluons le biais relatif et absolu à l'aide de simulations. Pour les SNP de puissance faible à modérée, les estimations bootstrap sont plus précises que les estimations non corrigées, mais la méthode a tendance à surcorriger lorsque la puissance est élevée. Pour les faux positifs, toutefois, la réduction du biais réduit convenablement les estimations vers zéro.

ANDRIY DERKACH & LEI SUN (University of Toronto), **JERRY LAWLESS** (University of Waterloo)

[Wednesday June 6 / mercredi 6 juin, 14:00]

Combining Linear and Quadratic Tests for Rare Variants Provides a Robust Test Across Genetic Models

Combiner des tests linéaires et quadratiques pour des variantes rares produit un test robuste pour différents modèles génétiques

Rare variants play an important role in complex human diseases and traits. Although many association tests have been proposed for rare variants, there is much confusion about the practical choice of a good test. Recent evaluations categorize methods into two classes: linear statistics sensitive to specific directional alternatives or the omnibus quadratic statistics. However, neither type of tests consistently outperforms the other. To achieve robustness, we consider the minimal p-value approach and the Fisher's method of combining p-values from linear and quadratic tests. Analytical and extensive simulation studies show that both methods are robust across genetic models with varying parameters.

Les variantes rares jouent un rôle important sur les traits et maladies complexes chez l'humain. Bien que plusieurs tests d'association aient été proposés pour les variantes rares, il y a beaucoup de confusion au sujet du choix d'un bon test en pratique. Des évaluations récentes catégorisent les méthodes en deux classes : les statistiques linéaires, sensibles aux directions des alternatives, et les statistiques quadratiques omnibus. Toutefois, aucun type de test n'est uniformément meilleur que l'autre. Pour assurer la robustesse, nous considérons l'approche de la valeur p minimale et l'approche de Fisher pour combiner des valeurs p de tests linéaires et quadratiques. Des études analytiques et par simulation montrent que les deux méthodes sont robustes pour des modèles génétiques avec différents paramètres.

3C-C3: Methods for Genetic Association Méthodes d'association génétique

ZHIJIAN CHEN (Samuel Lunenfeld Research Institute), **RADU V. CRAIU** (University of Toronto), **SHELLEY BULL** (Samuel Lunenfeld Research Institute)

[Wednesday June 6 / mercredi 6 juin, 14:15]

Sequential Two-Phase Stratified Designs for Regional Sequencing Following Genome-Wide Association Study: A Bayesian Approach

Plans stratifiés séquentiels à deux phases pour séquençage régional suite à une étude d'association pangénomique : une approche bayésienne

In focused follow-up studies, investigators may choose to sequence an entire genomic region of interest at the base pair level to identify potential causal variants. We consider Bayesian sequential two-phase designs, in which a subset of phase 1 subjects are selected from genotype categories of a genetic marker and sequenced in a target region in phase 2. At each sampling point, we apply Bayesian model averaging to account for genetic model uncertainty with an adaptive sample size allocation strategy to improve estimation efficiency of the genetic association parameter. We present numerical results for quantitative and binary traits.

Dans les études de suivi ciblées, les chercheurs séquentent parfois une région génomique d'intérêt au niveau des paires de bases pour identifier des variantes causales. Nous étudions les plans bayésiens séquentiels à deux phases où certains sujets sont sélectionnés parmi les catégories de génotype d'un marqueur génétique, puis séquencés dans une région cible. À chaque point d'échantillonnage, nous appliquons la combinaison bayésienne pour tenir compte de l'incertitude génétique du modèle et une stratégie de répartition de la taille de l'échantillon adaptative pour améliorer l'efficacité d'estimation du paramètre d'association génétique. Nous présentons des résultats numériques pour divers caractères quantitatifs et binaires.

ZEYNEP BASKURT (University of Toronto), **LISA J. STRUG** (The Hospital For Sick Children)

[Wednesday June 6 / mercredi 6 juin, 14:30]

A Composite Likelihood Approach Using the Evidential Paradigm to Analyze Genetic Association in Pedigrees

Approche de vraisemblance composite par le paradigme de preuves pour analyser l'association génétique dans les lignées

Royall (1997) proposed the evidential paradigm, an alternative to Frequentist and Bayesian paradigms for interpreting data as evidence. The evidential paradigm uses the likelihood ratio (LR) for two simple hypotheses as an objective measure of the strength of statistical evidence. In genetic studies, LRs are commonly used to measure evidence. However, evidential association of pedigrees is not straightforward due to complex family structures. We propose to use composite likelihoods to construct LRs for evidential analysis of families. We show how to make these LRs robust from model misspecification; that they have good operational characteristics; and are consistent with competing methods.

Royall (1997) a proposé le paradigme de preuves comme alternative aux paradigmes fréquentistes et bayésiens pour interpréter les données comme preuves. Celui-ci utilise le rapport de vraisemblance (RV) de deux hypothèses simples comme mesure objective de la solidité des preuves statistiques. Dans les études génétiques, les RV servent souvent à évaluer la preuve, mais celle-ci est difficile à établir en présence de structures familiales très complexes. Nous utilisons des vraisemblances composites pour construire des RV permettant l'analyse probante des familles. Nous expliquons comment protéger ces RV contre les erreurs de spécification du modèle et montrons qu'elles présentent de bonnes caractéristiques opérationnelles et sont compatibles avec les modèles concurrents.

Author List • Liste des auteurs

- Abarin, Taraneh, 44, 172 (2C-C1)
 Abdous, Belkacem, 41 (2B-C5), 45 (2C-C2), 157 (2B-C5), 174 (2C-C2)
 Abraham, Bovas, 53, 215 (3A-A3)
 Acar, Elif Fidan, 41, 157 (2B-C5)
 Adamic, Dr. Peter, 36, 134 (1D-C5)
 Adekambi, Franck, 60, 246 (3C-C1)
 Adjogou, Folly, 60, 248 (3C-C2)
 Aghababazadeh, Farnoosh Abbas, 46, 179 (2C-C4)
 Agnihotram, Ramana-Kumar, 35, 127 (1D-C2)
 Ahmed, Ejaz, 42, 161 (2C-I1)
 Ahmed, Sara, 28, 94 (1C-C1)
 Alam, S. M. Khurshid, 29, 100 (1C-C3)
 Albert-Green, Alisha, 46, 177 (2C-C3)
 Ali, Ayesha, 29 (1C-C3), 51 (2D-C3), 99 (1C-C3), 207 (2D-C3)
 Al Labadi, Luai, 55, 221 (3B-I4)
 Allen, Brian, 22 (1A-A1)
 Almalki, Tagreed, 31, 109 (1C-D2)
 Alsulami, Huda, 31, 109 (1C-D2)
 Alvo, Mayer, 46, 179 (2C-C4)
 Andrews, Jeffrey, 22, 65 (1B-I1)
 Angevaare, Justin, 46, 182 (2C-P)
 Apaloo, Joseph, 36, 134 (1D-C5)
 Arenovich, Tamara, 55, 223 (3B-I5)
 Asgharian, Masoud, 43 (2C-I2), 48 (2D-I2), 50 (2D-C1), 163 (2C-I2), 192, 193 (2D-I2), 201 (2D-C1)
 Atherton, Juli, 34 (1D-I5), 41 (2B-C4), 122 (1D-I5), 154 (2B-C4)
 Augustyniak, Maciej, 26, 83 (1B-C4)
- Babul, Arif, 31, 106 (1C-C5)
 Badescu, Alexandru, 44, 169 (2C-I5)
 Bae, Taehan, 50, 204 (2D-C2)
 Baek, Jangsun, 22, 65 (1B-I1)
 Bailey, Rosemary A., 59, 241 (3C-I3)
 Balakrishnan, Narayanaswamy, 50, 202 (2D-C1)
 Balendran, Prathayana, 31, 110 (1C-P)
- Bashir, Shaheena, 46, 182 (2C-P)
 Baskurt, Zeynep, 29 (1C-C3), 61 (3C-C3), 99 (1C-C3), 252 (3C-C3)
 Bayne, Erin, 54, 219 (3B-I3)
 Beaumont, Jean-François, 38, 142 (2B-I3)
 Becker, Devan, 46, 177 (2C-C3)
 Beg, Faisal, 31, 106 (1C-C5)
 Béliveau, Audrey, 31, 108 (1C-D1)
 Bellhouse, David R., 32 (1C-P), 33 (1D-I2), 113 (1C-P), 117 (1D-I2)
 Benach, Joan, 43, 163 (2C-I2)
 Benavides, Fernando G., 43, 163 (2C-I2)
 Berger, James O., 22, 63 (1A-A2)
 Bergeron, Pierre-Jérôme, 48, 192 (2D-I2)
 Berke, Olaf, 45, 176 (2C-C3)
 Bernard, Carole, 51, 205 (2D-C2)
 Beyene, Joseph, 30 (1C-C4), 32 (1C-P), 46 (2C-C4), 57 (3B-C3), 60 (3C-C3), 103 (1C-C4), 111 (1C-P), 180 (2C-C4), 231 (3B-C3), 250 (3C-C3)
 Bhattacharya, Sakyajit, 42, 161 (2C-I1)
 Biccard, Bruce M., 55, 225 (3B-C1)
 Bickel, David R., 40 (2B-C3), 46 (2C-C4), 151 (2B-C3), 179 (2C-C4)
 Bifolchi, Nadia, 52, 210 (2D-C4)
 Bingham, Derek, 24, 71, 72 (1B-I4)
 Bocci, Chiara, 38, 142 (2B-I3)
 Boente, Graciela, 59, 243 (3C-I4)
 Boiteau, Véronique, 55, 225 (3B-C1)
 Bolliger, Daniel, 55, 225 (3B-C1)
 Bombak, Andrea E., 28, 95 (1C-C1)
 Bonner, Ashley, 30, 103 (1C-C4)
 Borg Debono, Victoria, 47, 182 (2C-P)
 Bouchard-Côté, Alexandre, 48, 190 (2D-I1)
 Boudreau, Christian, 25, 77 (1B-C2)
 Braun, John, 31, 32 (1C-P), 32 (1D-I1), 110, 112 (1C-P), 115 (1D-I1)
 Brewster, John, 22 (1A-A1), 28, 93 (1C-I4)
 Brien, Chris, 59, 241 (3C-I3)

- Brillinger, David, 33, 118 (1D-I3)
 Briollais, Laurent, 56, 228 (3B-C2)
 Brodie, Pete, 23, 67 (1B-I2)
 Brown, Patrick, 32 (1C-P), 32, 33 (1D-I1), 113 (1C-P),
 115 (1D-I1)
 Browne, Ryan, 38 (2B-I2), 51 (2D-C3), 140, 141 (2B-I2),
 208 (2D-C3)
 Bull, Shelley, 60, 61, 250, 251 (3C-C3)
 Burke, Murray, 40, 149 (2B-C2)
 Burkhart, Christoph S., 55, 225 (3B-C1)
 Burr, Wesley, 26 (1B-C4), 57 (3B-C4), 84 (1B-C4), 235
 (3B-C4)
 Buse, Giovana A. Lurati, 55, 225 (3B-C1)
- Cabilio, Paul, 39, 145 (2B-I4)
 Cadigan, Noel, 29, 97 (1C-C2)
 Cai, Eric, 40, 152 (2B-C3)
 Cai, Song, 30, 105 (1C-C5)
 Cai, Tianxi, 39, 147 (2B-C1)
 Callaghan, Liam, 51, 207 (2D-C3)
 Canty, Angelo, 47, 184 (2C-P)
 Cao, Jiguo, 37, 138 (2B-I1)
 Caron, Sylvain, 36, 134 (1D-C5)
 Carone, Marco, 48, 192 (2D-I2)
 Carriere, Keumhee Chough, 36, 132 (1D-C4)
 Cartea, Alvaro, 44, 169 (2C-I5)
 Castura, John, 38, 141 (2B-I2)
 Celini, Léonide, 29, 97 (1C-C2)
 Chakraborty, Bibhas, 58, 237 (3C-I1)
 Chambers, Larry, 24, 73 (1B-C1)
 Chapman, Judy-Anne W., 42, 158 (2B-C5)
 Charest, Anne-Sophie, 53, 216 (3B-I1)
 Charrad, Malika, 55, 225 (3B-C1)
 Chen, Bei, 53, 215 (3A-A3)
 Chen, Fuqi, 30, 105 (1C-C5)
 Chen, Gemai, 50, 201 (2D-C1)
 Chen, Jiahua, 53, 214 (3A-A2)
 Chen, Maggie, 31, 108 (1C-D1)
 Chen, Min, 40, 151 (2B-C3)
 Chen, Xin, 31, 109 (1C-D2)
 Chen, Yang, 31, 109 (1C-D2)
 Chen, Zhengfei, 47, 183 (2C-P)
 Chen, Zhijian, 61, 251 (3C-C3)
 Chipman, Hugh, 39, 146 (2B-C1)
 Choi, Yun-Hee, 25 (1B-C2), 57 (3B-C3), 77 (1B-C2), 233
 (3B-C3)
 Chowdhury, Rafiqul, 57, 233 (3B-C3)
 Chu, Maggie, 46, 180 (2C-C4)
- Chu, Rong, 57, 232 (3B-C3)
 Ciampi, Antonio, 60 (3C-C2), 60 (3C-C3), 248 (3C-C2),
 250 (3C-C3)
 Claggett, Brian, 39, 147 (2B-C1)
 Clèries, Ramon, 24 (1B-C1), 43 (2C-I2), 74 (1B-C1), 163
 (2C-I2)
 Coia, Vincenzo, 59, 245 (3C-C1)
 Connolly, Jessica, 31, 110 (1C-P)
 Connolly, Stuart J., 47, 186 (2C-P)
 Cook, Richard J., 24 (1B-C1), 25 (1B-C2), 74 (1B-C1),
 78 (1B-C2)
 Corey, Paul, 28, 93 (1C-I4)
 Correll, Ray L., 59, 241 (3C-I3)
 Cotton, Cecilia, 24 (1B-C1), 35 (1D-C2), 74 (1B-C1), 127
 (1D-C2)
 Coulombe, Janie, 31, 109 (1C-D2)
 Cowling, Ben, 52, 212 (2D-C4)
 Craiu, Radu V., 41 (2B-C5), 61 (3C-C3), 157 (2B-C5),
 251 (3C-C3)
 Crawford, Stephen, 47, 185 (2C-P)
 Crea, Catherine, 29, 99 (1C-C3)
 Csörgő, Miklós, 50, 199, 200 (2D-I5)
 Cuerden, Meaghan S., 24, 74 (1B-C1)
 Cui, Zhenyu, 51, 205 (2D-C2)
 Currie, Lauren, 29, 98 (1C-C2)
- Dadimuni, Shyam, 34, 124 (1D-C1)
 Dang, Sanjeena, 60, 249 (3C-C2)
 Dansu, Bolanle, 55, 226 (3B-C1)
 Darlington, Gerarda, 40, 148 (2B-C2)
 Datta, Gauri, 33, 120 (1D-I4)
 Dawson, Don, 26, 87 (1C-I1)
 Dean, Charmaine, 46, 177 (2C-C3)
 Deardon, Rob, 32 (1C-P), 35 (1D-C3), 51, 52 (2D-C4),
 54 (3B-I3), 111 (1C-P), 129 (1D-C3), 210–212
 (2D-C4), 219 (3B-I3)
 Deeth, Lorna, 52, 211 (2D-C4)
 de Leon, Alexander R., 29 (1C-C3), 42 (2B-C5), 99 (1C-
 C3), 158 (2B-C5)
 Dendukuri, Nandini, 35, 127, 128 (1D-C2)
 Derkach, Andriy, 61, 251 (3C-C3)
 Devereaux, P.J., 57, 232 (3B-C3)
 De Villa, Eileen, 52, 210 (2D-C4)
 Dewey, Cate E., 32, 111 (1C-P)
 Diao, Liqun, 25, 78 (1B-C2)
 Dimitromanolakis, Apostolos, 60, 250 (3C-C3)
 Ding, Jing, 47, 183 (2C-P)
 Doksum, Kjell A., 42, 161 (2C-I1)

- Dolovich, Lisa, 24 (1B-C1), 28 (1C-C1), 73 (1B-C1), 96 (1C-C1)
- Dong, Bin, 42, 158 (2B-C5)
- Donner, Allan, 54, 218 (3B-I2)
- Dorri, Fatemeh, 25, 81 (1B-C3)
- Doss, Forrest W., 24, 71 (1B-I4)
- Doucas, Marie-Christine, 31, 109 (1C-D2)
- Doucet, Arnaud, 48, 190 (2D-I1)
- Dowd, Michael, 23, 69 (1B-I3)
- Drignei, Dorin, 24, 71 (1B-I4)
- Drikvandi, Reza, 56, 228 (3B-C2)
- Du, Yeting, 37, 138 (2B-I1)
- Dubin, Joel A., 45 (2C-C3), 56 (3B-C2), 176 (2C-C3), 230 (3B-C2)
- Duchesne, Simon, 28, 95 (1C-C1)
- Duchesne, Thierry, 39 (2B-C1), 45 (2C-C2), 146 (2B-C1), 174 (2C-C2)
- Dunham, Bruce, 49, 195 (2D-I3)
- Eberhardt, Frederick, 58, 237 (3C-I1)
- Edinger, B. Johannes, 29, 98 (1C-C2)
- Eikelboom, John, 47, 186 (2C-P)
- El Shaarawi, Abdel, 39, 144 (2B-I4)
- Ergun, Tolga, 56, 231 (3B-C3)
- Ernst, Pierre, 28, 94 (1C-C1)
- Ertefaei, Ashkan, 48, 193 (2D-I2)
- Eustace, Jayde, 47, 184 (2C-P)
- Evans, Michael, 29, 99 (1C-C3)
- Fan, Chun-Po Steve, 32, 115 (1D-I1)
- Fan, Jianqing, 42, 161 (2C-I1)
- Fan, Lingling, 31, 109 (1C-D2)
- Farahat, Ahmed, 41, 154 (2B-C4)
- Faye, Laura L., 60, 250 (3C-C3)
- Feder, Moshe, 25, 80 (1B-C3)
- Feng, Rui, 46, 179 (2C-C4)
- Feng, Shui, 55, 221 (3B-I4)
- Feng, Yang, 42, 161 (2C-I1)
- Feng, Zeny, 35 (1D-C3), 52 (2D-C4), 56 (3B-C2), 129 (1D-C3), 210 (2D-C4), 229 (3B-C2)
- Ferland, René, 50, 204 (2D-C2)
- Field, Chris, 28 (1C-I4), 29 (1C-C3), 59 (3C-I4), 93 (1C-I4), 101 (1C-C3), 243 (3C-I4)
- Findlay, Chris, 38, 140 (2B-I2)
- Fode, Tounkara, 41, 157 (2B-C5)
- Foster, Gary, 28, 96 (1C-C1)
- Franco, Eduardo, 35, 127 (1D-C2)
- Franczak, Brian, 51, 208 (2D-C3)
- Fredette, Marc, 35, 129 (1D-C3)
- Fujikoshi, Yasunori, 41, 155 (2B-C4)
- Gagnon, Éric, 54, 216 (3B-I1)
- Gallagher, Marcus, 37, 138 (2B-I1)
- Gangeh, Mehrdad, 45, 174 (2C-C2)
- Gao, Xin, 39, 144 (2B-I4)
- Gel, Yulia, 26 (1B-C4), 39 (2B-I4), 52 (2D-C4), 53 (3A-A3), 84 (1B-C4), 144 (2B-I4), 210 (2D-C4), 215 (3A-A3)
- Genest, Christian, 33 (1D-I2), 37 (2A-A), 58 (3C-I2), 117 (1D-I2), 136 (2A-A), 239 (3C-I2)
- George, Michailidis, 44, 168 (2C-I4)
- Ghazzali, Nadia, 55, 225 (3B-C1)
- Ghodsai, Ali, 25 (1B-C3), 30 (1C-C4), 41 (2B-C4), 45 (2C-C2), 51 (2D-C3), 81 (1B-C3), 102 (1C-C4), 154 (2B-C4), 174 (2C-C2), 207 (2D-C3)
- Gill, Paramjit, 21, 48, 190 (2D-I1)
- Gillis, Daniel, 46, 47 (2C-P), 52 (2D-C4), 56 (3B-C1), 182, 185 (2C-P), 211 (2D-C4), 226 (3B-C1)
- Goeree, Ron, 24, 75 (1B-C1)
- Goh, Joslin, 24, 71 (1B-I4)
- Gold, Jourdan, 35, 129 (1D-C3)
- Goodrich, Ben, 27, 91 (1C-I3)
- Gosh, Malay, 38, 142 (2B-I3)
- Goss, Paul E., 42, 158 (2B-C5)
- Greenwood, Celia, 46 (2C-P), 60 (3C-C3), 182 (2C-P), 250 (3C-C3)
- Groendyke, Chris, 44 (2C-I4), 51 (2D-C2), 167 (2C-I4), 205 (2D-C2)
- Grosskopf, Michael J., 24, 71 (1B-I4)
- Gu, Chong, 53, 213 (3A-A1)
- Guillotte, Simon, 48, 190 (2D-I1)
- Guo, Qing, 24, 75 (1B-C1)
- Guo, Shu, 31, 109 (1C-D2)
- Gustafson, Paul, 43, 163 (2C-I2)
- Guyatt, Gordon, 57, 232 (3B-C3)
- Halder, Avik, 25, 77 (1B-C2)
- Haley, Charlotte L., 57, 234 (3B-C4)
- Hall, Geoffrey, 24, 75 (1B-C1)
- Hall, Kevin, 22 (1A-A1)
- Hall, Peter, 33, 120 (1D-I4)
- Hamid, Jemila, 32 (1C-P), 60 (3C-C3), 111 (1C-P), 250 (3C-C3)
- Han, L. Spectrum, 31, 110 (1C-P)
- Hanley, James A., 35, 127, 128 (1D-C2)
- Hann, Crystal, 35, 127 (1D-C2)
- Harch, Bronwyn D., 59, 241 (3C-I3)
- Hartman, Brian M., 51, 205 (2D-C2)

- Hatefi, Armin, 51, 207 (2D-C3)
 Haziza, David, 38, 142 (2B-I3)
 He, Wenqing, 46 (2C-C4), 48 (2C-P), 180 (2C-C4), 188 (2C-P)
 Hidioglou, Mike, 23 (1B-I2), 43, 165 (2C-I3)
 Hoang, Thi Lam Uyen, 31, 109 (1C-D2)
 Hodgson, David, 35, 127 (1D-C2)
 Hoeninger, Julia, 54, 216 (3B-I1)
 Hoerl, Roger, 42, 160 (2C-A)
 Höhne, Jörg, 54, 216 (3B-I1)
 Holloway, James Paul, 24, 71 (1B-I4)
 Hooker, Giles, 23, 69 (1B-I3)
 Horrocks, Julie, 40 (2B-C2), 47 (2C-P), 148 (2B-C2), 185 (2C-P)
 Horton, Nick, 23, 67 (1B-I2)
 Hossain, Ahmed, 46, 180 (2C-C4)
 Hossain, M. Shakhawat, 42, 161 (2C-I1)
 Hoyer, Patrik, 58, 237 (3C-I1)
 Hu, Joan, 32, 115 (1D-I1)
 Huang, Longlong, 36, 135 (1D-C5)
 Huang, Mei Ling, 59, 245 (3C-C1)
 Huh, Ick, 50, 202 (2D-C1)
 Hunt, Scott, 36, 134 (1D-C5)
 Hunter, David R., 44, 167 (2C-I4)
 Hurlburt, Lindsay, 47, 182 (2C-P)
 Hussein, Abdulkadir, 40, 148 (2B-C2)
 Hyttinen, Antti, 58, 237 (3C-I1)
- Imori, Shinpei, 41, 154 (2B-C4)
 Ingrassia, Salvatore, 23, 65 (1B-I1)
 Iscoe, Ian, 50, 204 (2D-C2)
 Islam, Mohammad Shahidul, 31, 109 (1C-D2)
 Islam, Muhammad, 45, 174 (2C-C2)
 Islam, Shofiqul, 32, 111 (1C-P)
 Ismail, Salimah, 31, 106 (1C-C5)
- Jafari Jozani, Mohammad, 45 (2C-C2), 51 (2D-C3), 175 (2C-C2), 207 (2D-C3)
 Jahanshahloo, Gholamreza, 50, 201 (2D-C1)
 Jaimungal, Sebastian, 44, 169 (2C-I5)
 Jang, Gun Ho, 46, 179 (2C-C4)
 Jankowski, Hanna, 31, 110 (1C-P)
 Ji, Yunqi, 60, 248 (3C-C2)
 Jiang, Depeng, 28, 95 (1C-C1)
 Jiang, Jiming, 34, 120 (1D-I4)
 Jiang, Weibin, 33, 118 (1D-I3)
 John, Remya M., 47, 184 (2C-P)
 Johnson, Brad, 45, 175 (2C-C2)
 Junxin, Liu, 44, 171 (2C-C1)
- Kaczorowski, Janusz, 24, 73 (1B-C1)
 Kamel, Mohamed, 41 (2B-C4), 45 (2C-C2), 154 (2B-C4), 174 (2C-C2)
 Katsahian, Sandrine, 25, 77 (1B-C2)
 Ker, Alan, 34 (1D-C1), 56 (3B-C3), 124 (1D-C1), 231 (3B-C3)
 Kerby, Matthew, 47, 187 (2C-P)
 Khalili, Abbas, 36 (1D-C3), 37 (2B-I1), 131 (1D-C3), 138 (2B-I1)
 Khan, Shahedul A., 45, 176 (2C-C3)
 Khodadadi, Ahmad, 56, 228 (3B-C2)
 Khribi, Lotfi, 35, 129 (1D-C3)
 Kiapour, Azadeh, 60, 246 (3C-C1)
 Kim, Jae-Kwang, 43 (2C-I3), 53 (3A-A2), 165 (2C-I3), 214 (3A-A2)
 King, Aaron A., 23, 70 (1B-I3)
 Koch, Dean, 31, 108 (1C-D1)
 Kolkiewicz, Adam, 59, 245 (3C-C1)
 Kopciuk, Karen, 36, 135 (1D-C5)
 Kopp, Gregory, 32, 112 (1C-P)
 Krieg, Sabine, 38, 142 (2B-I3)
 Krishnamurthy, Ashok, 32, 112 (1C-P)
 Krislock, Nathan, 30, 102 (1C-C4)
 Kulik, Rafal, 33, 119 (1D-I3)
 Kulmatitskiy, Nikolay, 44, 167 (2C-I4)
 Kulperger, Reg, 26 (1B-C4), 45 (2C-C3), 84 (1B-C4), 176 (2C-C3)
 Kuranz, Carolyn C., 24, 71 (1B-I4)
 Kwong, Grace P.S., 32 (1C-P), 52 (2D-C4), 111 (1C-P), 212 (2D-C4)
- Labbé, Chantal, 50, 204 (2D-C2)
 Lafaye de Micheaux, Pierre, 41, 155 (2B-C4)
 Lahiri, Partha, 34, 121 (1D-I4)
 Lakhali-Chaieb, Lajmi, 39 (2B-C1), 41 (2B-C5), 45 (2C-C2), 46 (2C-C4), 146 (2B-C1), 157 (2B-C5), 174 (2C-C2), 181 (2C-C4)
 Laniel, Normand, 49, 197 (2D-I4)
 Laroche, François, 56, 231 (3B-C3)
 Lawless, Jerry, 28 (1C-C1), 28 (1C-I4), 61 (3C-C3), 93 (1C-I4), 95 (1C-C1), 251 (3C-C3)
 Le, Vu Thien Huong, 31, 109 (1C-D2)
 Leclerc, Martin, 46, 181 (2C-C4)
 Lee, Jonathan, 45, 176 (2C-C3)
 Lefebvre, Geneviève, 34 (1D-I5), 41 (2B-C4), 122 (1D-I5), 154 (2B-C4)
 Léger, Christian, 41, 155 (2B-C4)
 Lele, Subash, 54, 219 (3B-I3)

- Lemieux, Christiane, 35, 129 (1D-C3)
 Letac, Gerard, 58, 237 (3C-I1)
 Leung, Andy, 31, 108 (1C-D1)
 Li, Deli, 29 (1C-C3), 54 (3B-I4), 100 (1C-C3), 221 (3B-I4)
 Li, Haocheng, 56, 229 (3B-C2)
 Li, Jingya, 44, 170 (2C-I5)
 Li, Lennon, 33, 115 (1D-I1)
 Li, Longhai, 45, 176 (2C-C3)
 Li, Moyi, 38, 141 (2B-I2)
 Li, Pengfei, 47, 185 (2C-P)
 Li, Xuan, 36, 133 (1D-C4)
 Li, Yin, 36, 132 (1D-C4)
 Lian, Heng, 40, 149 (2B-C2)
 Liang, Hua, 40, 149 (2B-C2)
 Lim, Hyun (June), 28, 94 (1C-C1)
 Lin, Kevin, 23, 69 (1B-I3)
 Lin, Wei, 31, 109 (1C-D2)
 Liu, Juxin, 43, 163 (2C-I2)
 Liu, Wei, 45, 172 (2C-C1)
 Liu, Weiwei, 26, 83 (1B-C4)
 Liu, Zhihui (Amy), 35, 127, 128 (1D-C2)
 Liu, Zi Zhen, 26, 84 (1B-C4)
 Lix, Lisa, 44 (2C-C1), 55 (3B-C1), 171 (2C-C1), 226 (3B-C1)
 Loeppky, Jason, 24, 72 (1B-I4)
 Loredó-Osti, J. Concepción, 47, 184, 188 (2C-P)
 Lourenço, Vanda, 30, 106 (1C-C5)
 Lu, Xin, 31, 109 (1C-D2)
 Lu, Xuwen, 36 (1D-C5), 40 (2B-C2), 135 (1D-C5), 149 (2B-C2)
 Luo, Wei, 37, 138 (2B-I1)
 Lyubchich, Vyacheslav, 26 (1B-C4), 39 (2B-I4), 84 (1B-C4), 144 (2B-I4)
- Ma, Jinhui, 57, 231 (3B-C3)
 Ma, Ping, 53, 213 (3A-A1)
 Ma, Wenkai, 32, 112 (1C-P)
 MacEachern, Steve, 22, 64 (1B-V)
 MacGibbon, Brenda, 35, 129 (1D-C3)
 Makosso-Kallyth, Sun, 30, 102 (1C-C4)
 Makvandi-Nejad, Ewa, 48, 192 (2D-I2)
 Malik, Rajat, 52, 212 (2D-C4)
 Mandal, Saumen, 36, 132 (1D-C4)
 Martell, David, 46, 177 (2C-C3)
 Martín, Nirian, 25, 80 (1B-C3)
 Martinez, Jose Miguel, 24 (1B-C1), 43 (2C-I2), 74 (1B-C1), 163 (2C-I2)
 Martsynyuk, Yuliya, 50, 199 (2D-I5)
- Massam, Helene, 58, 237 (3C-I1)
 Mathieu, Patrice, 49, 197 (2D-I4)
 Mayhew, Linda, 35, 127 (1D-C2)
 McCandless, Lawrence, 29, 98 (1C-C2)
 McGillivray, Annaliza, 36, 131 (1D-C3)
 McGregor, Maurice, 35, 128 (1D-C2)
 McKinnell, Skip, 47, 187 (2C-P)
 McKinnon, Margaret, 24, 75 (1B-C1)
 McLeish, Don, 27 (1C-I1), 51 (2D-C2), 87 (1C-I1), 205 (2D-C2)
 McLeod, Ian A., 32, 112 (1C-P)
 McNicholas, Paul, 22, 23 (1B-I1), 38 (2B-I2), 42 (2C-I1), 47, 48 (2C-P), 51 (2D-C3), 60 (3C-C2), 65 (1B-I1), 140, 141 (2B-I2), 161 (2C-I1), 187, 189 (2C-P), 208, 209 (2D-C3), 249 (3C-C2)
 McPherson, Matt, 52, 210 (2D-C4)
 McRae, Ken, 59 (3C-I3)
 Metzler, Adam, 44, 170 (2C-I5)
 Mihram, Danielle, 40, 152 (2B-C3)
 Mihram, G. Arthur, 40, 152 (2B-C3)
 Millar, Anne Michele, 45, 173 (2C-C1)
 Milliken, George, 59, 241 (3C-I3)
 Misir, Vachan, 31, 109 (1C-D2)
 Mitsakakis, Nicholas, 30, 105 (1C-C5)
 Mohammad, Nagham, 32, 112 (1C-P)
 Mohsenipour, Aliakbar, 34, 125 (1D-C1)
 Moiseev, Alexander, 31, 106 (1C-C5)
 Molina, Isabel, 25, 80 (1B-C3)
 Mondal, Prosanta, 28, 94 (1C-C1)
 Monette, Georges, 56, 228 (3B-C2)
 Moniruzzaman, Akm, 29, 98 (1C-C2)
 Moodie, Erica, 24 (1B-C1), 58 (3C-I1), 74 (1B-C1), 237 (3C-I1)
 Morgan, Melissa, 32, 112 (1C-P)
 Mostafa, Davtalab Olyaie, 50, 201 (2D-C1)
 Mouiha, Abderazzak, 28, 95 (1C-C1)
 Mudalige, Nishan, 31, 109 (1C-D2)
 Mulayath Variyath, Asokan, 40, 148 (2B-C2)
 Murdoch, Duncan, 40, 153 (2B-C3)
 Murray, Lori L., 32, 113 (1C-P)
 Murthy, Yamini, 47, 182 (2C-P)
 Murua, Alejandro, 46 (2C-C4), 60 (3C-C2), 180 (2C-C4), 248 (3C-C2)
 Mutch, David M., 47, 187 (2C-P)
 Muthukumarana, Saman, 34, 122 (1D-I5)
- Nadarajah, Tharshanna, 40, 148 (2B-C2)
 Nasari, Masoud, 50, 199 (2D-I5)

- Nathoo, Farouk, 31, 106 (1C-C5)
 Nematollahi, Nader, 60, 246 (3C-C1)
 Nešlehová, Johanna, 37 (2B-I1), 58 (3C-I2), 138 (2B-I1), 239 (3C-I2)
 Neupane, Binod, 60, 250 (3C-C3)
 Nguyen, Paul, 32 (1C-P), 32, 33 (1D-I1), 113 (1C-P), 115 (1D-I1)
 Nguyen, Thuan, 34, 120 (1D-I4)
 Niknafs, Azam, 55, 225 (3B-C1)
 Niu, Xiaoqing, 47, 185 (2C-P)
 Nkurunziza, Sévérien, 30 (1C-C5), 40 (2B-C2), 105 (1C-C5), 148 (2B-C2)
 Nourmohammadi, Mohammad, 45, 175 (2C-C2)
- Oldford, R. Wayne, 40, 152 (2B-C3)
 O'Reilly, Rachel, 51, 209 (2D-C3)
 Oros Klein, Kathleen, 60, 250 (3C-C3)
 Ortega, Juan-Pablo, 44, 169 (2C-I5)
 Oyet, Alwell, 52, 211 (2D-C4)
- Padilla, Marta, 40, 151 (2B-C3)
 Pan, Yiyang, 31, 108 (1C-D1)
 Panzarella, Tony, 55, 223 (3B-I5)
 Park, Seunghwan, 43, 165 (2C-I3)
 Partovi Nia, Vahid, 50 (2D-C1), 56 (3B-C2), 201 (2D-C1), 228 (3B-C2)
 Patterson, Michelle, 29, 98 (1C-C2)
 Paul, James, 47, 182 (2C-P)
 Paul, Sudhir, 29, 100 (1C-C3)
 Peng, Jianan, 39, 146 (2B-C1)
 Perron, François, 48, 190 (2D-I1)
 Perry, Patrick, 44, 167 (2C-I4)
 Petkau, John, 48, 190 (2D-I1)
 Pfeffermann, Danny, 25, 80 (1B-C3)
 Phipps, Polly, 23, 67 (1B-I2)
 Pichika, Sathish Chandra, 30, 103 (1C-C4)
 Picka, Jeffrey, 54, 219 (3B-I3)
 Pires, Ana Maria, 30, 106 (1C-C5)
 Platt, Robert, 54 (3B-I2)
 Pogue, Janice, 47, 186 (2C-P)
 Poljak, Zvonimir, 32, 111 (1C-P)
 Pollak, Michael N., 42, 158 (2B-C5)
 Pooneh, Pordeli, 40, 149 (2B-C2)
 Pramanik, Santanu, 34, 121 (1D-I4)
 Provost, Serge, 34, 124, 125 (1D-C1)
 Pullenayegum, Eleanor M., 24 (1B-C1), 56 (3B-C2), 75 (1B-C1), 228 (3B-C2)
 Punzo, Antonio, 23, 65 (1B-I1)
- Qi, Ying, 31 (1C-D1), 31 (1C-D2), 108 (1C-D1), 109 (1C-D2)
 Quan, Joanne, 49, 195 (2D-I3)
- Rad, Narges Nazeri, 28, 95 (1C-C1)
 Raffa, Jesse D., 56, 230 (3B-C2)
 Raifu, Amidu Olalekan, 35, 127 (1D-C2)
 Raina, Parminder, 57, 231 (3B-C3)
 Ramadan, Anas, 34, 124 (1D-C1)
 Ramandi, Babak Alipanahi, 30, 102 (1C-C4)
 Ramirez Ramirez, Lilia Leticia, 52, 210 (2D-C4)
 Ramsay, James, 60, 248 (3C-C2)
 Rana, Masud, 45, 176 (2C-C3)
 Rao, J. Sunil, 34, 120 (1D-I4)
 Rao, Jon N.K., 27, 88 (1C-I1)
 Rathouz, Paul, 21
 Raymond-Blaess, Virginie, 38, 142 (2B-I3)
 Reesor, R. Mark, 44, 170 (2C-I5)
 Reid, Nancy, 27, 89 (1C-I2)
 Reiter, Jerome P., 27, 92 (1C-I3)
 Rémillard, Bruno, 58, 239 (3C-I2)
 Ren, Zhaoxia (Jeana), 59, 245 (3C-C1)
 Ribes, Josepa, 24, 74 (1B-C1)
 Richards, J. Brent, 60, 250 (3C-C3)
 Riegert, David, 57, 235 (3B-C4)
 Rivest, Louis-Paul, 21, 41 (2B-C5), 58 (3C-I2), 157 (2B-C5), 239 (3C-I2)
 Robert Lambert, Marc-Olivier, 55, 225 (3B-C1)
 Rodseth, Reitze N., 55, 225 (3B-C1)
 Rogers, Bruce, 23, 69 (1B-I3)
 Romdhani, Héla, 41, 157 (2B-C5)
 Romo, Rebecca, 45, 173 (2C-C1)
 Roy, Valérie, 54, 216 (3B-I1)
 Rueffer, Matthew, 47, 185 (2C-P)
 Ruiz-Gazen, Anne, 38, 142 (2B-I3)
 Rutter, Erica, 24, 71 (1B-I4)
 Ryan, Kathleen, 47, 185 (2C-P)
- Saarela, Olli, 24, 74 (1B-C1)
 Saeedi, Ardavan, 48, 190 (2D-I1)
 Sajobi, Tolulope, 55, 226 (3B-C1)
 Salibián-Barrera, Matías, 59, 243 (3C-I4)
 Sallier, Kenza, 31, 109 (1C-D2)
 Sam, Abdoul, 56, 231 (3B-C3)
 Sarhan, Ammar, 36, 134 (1D-C5)
 Scharfstein, Daniel, 48, 192 (2D-I2)
 Schau, Candace, 45, 173 (2C-C1)
 Schenker, Nathaniel, 27, 91 (1C-I3)
 Shandizi, Amir Hosein Homaie, 50, 201 (2D-C1)

- Shen, Hua, 25, 78 (1B-C2)
 Sheng, Susan Zhe, 34, 125 (1D-C1)
 Shepherd, Lois E., 42, 158 (2B-C5)
 Shestakovska, Olga, 47, 186 (2C-P)
 Shin, Hwashin H., 26, 84 (1B-C4)
 Shinohara, Taki R., 48, 193 (2D-I2)
 Shojaie, Ali, 44, 168 (2C-I4)
 Si, Yajuan, 27, 92 (1C-I3)
 Sinha, Sanjoy, 59 (3C-I4)
 Sirski, Monica, 47, 186 (2C-P)
 Skinner, Stuart, 28, 94 (1C-C1)
 Smeets, Marc, 38, 142 (2B-I3)
 Sneddon, Gary, 47, 187 (2C-P)
 Soave, David, 55, 224 (3B-I5)
 Solymos, Peter, 54, 219 (3B-I3)
 Somers, Julian, 29, 98 (1C-C2)
 Song, Peter X.-K., 27, 89, 90 (1C-I2)
 Soulier, Philippe, 33, 119 (1D-I3)
 Springford, Aaron, 57, 234, 235 (3B-C4)
 Srivastava, Muni, 41, 156 (2B-C4)
 Stafford, James, 32 (1C-P), 32, 33 (1D-I1), 113 (1C-P),
 115 (1D-I1)
 Steele, Russell, 37, 138 (2B-I1)
 Stephens, David, 24 (1B-C1), 48 (2D-I2), 74 (1B-C1),
 193 (2D-I2)
 Stewart, Connie, 29, 101 (1C-C3)
 Strug, Lisa J., 61, 252 (3C-C3)
 Strumpf, Erin, 35, 127, 128 (1D-C2)
 Su, Wanhua, 39, 146 (2B-C1)
 Subedi, Sanjeena, 23, 65 (1B-I1)
 Sultana, Nahid, 31, 109 (1C-D2)
 Sumanta, Basu, 44, 168 (2C-I4)
 Sun, Bingrui, 57, 232 (3B-C3)
 Sun, Guang, 44, 172 (2C-C1)
 Sun, Lei, 60, 61, 250, 251 (3C-C3)
 Sun, Wenqi, 31, 106 (1C-C5)
 Sun, Zheng, 31, 108 (1C-D1)
 Suresh, Krithika, 35, 127 (1D-C2)
 Sutradhar, Brajendra, 44 (2C-C1), 57 (3B-C3), 172 (2C-
 C1), 232 (3B-C3)
 Swaminathan, Balakumar, 25, 77 (1B-C2)
 Swartz, Tim, 21, 48, 190 (2D-I1)
 Tadavani, Pooyan Khajehpour, 51, 207 (2D-C3)
 Takahara, Glen, 25 (1B-C2), 26 (1B-C4), 77 (1B-C2), 84
 (1B-C4)
 Talbot, Denis, 41, 154 (2B-C4)
 Taleban, Julia, 60, 250 (3C-C3)
 Tan, Qier, 28, 95 (1C-C1)
 Tanha, Azadeh, 31, 109 (1C-D2)
 Tekougang, Thierry Chekouo, 46, 180 (2C-C4)
 Thabane, Lehana, 24 (1B-C1), 28 (1C-C1), 47 (2C-P),
 55 (3B-C1), 57 (3B-C3), 73, 75 (1B-C1), 96 (1C-
 C1), 182 (2C-P), 225 (3B-C1), 231, 232 (3B-
 C3)
 Thaddeus, Edens, 34, 122 (1D-I5)
 Thiess, Mary, 47, 187 (2C-P)
 Thompson, Angela, 25, 80 (1B-C3)
 Thompson, Mary, 27 (1C-I2), 40 (2B-C3), 43 (2C-I3), 52
 (2D-C4), 90 (1C-I2), 151 (2B-C3), 165 (2C-I3),
 210 (2D-C4)
 Thomson, David J., 57, 234, 235 (3B-C4)
 Thorbergsson, Liefur, 23, 69 (1B-I3)
 Thorlund, Kristian, 57, 232 (3B-C3)
 Tian, Lu, 39, 147 (2B-C1)
 Tian, Yanhua, 31, 109 (1C-D2)
 Tingting, Gou, 56, 231 (3B-C3)
 Tomal, Javed Hossain, 55, 226 (3B-C1)
 Tomanelli, Katrina, 40, 148 (2B-C2)
 Tomlinson, George, 30, 105 (1C-C5)
 Tong, Xin, 42, 161 (2C-I1)
 Torabi, Mahmoud, 37, 139 (2B-I1)
 Toth, Daniell, 23, 67 (1B-I2)
 Trépanier, Julie, 49, 197 (2D-I4)
 Troupé, Marylène, 29, 97 (1C-C2)
 Trudel, Marc, 47, 187 (2C-P)
 Tzavidis, Nikolas, 38, 142 (2B-I3)
 Vaillant, Jean, 29 (1C-C2), 35 (1D-C3), 97 (1C-C2), 130
 (1D-C3)
 Valmy, Larissa, 35, 130 (1D-C3)
 van Bommel, Christopher, 25, 80 (1B-C3)
 Van Dam-Bates, Paul, 31, 108 (1C-D1)
 van der Laan, Mark, 48, 192 (2D-I2)
 Vanniyasingam, Thuvaraha, 55, 225 (3B-C1)
 Varet, Hugo, 25, 77 (1B-C2)
 Variyath, Asokan Mulayath, 50, 202 (2D-C1)
 Vatathoor, Jayasankar, 50, 202 (2D-C1)
 Velarde, José Maria, 43, 163 (2C-I2)
 Verbeke, Geert, 56, 228 (3B-C2)
 Virji-Babul, Naznin, 31, 106 (1C-C5)
 Viveros-Aguilera, Roman, 50, 202 (2D-C1)
 Vrbik, Irene, 51, 208 (2D-C3)
 Waddell, Adrian, 40, 152 (2B-C3)
 Wadsworth, Laurie, 25, 80 (1B-C3)
 Wakaki, Hirofumi, 41, 154, 155 (2B-C4)

- Walsh, Michael, 57, 232 (3B-C3)
Walter, Stephen D., 35 (1D-C2), 57 (3B-C3), 127 (1D-C2), 232 (3B-C3)
Wang, Daoxing, 31, 109 (1C-D2)
Wang, Fei, 27, 89 (1C-I2)
Wang, Lily, 33, 120 (1D-I4)
Wang, Lu, 27, 89 (1C-I2)
Wang, Mei-Cheng, 48, 193 (2D-I2)
Wang, Weiqiang, 56, 229 (3B-C2)
Wang, Xiaogang (Steven), 30, 102 (1C-C4)
Wang, Xikui, 36, 133 (1D-C4)
Wang, Xu, 25, 80 (1B-C3)
Watier, François, 50, 204 (2D-C2)
Wei, Lee-Jen, 39, 147 (2B-C1)
Welch, David, 44, 167 (2C-I4)
Welch, William, 24 (1B-I4), 55 (3B-C1), 72 (1B-I4), 226 (3B-C1)
Weldon, Larry, 28, 93 (1C-I4)
Weng, Chengguo, 51, 205 (2D-C2)
White, Bethany, 45 (2C-C1), 49 (2D-I3), 173 (2C-C1), 195 (2D-I3)
Wiles, Janet, 37, 138 (2B-I1)
Willan, Andrew, 46, 180 (2C-C4)
Wilson, David Edward, 59, 245 (3C-C1)
Withanage, Niroshan, 29, 99 (1C-C3)
Wolfe, Patrick, 44, 167 (2C-I4)
Wolfson, Christina, 43, 163 (2C-I2)
Wolfson, David, 34 (1D-I5), 43 (2C-I2), 122 (1D-I5), 163 (2C-I2)
Wong, Augustine, 56, 228 (3B-C2)
Wong, Monica H.T., 47, 187 (2C-P)
Woodard, Roger, 49, 196 (2D-I3)
Woolford, Douglas, 46, 177 (2C-C3)
Wotton, Mike, 32, 112 (1C-P)
Wu, Beilei, 42, 158 (2B-C5)
Wu, Changbao, 40 (2B-C3), 53 (3A-A2), 58 (3C-A), 151 (2B-C3), 214 (3A-A2), 236 (3C-A)
Wu, Fan, 31, 108 (1C-D1)
Wu, Lang, 45, 172 (2C-C1)
Wu, Yan Yan, 56, 228 (3B-C2)
Xie, Feng, 56, 228 (3B-C2)
Xin, Xin, 40, 148 (2B-C2)
Xu, Fang, 55, 221 (3B-I4)
Xu, Jingxiong, 31, 109 (1C-D2)
Xu, Xiaojian, 36, 134 (1D-C5)
Xu, Ximing, 31, 109 (1C-D2)
Yamamura, Mariko, 41, 156 (2B-C4)
Yan, Ying, 44, 171 (2C-C1)
Yanagihara, Hirokazu, 41, 154–156 (2B-C4)
Yang, Haiyan, 47, 188 (2C-P)
Yao, Fang, 41, 157 (2B-C5)
Yao, Xue, 44, 171 (2C-C1)
Yasui, Yutaka, 24 (1B-C1), 43 (2C-I2), 74 (1B-C1), 163 (2C-I2)
Ye, Chenglin, 28 (1C-C1), 47 (2C-P), 96 (1C-C1), 182 (2C-P)
Yi, Grace Y., 27 (1C-I2), 39 (2B-I4), 44 (2C-C1), 56 (3B-C2), 89 (1C-I2), 144 (2B-I4), 171 (2C-C1), 229 (3B-C2)
Yi, Yanqing, 36, 132 (1D-C4)
You, Yong, 43, 165 (2C-I3)
Yu, Eugenia, 49, 195 (2D-I3)
Yu, Hao, 26 (1B-C4), 45 (2C-C3), 84 (1B-C4), 176 (2C-C3)
Yuan, Yan, 25, 81 (1B-C3)
Yuan, Yuan, 36, 132 (1D-C4)
Yue, Li Hua, 48, 188 (2C-P)
Yuhong, Wei, 48, 189 (2C-P)
Zaihra, Tasneem, 28, 94 (1C-C1)
Zamar, Ruben H., 55, 226 (3B-C1)
Zarepour, Mahmoud, 55, 221 (3B-I4)
Zhang, Chen, 52, 211 (2D-C4)
Zhang, Lin, 51, 210 (2D-C4)
Zhang, Shiyuan (William), 24 (1B-C1), 47 (2C-P), 73 (1B-C1), 182 (2C-P)
Zhang, Xuemao, 29, 100 (1C-C3)
Zhang, Xueshan, 31, 109 (1C-D2)
Zhang, Ying, 39, 145 (2B-I4)
Zhao, Lihui, 39, 147 (2B-C1)
Zhao, Tingting, 31, 108 (1C-D1)
Zhou, Michelle Qian, 27, 90 (1C-I2)
Zhu, Chao, 36, 132 (1D-C4)
Zhu, Mu, 39, 146 (2B-C1)
Zidek, Jim, 37, 137 (2B-V)
Zou, Jialin, 31, 109 (1C-D2)
Zou, Liwen, 36, 130 (1D-C3)
Zwiers, Francis, 26, 86 (1C-A)