

Société Statistical
statistique Society
du Canada of Canada

45th Annual Meeting
of the
Statistical Society of Canada

45^e Congrès annuel
de la
Société statistique du Canada

June 11 - June 14, 2017
11 juin au 14 juin 2017

University of Manitoba

Table of Contents • Table des matières

Table of Contents • Table des matières	1
Welcome to Winnipeg • Bienvenue à Winnipeg	2
Sponsors • Commanditaires	3
Exhibitors • Exposants	3
Organizers • Organisateurs	4
General Information • Informations générales	5
The Conference • Le congrès	8
Social Events • Activités sociales	9
Committees and Meetings • Comités et réunions	12
Social and Information Events • Événements sociaux et informatifs	16
Program • Programme	17
Abstracts • Résumés	55
Author List • Liste des auteurs	247

Welcome to Winnipeg • Bienvenue à Winnipeg

It is with great pleasure that the members of the Department of Statistics at the University of Manitoba welcome you to the 45th Annual Meeting of the Statistical Society of Canada.

The Department of Statistics currently hosts 17 faculty members (including twelve research faculty and five full-time teaching instructors). It was created in 1967 with a teaching staff of 10 members, and is celebrating its 50th Anniversary this year. This being said, the teaching of Statistics at the University of Manitoba began in the Department of Mathematics in the Faculty of Arts as early as 1912-13, more than 100 years ago!

The University of Manitoba is western Canada's first university. Founded in 1877 and located in the heart of the country, it is the region's largest and only research intensive university offering well over 100 degrees, diplomas and certificates, including professional disciplines such as medicine, law, and engineering. With over 30,000 students and 8,900 faculty and staff, the University of Manitoba plays a key role in the social, cultural, and economic well-being of its community and the world.

Winnipeg has a rich cultural environment, including symphony, opera, dance, theatre and ethnic festivals. It is also host to many museums, the most prominent certainly being the brand new Museum of Human Rights. The region provides ample opportunities for outdoor recreation, the most popular destination perhaps being the sandy beaches and dunes at Grand Beach Provincial Park, on Lake Winnipeg.

For more information to help you plan your trip, please visit the conference local attractions page at ssc.ca/en/meeting/annual/2017, as well as www.tourismwinnipeg.com and www.travelmanitoba.com.

C'est avec un très grand plaisir que les membres du département de statistique de l'Université du Manitoba vous accueillent au 45^e congrès annuel de la Société statistique du Canada.

Le département de statistique de l'université, hôte du congrès annuel de la Société, compte actuellement 17 membres (incluant douze professeurs et cinq instructeurs à plein temps). Il a été créé en 1967, avec un corps enseignant comptant 10 membres, et célèbre cette année son 50^e anniversaire. Ceci étant dit, l'enseignement de la statistique à l'Université du Manitoba a débuté en 1912-13, au département de mathématiques dans la Faculté des Arts, il y a de cela plus de 100 ans!

L'Université du Manitoba of Manitoba, fondée en 1877, a été la première université de l'ouest canadien. Elle est située à Winnipeg, en plein coeur du pays, et compte plus de 30 000 étudiants et 8 900 chercheurs, enseignants et employés de soutien. Elle est de loin la plus grande université de la région et joue un rôle crucial dans le développement culturel, social et économique de sa communauté et du reste du monde.

Winnipeg offre un environnement culturellement très riche, incluant danse, symphonies, opéra, théâtre et festivals de toute sorte. La ville abrite aussi plusieurs musées, dont le tout nouveau musée canadien pour les droits de la personne. Finalement, de nombreuses activités de plein air sont offertes dans la région, la destination la plus populaire étant peut-être le parc provincial de Grand Beach sur le Lac Winnipeg, avec ses impressionnantes plages et ses dunes de sable fin.

Pour vous aider à planifier votre séjour, veuillez consulter la page des attractions locales du congrès à ssc.ca/fr/congres/annuel/2017, ainsi que www.tourismwinnipeg.com et www.travelmanitoba.com.

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 les généreuses contributions ont rendu possible la tenue de ce congrès :

- Office of the President, University of Manitoba
- Faculty of Science, University of Manitoba
- Department of Statistics, University of Manitoba
- Fields Institute
- Pacific Institute for the Mathematical Sciences
- Centre de recherches mathématiques
- Canadian Statistical Sciences Institute /
Institut canadien des sciences statistiques

Exhibitors • Exposants

The exhibitors, providing information and displays for examination and purchase, will be located in the Atrium of the Engineering Information and Technology Complex (EITC) on the Fort Garry Campus of the University of Manitoba, a short distance away from the rooms where all scientific sessions will be held, and in the same location as where all poster sessions will take place. The exhibitors will be available Monday, June 12 and Tuesday, June 13 from 9:00 am to 4:00pm.

Les kiosques des exposants, où vous trouverez différentes informations et pourrez bouquiner ou acheter, seront situés dans l'Atrium de l'Engineering Information and Technology Complex (EITC) sur le campus Fort Garry de l'Université du Manitoba, tout juste à côté des locaux où toutes les séances de présentations scientifiques auront lieu et où les séances de présentation par affiches se tiendront. Les exposants seront présents les lundi et mardi, 12 et 13 juin, de 9h à 16h.

- CRC Press / Taylor & Francis www.crcpress.com
- Wiley ca.wiley.com/WileyCDA
- Department of Epidemiology, Biostatistics & Occupational Health, McGill University www.mcgill.ca/epi-biostat-occh
- Canadian Statistical Sciences Institute / Institut canadien des sciences statistiques www.canssi.ca / www.incass.ca
- Save the Stats Book – by McDougall Scientific savethestatsbook.com

Organizers • Organisateurs

Local Arrangements Committee • Comité des arrangements locaux

- | | |
|---|------------------|
| • Alexandre Leblanc (Chair • Président) | • Margaret Smith |
| • Saman Muthukumarana | • Elif Acar |
| • Katherine Davies | • Saumen Mandal |
| • Carrie Paquette | • Po Yang |
| • Liqun Wang | • John Brewster |
| • Brad Johnson | • Jenna Tichon |
| • Mohammad Jafari Jozani | • Xikui Wang |
| • Aerambamoorthy Thavaneswaran | • Andrew Morris |
| • Yuliya Martsynyuk | • Zeny Mateo |

It is impossible to organize an event of the size of the Annual Meeting of the SSC without the help of several individuals and organizations. The local arrangements committee would like to thank all those who helped pull this event together.

Among other responsibilities, our student volunteers have handled AV support and manned the registration desk. Maria Vieira and Amartya Chatterjee of Conference Services at the University of Manitoba have been helpful and supportive over the past two years. At the SSC Office, Larysa Valachko and, especially, Miaclaire Woodland have provided crucial support for many aspects, including registration and exhibitors. The SSC Meetings Coordinator Changbao Wu and other SSC executive members, and John Yuen and other previous local arrangements chairs all shared their experience, offered useful advice and answered our numerous questions. Finally, Hugh Chipman, Jason Loeppky and Kevin Simpson managed all electronic services related to the meeting and put together the PDF version of the conference program.

Il est impossible d'organiser un événement de l'envergure du congrès annuel de la SSC sans l'aide de nombreux individus et organismes. Le comité des arrangements locaux est très reconnaissant à tous ceux qui ont aidé à mettre sur pied cet événement.

Les étudiants bénévoles qui ont contribué, entre autres, au soutien audiovisuel et au bureau des inscriptions. Maria Vieira et Amartya Chatterjee des Services de conférence de l'Université du Manitoba nous ont donné leur aide et leur support au cours des deux dernières années. Au bureau de la SSC, Larysa Valachko et, tout spécialement, Miaclaire Woodland ont fourni un appui essentiel pour divers aspects, dont l'inscription et les exposants. Le coordonnateur des congrès Changbao Wu et les autres membres de l'exécutif de la SSC, ainsi que John Yuen et les autres anciens présidents des comités des arrangements locaux ont généreusement partagé leurs expériences, offert leurs recommandations et répondu à nos nombreuses questions. Finalement, Hugh Chipman, Jason Loeppky et Kevin Simpson ont géré tous les services électroniques reliés au congrès et ont préparé la version PDF du programme.

Program Committee • Comité du programme

- Erica E.M. Moodie (Chair • Présidente) *McGill University*
- Grace Y. Yi *University of Waterloo*, Biostatistics - Biostatistique
- Neil Madras *York University*, Probability - Probabilité
- François Bellavance *HEC Montréal*, Business and Industrial Statistics- Industrielle et Gestion
- Matthias Schonlau *University of Waterloo*, Survey Methods - Enquête
- Jim Stallard *University of Calgary*, Statistical Education - Éducation en statistique
- Sheldon Lin *University of Toronto*, Actuarial Science - Science Actuarielle

General Information • Informations générales

Directions • Emplacement

Campus maps are available on the conference website, in the conference bag and on the back cover of this program. An interactive version of the campus map is also accessible on the **UManitoba** Mobile App that can be downloaded for free from the Apple App store and from Google Play.

Those staying at any of the conference hotels can refer to the Accommodations page of the conference at ssc.ca/en/meeting/annual/2017

for appropriate bus routes and distances to the Fort Garry campus.

Des cartes du campus sont accessibles sur le site web du congrès et dans ce programme. Une version imprimée de la carte du campus vous sera aussi remise lors de l'inscription. Finalement, une version interactive de la carte du campus est disponible sur l'application mobile **UManitoba** qui peut être installée gratuitement à partir du Apple App store et de Google Play.

Ceux qui logent à l'un des hôtels associés au congrès peuvent se référer à la page du congrès dédiée à l'hébergement à ssc.ca/fr/congres/annuel/2017 pour obtenir de l'information à propos des circuits d'autobus et des distances approximatives entre les hôtels et le campus Fort Garry.

Registration • Inscription

Registered participants can pick up their badges at the registration desk at the times and locations indicated below. Walk-ins should register online and present their receipt at the registration desk.

- Sunday, June 11: 8:00 am - 5:30 pm
EITC Atrium
- Sunday, June 11: 5:45 pm - 8:00 pm
Marshall McLuhan Hall in University Centre
- Monday, June 12: 7:30 am - 5:00 pm
EITC Atrium
- Tuesday, June 13: 8:15 am - 5:00 pm
EITC Atrium
- Wednesday, June 14: 8:15 am - 4:00 pm
EITC Atrium

Les participants déjà inscrits peuvent venir chercher leur matériel de congrès à l'endroit et aux heures indiqués ci-dessous. Ceux qui veulent s'inscrire au moment du congrès devraient le faire en ligne et présenter leur reçu au kiosque d'inscription.

- Dimanche 11 juin : 8h - 17h30
Atrium de l'EITC
- Dimanche 11 juin : 17h45 - 20h
Marshall McLuhan Hall au University Centre
- Lundi 12 juin : 7h30 - 17h
Atrium de l'EITC
- Mardi 13 juin : 8h15 - 17h
Atrium de l'EITC
- Mercredi 14 juin : 8h15 - 16h
Atrium de l'EITC

Parking on Campus • Stationnement sur le campus

Updated parking information (including an interactive map) can be found online at umanitoba.ca/campus/parking/visitor. Note that parking on campus is generally free on weekends and after 4:30 pm on weekdays.

De l'information actualisée sur le stationnement pour les visiteurs du campus Fort Garry (incluant une carte interactive) est disponible en ligne à umanitoba.ca/campus/parking/visitor. Veuillez noter que le stationnement est généralement gratuit sur le campus les soirs de semaine après 16h30 et la fin de semaine.

Campus Security • Sécurité sur le campus

- Fire/Police/Ambulance: 911
- Campus Security Services:
 - 555 on any university phone
 - #555 for MTS or Rogers Wireless customers
 - any blue or red emergency phone on campus
 - 204-474-9312 from all other phones
- Winnipeg Police Service: 204-986-6222
- Pompiers/Police/Ambulance : 911
- Force policière du campus :
 - 555 sur tout appareil de l'université
 - #555 pour les clients de MTS et Rogers
 - tout téléphone d'urgence rouge ou bleu sur le campus
 - 204-474-9312 sur tout autre téléphone
- Service de police de Winnipeg : 204-986-6222

Internet Access • Accès internet

Wireless access is available for all attendees. The wireless network name is **uofm-guest**. Once you are connected to the network, open a browser and you should be redirected to a User Agreement Policy page. After clicking on the “I Agree” button located at the bottom of the page, you will be connected to the internet.

L'accès sans fil est disponible pour tous les participants. Le nom du réseau à utiliser est **uofm-guest**. Après vous être connecté sur ce réseau, ouvrez une fenêtre de navigateur : vous serez redirigé vers une page d'accord utilisateur. Cliquez sur le bouton « I Agree » en bas de cette page pour vous connecter à l'internet.

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

Lunch Monday to Wednesday is provided with registration and is served at Pembina Hall. You just need to show your conference badge to enter the facility. Lunch will be served from 11:30 am to 2:00 pm so people who wish to have an early or late lunch can do so. Lunch-time committee meetings will occur in the scheduled rooms: participants are invited to pick up boxed lunches at the registration desk in the EITC Atrium.

Other venues that provide food on a cash basis on campus include

- the University Centre (UC) Food Court, opening hours are available at umanitoba.ca/campus/food.html,
- Subway in University Centre Food Court, Monday to Friday, 10:00 am - 3:00 pm,
- Tim Horton's on UC second floor, Monday to Friday, 7:30 am - 3:30 pm,
- Starbucks on UC first floor, Monday to Friday, 8:00 am - 3:00 pm,
- Degrees on UC third floor, Monday to Friday, 8:00 am - 5:00 pm.

Note that dinner options are very limited on campus, but that Winnipeg is known for hosting many excellent restaurants. A list of restaurants recommended by our friends and colleagues is available online at ssc.ca/en/meeting/annual/2017.

Le repas de midi est fourni avec l'inscription du lundi au mercredi ; il sera servi à Pembina Hall. Il vous suffira de montrer votre badge du congrès pour entrer sur les lieux. Le lunch sera servi de 11h30 à 14h, donnant l'option aux congressistes de manger un peu plus tôt ou un peu plus tard. Les diners-réunions de comités se tiendront dans les salles prévues au programme : les participants sont invités à aller chercher une boîte à lunch au comptoir des inscriptions dans l'Atrium de l'EITC.

Vous pouvez également acheter à manger ailleurs sur le campus, incluant

- le University Centre (UC) Food Court, les heures d'ouverture étant disponibles à umanitoba.ca/campus/food.html,
- Subway au University Centre Food Court, lundi à vendredi, 10h - 15h,
- Tim Horton's au deuxième étage de l'UC, lundi à vendredi, 7h30 - 15h30,
- Starbucks au premier étage de l'UC, lundi à vendredi, 8h - 15h,
- Degrees au troisième étage de l'UC, lundi à vendredi 8h - 17h.

Veuillez noter que les options sont limitées sur le campus quant à l'offre de souper, mais que Winnipeg est reconnue pour ces nombreux excellents restaurants. Une liste de restaurants recommandés par nos amis et collègues est disponible en ligne à ssc.ca/fr/congres/annuel/2017.

Athletics Facilities • Installations sportives

The Active Living Centre is the main fitness facility at the University of Manitoba. It offers a fitness centre (cardio room, weight and strength training equipment), an indoor running track, a pool, courts for squash/racketball and gyms. Full access day passes are available for \$13. More details can be found at umanitoba.ca/activelivingcentre.





L'Active Living Centre est le centre d'entraînement principal de l'université. Il comprend un gymnase, une salle de cardio (offrant l'équipement pour l'entraînement en force et en puissance), une piste d'athlétisme intérieure, une piscine, des courts de squash/racketball et des gymnases. Des passes de jour sont disponibles au prix de 13 \$. Pour plus de détails, visitez

umanitoba.ca/activelivingcentre.





The Conference • Le congrès

Language • Langue

An important feature of our meetings is the presentation of the abstracts and of the plenary session visual aids in both official languages. This translation was once again very ably carried out under the supervision of the Bilingualism Committee (chaired by Geneviève Lefebvre) by the translators Caroline Gras, Catherine Cox, Caroline Pelletier, Michelle Blaquièrre and Caroline Petit-Turcotte.

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

Une caractéristique importante de nos congrès est la présentation des résumés et des supports visuels des sessions plénières dans les deux langues officielles. Cette traduction a été encore une fois très habilement menée sous la supervision du Comité du bilinguisme (présidé par Geneviève Lefebvre) par les traductrices Caroline Gras, Catherine Cox, Caroline Pelletier, Michelle Blaquièrre et Caroline Petit-Turcotte.

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

Rooms • Salles

Exhibits and poster presentations will take place in the Atrium of the Engineering Information and Technology Complex (EITC) on Monday and Tuesday. Coffee breaks will take place in the EITC Atrium (ground level) and in the hallways leading to the rooms used for the scientific sessions in EITC (ground and basement levels). Plenary sessions on Monday and Tuesday mornings will take place in room 210 of the University Centre. These will be the only scientific sessions not taking place in EITC.

Les exposants seront présents dans l'Atrium de l'Engineering Information and Technology Complex (EITC) lundi et mardi. Les présentations par affichage se feront aussi dans l'Atrium de l'EITC les mêmes jours. Les pauses-café se tiendront dans les corridors menant aux salles de l'EITC où les sessions scientifiques auront lieu (niveaux sous-sol et rez-de-chaussée). Les séances plénières des lundi et mardi matins auront lieu dans la salle 210 du University Centre. Ces séances seront les seules sessions scientifiques n'ayant pas lieu dans l'EITC.

Poster Sessions • Séances d'affichage

Contributed posters and Case Study posters will be displayed in the EITC Atrium, with Case Study posters being displayed Monday afternoon and contributed research posters appearing on Tuesday afternoon. In both cases, posters will be displayed between noon and 5:30 pm, the authors being with their posters from 1:30 pm until the end of coffee break at 3:30 pm.

Les séances d'affiches scientifiques et d'études de cas se feront dans l'Atrium de l'EITC, les affiches d'étude de cas étant exposées le lundi après-midi et les affiches libres exposées le mardi après-midi. Dans les deux cas, les affiches seront exposées de midi à 17h30, les auteurs devant être présents avec leur affiche de 13h30 jusqu'à la fin de la pause-café à 15h30.

Workshops • Ateliers

Sunday, June 11, 9:00 am - 4:00 pm • Dimanche 11 juin, 9h - 16h

Engineering Information and Technology Complex (EITC)

Workshops organized by the sections will be held Sunday 9:00 am to 4:00 pm in EITC. Coffee breaks will take place in the hallway, outside the rooms where the workshops are held, from 10:15 to 10:45 am and 2:30 to 3:00 pm. The lunch break will be between noon and 1:00 pm in the same locations as coffee breaks.

Les ateliers organisés par les groupes auront lieu le dimanche de 9h00 à 16h00 dans l'EITC. Les pauses-café se feront dans les corridors, juste à l'extérieur des salles où les ateliers auront lieu, de 10h15 à 10h45 et de 14h30 à 15h00. Le lunch sera entre midi et 13h00 au même endroit que les pauses-café.

Other Meetings • Autres réunions

NSERC – Updates and 2017 Competition Results:

Sunday 5:00 - 6:00 pm, in E3 270 (EITC)

CRSNG – mise-à-jour et résultats du concours de 2017 : dimanche 17h - 18h, salle E3 270 (EITC)

Committee meetings: Lunchtime business meetings will occur in EITC, in the same rooms that are used for the scientific sessions. Business meeting participants should pick up their boxed lunches in the Atrium of EITC, close to the registration desk.

Réunions de comités : Les réunions d'affaires ayant lieu à l'heure du lunch se tiendront dans l'EITC, dans les mêmes salles où les séances de présentations auront lieu. Les participants à ces réunions sont invités à prendre une boîte à lunch dans l'ATRIUM de l'EITC, tout près du comptoir des inscriptions avant de se rendre aux réunions.

NSERC Discovery Grant Application Workshop:

Monday 12:30 - 1:30 pm, in E2 110 (EITC)

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.

Atelier subventions du CRSNG : lundi 12h30 - 13h30, salle E2 110 (EITC)

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 présentera certains conseils pour la préparation d'une demande de subvention à la découverte. À la fin de l'atelier, les participants seront invités à poser leurs questions.

SSC Annual General Meeting: Monday 5:00 - 6:00 pm, in E3 270 (EITC)

Assemblée générale annuelle : lundi 17h - 18h, salle E3 270 (EITC)

Section AGM's: Tuesday 5:00 - 6:00 pm, in various rooms in EITC

AGM des groupes : mardi 17h - 18h, dans diverses salles de l'EITC

Social Events • Activités sociales

Welcome Reception • Réception de bienvenue

Sunday, June 11, 6:00 - 8:00 pm • Dimanche 11 juin, 18h00 - 20h00 . Marshall McLuhan Hall (UC)

The welcome reception will be held in Marshall McLuhan Hall, in University Centre. All conference attendees are welcome to join us to share a few drinks and excellent food in good company.

La réception de bienvenue aura lieu au Marshall McLuhan Hall, dans le University Centre. Tous les congressistes sont bienvenus et sont invités à venir partager quelques verres et d'excellentes bouchées en bonne compagnie.

Barbeque • Barbecue

Monday, June 12, 6:30 - 8:30 pm • Lundi 12 juin, 18h30 - 20h30 Marshall McLuhan Hall (UC)

The student BBQ is free for undergraduate and graduate students; registration is required. The BBQ will take place at Marshall McLuhan Hall, in University Centre.

Le barbecue étudiant est gratuit pour les étudiants du premier cycle et des cycles supérieurs; l'inscription est requise. Le BBQ aura lieu au Marshall McLuhan Hall, dans le University Centre.

Banquet

Tuesday, June 13, 6:15 pm • Mardi 13 juin, 18h15 Club Regent Event Centre

The banquet will be held at the Club Regent Event Centre, with cocktails and a cash bar at 6:15 pm and dinner service beginning at 7:00 pm. All conference participants who have selected a meal for the banquet will find a banquet ticket in their registration envelope. You are asked to place this on the table in front of you. If you have a ticket but do not wish to attend the banquet, please return it to the registration desk by noon on Tuesday, June 13 - this will help us better plan for attendance at the banquet.

Buses are arranged to transport participants from the conference location, just outside EITC, to the banquet location leaving at 5:30, 5:45, 6:00 and 6:15 pm. After the banquet, buses will return participants to the conference location or to one of the conference hotels (Fairmont and Fort Garry located downtown – Holiday Inn, Best Western, Canad'Inns and Sheraton located close to campus). It is expected that buses will leave the banquet location between 10:00 and 10:30 pm.

Le banquet se tiendra au Club Regent Event Centre, avec cocktails et bar payant à partir de 18h15, suivi du souper servi à table à partir de 19h. Les participants au congrès qui ont choisi de prendre part au banquet trouveront un ticket repas dans leur enveloppe d'inscription. Nous vous demanderons de le placer devant vous à la table. Si vous avez un ticket mais que vous ne souhaitez plus participer au banquet, veuillez remettre votre ticket au bureau des inscriptions, au plus tard mardi 13 juin à midi.

Des autobus ont été loués pour transporter les participants du site du congrès, devant l'EITC, jusqu'au Club Regent. Des départs se feront à 17h30, 17h45, 18h et 18h15. À la fin du banquet, les autobus ramèneront les participants au site du congrès ou à l'un des hôtels associés au congrès (Fairmont et Fort Garry situés au centre ville – Holiday Inn, Best Western, Canad'Inns et Sheraton situés près du campus. Il est prévu que les autobus quitteront le site du banquet entre 22h et 22h30.

Other Social Events • Autres événements sociaux

Waterloo Alumni Reception

Monday, June 12, 5:30 - 7:30 pm, at The Hub, on the third floor of University Centre. This event has limited seating. Register online at

uwaterloo.ca/math/events

Réception des anciens de Waterloo

lundi 12 juin, 17h30 - 19h30, à The Hub, au 3e étage du University Centre. Les places sont limitées. Veuillez vous inscrire en ligne à

uwaterloo.ca/math/events

Informal Gathering of New Investigators

Monday, June 12, 7:00 pm, at The Hub, on the third
floor of University Centre

Repas des nouveaux chercheurs

lundi 12 juin à 19h00, à The Hub, au 3e étage du
University Centre

Committees and Meetings • Comités et réunions

Saturday June 10

Samedi 10 juin

11:00-17:00

MH 312 (MH)

Executive Committee 1/Comité exécutif 1

Sunday June 11

Dimanche 11 juin

09:30-11:00

E1 270 (EITC)

Ad hoc Fundraising Committee/Comité ad hoc de collecte de fonds

11:00-12:00

E1 270 (EITC)

Board of Directors 1 - Board Members Only/Conseil d'administration 1 - membres du Conseil seulement

12:00-17:00

E1 270 (EITC)

Board of Directors 1 - Board Members and Invited Guests/Conseil d'administration 1 - membres du Conseil et invités

14:30-17:00

MH 312 (MH)

Department Heads Meeting/Rencontre des directeurs de départements

17:00-18:00

E3 270 (EITC)

NSERC – Updates and 2017 Competition Results/CRSNG – mise-à-jour et résultats du concours de 2017

Monday June 12

Lundi 12 juin

12:00-13:30

E2 165 (EITC)

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

12:00-13:30

E2 125 (EITC)

Actuarial Science Section Executive Committee 1/Comité exécutif du groupe d'actuariat 1

12:00-13:30

E2 130 (EITC)

Biostatistics Section Executive Committee 1/Comité exécutif du groupe de biostatistique 1

12:00-13:30

E2 155 (EITC)

Finance Committee and Committee of Section Treasurers/Comité des finances

12:00-13:30

E2 150 (EITC)

Business and Industrial Statistics Section Executive Committee 1/Comité exécutif du groupe de statistique industrielle et de gestion 1

12:00-13:30	E2 105 (EITC)
Publications Committee/Comité des publications	

12:00-13:30	E2 160 (EITC)
Probability Section Executive Committee 1/Comité exécutif du groupe de probabilité 1	

12:00-18:00	E1 270 (EITC)
Award for Case Studies and Data Analysis Committee/Comité du prix pour les études de cas et l'analyse de données	

12:30-13:30	E2 110 (EITC)
How to Prepare an NSERC Discovery Grant Application/Assistance pour les demandes de subvention à la découverte du CRSNG	

17:00-18:00	E3 270 (EITC)
SSC Annual General Meeting/Assemblée générale annuelle de la SSC	

Tuesday June 13	Mardi 13 juin
------------------------	----------------------

09:00-18:00	E2 525 (EITC)
Student Research Presentation Award Committee/Comité du prix pour la présentation de recherche étudiante	

12:00-13:30	E2 155 (EITC)
Office Committee/Comité du Bureau	

12:00-13:30	MH 312 (MH)
CANSSI Health Sciences Collaborating Centres/INCASS Centres de collaboration des sciences de la santé	

12:00-13:30	E2 165 (EITC)
Statistics Education Committee/Comité d'éducation en statistique	

12:00-13:30	E2 105 (EITC)
Accreditation General Meeting/Assemblée générale des membres accrédités	

12:00-13:30	E2 125 (EITC)
Census at School Committee/Comité de recensement à l'école	

12:00-13:30	E2 110 (EITC)
CJS Editorial Board/Comité de rédaction de la RCS	

12:00-13:30	E2 150 (EITC)
Committee on Membership/Comité de recrutement	

12:00-13:30	E2 130 (EITC)
Committee on New Investigators /Comité des nouveaux chercheurs	

12:00-13:30	E2 160 (EITC)
Public Relations Committee/Comité des relations publiques	

17:00-18:00	E2 165 (EITC)
Statistical Education Section General Meeting/Groupe d'éducation en statistique assemblée générale	

17:00-18:00	E2 130 (EITC)
Survey Methods Section General Meeting/Groupe de méthodes d'enquête assemblée générale	

17:00-18:00	E2 125 (EITC)
Actuarial Science Section General Meeting/Groupe d'actuariat assemblée générale	

17:00-18:00	E2 105 (EITC)
Biostatistics Section General Meeting/Groupe de biostatistique assemblée générale	

17:00-18:00	E2 150 (EITC)
Business and Industrial Statistics Section General Meeting/Groupe de statistique industrielle et de gestion assemblée générale	

17:00-18:00	E2 160 (EITC)
Probability Section General Meeting/Groupe de probabilité assemblée générale	

Wednesday June 14	Mercredi 14 juin
--------------------------	-------------------------

12:00-13:30	E2 130 (EITC)
Ad hoc Financial Procedures Committee/Comité ad hoc des procédures financières	

12:00-13:30	E2 160 (EITC)
Probability Section Executive Committee 2/Comité exécutif du groupe de probabilité 2	

12:00-13:30	E2 155 (EITC)
Student and Recent Graduate Committee/Comité des étudiants et diplômés récents	

12:00-13:30	E2 165 (EITC)
Statistical Education Section Executive Committee 2/Comité exécutif du groupe d'éducation en statistique 2	

12:00-13:30	E2 125 (EITC)
Actuarial Science Section Executive Committee 2/Comité exécutif du groupe d'actuariat 2	

12:00-13:30 Liaison Committee/Comité Liaison	E2 150 (EITC)
<hr/>	
12:00-13:30 Biostatistics Section Executive Committee 2/Comité exécutif du groupe de biostatistique 2	E2 105 (EITC)
<hr/>	
12:00-13:30 Program Committee/Comité du programme	E2 110 (EITC)
<hr/>	
16:45-18:00 Board of Directors 2/Conseil d'administration 2	MH 312 (MH)
<hr/>	
18:00-19:00 Executive Committee 2/Comité exécutif 2	MH 312 (MH)

Social and Information Events • Événements sociaux et informatifs

Sunday June 11

Dimanche 11 juin

18:00-20:00

Marshall McLuhan Hall (UC)

Welcoming Reception/Réception de bienvenue

Monday June 12

Lundi 12 juin

09:50-10:20

Hallways of EITC (EITC)

Coffee Break/Pause-café

15:00-15:30

Hallways of EITC (EITC)

Coffee Break/Pause-café

18:30-20:30

Marshall McLuhan Hall (UC)

Monday Evening Student BBQ/BBQ étudiant

19:00-22:00

The Hub (UC)

New Investigators Social Gathering/Rassemblement social des nouveaux chercheurs

Tuesday June 13

Mardi 13 juin

09:50-10:20

Hallways of EITC (EITC)

Coffee Break/Pause-café

15:00-15:30

Hallways of EITC (EITC)

Coffee Break/Pause-café

18:15-18:15

Event Centre (Club Regent)

Banquet/Banquet

Wednesday June 14

Mercredi 14 juin

09:50-10:20

Hallways of EITC (EITC)

Coffee Break/Pause-café

14:30-15:00

Hallways of EITC (EITC)

Coffee Break/Pause-café

Program • Programme

Sunday June 11

Dimanche 11 juin

09:00-16:00

E2 130 (EITC)

Survey Methods Workshop

Atelier du Groupe des méthodes d'enquête

09:00-16:00

Mark Stinner (Statistics Canada)

Disclosure Control Methods / Méthodes de contrôle de la divulgation



09:00-16:00

E3 270 (EITC)

Biostatistics Workshop

Atelier du Groupe de biostatistique

09:00-16:00

Stef van Buuren (Utrecht University)

Handling Missing Data in R with MICE / Gestion de données manquantes dans R à l'aide de MICE



09:00-16:00

E2 125 (EITC)

Statistical Education Workshop

Atelier du Groupe d'éducation en statistique

09:00-16:00

Nathan Tintle (Dordt College)

Teaching Introductory Statistics with Simulation Based-Inference / Enseignement d'un cours d'introduction à la statistique avec de l'inférence basée sur les simulations



09:00-16:00

E2 105 (EITC)

Business and Industrial Statistics Workshop

Atelier du Groupe de statistique industrielle et de gestion

09:00-16:00

Nick Pizzi (InfoMagnetis Technologies Corporation)

An Introduction to Data Science: Graph Databases and Entity Analytics / Introduction à la science des données : bases de données orientées graphes et analytique des entités



Monday June 12**Lundi 12 juin****08:25-08:40****UC 210 (UC)****Inaugural session****Session inaugurale****08:25-08:30 Jack Gambino** (Statistics Canada)

Welcome / Bienvenue

 **EF****08:30-08:40 Alexandre Leblanc** (University of Manitoba)

Welcome / Bienvenue

 **EF****08:40-09:45****Invited / Sur invitation** (abstract/résumé 55)**UC 210 (UC)****SSC Presidential Invited Address****Allocution de l'invité du président de la SSC**

Organizer/Responsable: Jack Gambino

08:40-09:45 Don A. Dillman (Washington State University)

The Challenge of Creating Data Collection Methods that are Neither Too Far Ahead nor Behind our Survey Respondents / Le défi de construire une méthode de collecte de données qui n'est ni en avance, ni en retard pour les répondants

 **EF****10:20-11:50****Invited / Sur invitation** (abstract/résumé 56)**E3 270 (EITC)****Analysis of Complex Traits in Families and Populations****Analyse des caractères complexes dans les familles et populations**

Chair/Président: Jinko Graham

Organizer/Responsable: Jinko Graham

Sponsor/Commanditaires: Biostatistics Section / Groupe de biostatistique

10:20-10:50 J. Concepcion Loredó-Osti (Memorial University of Newfoundland)

The Analysis of Longitudinal Multivariate (Discrete or Continuous) Traits under Irregular Time Measurements / Analyse longitudinale de traits multivariés (discrets ou continus) avec des temps de mesure irréguliers

 **E****10:50-11:20 Kelly M. Burkett** (University of Ottawa)

An Ancestral Tree-Based Approach to Detect Rare and Common Variants / Approche arborescente ancestrale pour détecter les variants rares et communs

 **E****11:20-11:50 Fabrice Larribe** (Université du Québec à Montréal)

Mapping Complex Traits, Rare Variants and Interaction via the Coalescent Process with Recombination / Cartographie de traits complexes, de variants rares et d'interaction par le processus de coalescence avec recombinaison

 **E****10:20-11:50****Invited / Sur invitation** (abstract/résumé 58)**E2 125 (EITC)****Statistical Consulting: Bridging Innovation and Practice****Consultation statistique : entre innovation et pratique**

Chair/Président: Sanjeena Dang

Organizer/Responsable: Sanjeena Dang

Sponsor/Commanditaires: Business and Industrial Statistics Section / Groupe de statistique industrielle et de gestion

10:20-11:05 Jeffrey A. Bakal (Alberta Health Services)

How Best to Bridge the Gap between Statistics and Medical Practice / Comment combler le fossé entre statistique et pratique médicale

 **E**

11:05-11:50 **Paul D. McNicholas** (McMaster University)
Some Experiences From an Industry Collaboration / Expériences tirées d'une collaboration en milieu industriel  

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 59) **E2 130 (EITC)**



Confidentiality



Confidentialité



Chair/Président: Karla Fox

Organizer/Responsable: Karla Fox

Sponsor/Commanditaires: Survey Section / Groupe des méthodes d'enquête

10:20-10:50 **Natalie Shlomo** (University of Manchester)
Statistical Disclosure Control for Statistical Outputs: Where Do We Go From Here? /
Contrôle de la divulgation de résultats statistiques : vers quoi allons-nous ?  

10:50-11:20 **Anne-Sophie Charest** (Université Laval)
What is Synthetic about Synthetic Datasets? / Qu'y a-t-il de synthétique dans un jeu de données synthétiques ?  

11:20-11:50 **Mark Stinner** (Statistics Canada)
Disclosure Control and Random Tabular Adjustment / Contrôle de la divulgation et ajustement tabulaire aléatoire  



10:20-11:50 **Invited / Sur invitation** (abstract/résumé 61) **E2 105 (EITC)**



Spatial and Spatiotemporal Statistics: Novel Methods and Applications



Statistique spatiale et spatiotemporelle : nouvelles méthodes et applications



Chair/Président: Xin (Cindy) Feng

Organizer/Responsable: Xin (Cindy) Feng

10:20-10:42 **Farouk Nathoo** (University of Victoria), **Yin Song** (University of Victoria)
A Potts Mixture Spatiotemporal Joint Model for Combined MEG and EEG Data / Modèle de mélange conjoint spatio-temporel de Potts pour données MEG et EEG combinées  

10:42-11:04 **Andrew Lawson** (Medical University of South Carolina), **Ray Boaz** (Medical University of South Carolina)
Spatially Structured Sparseness in Bayesian Spatial Health Modeling / Dispersion spatiale des données en modélisation bayésienne spatiale en santé  

11:04-11:26 **Patrick E. Brown** (St. Michael's Hospital, Toronto), **Ofir Hariri** (University of Toronto), **Alisha Albert-Green** (University of Toronto), **Jamie Stafford** (University of Toronto), **Paul Nguyen** (Queen's University), **Lennon Li** (Public Health Ontario)
Continuous Models for Aggregated Spatio-Temporal Data in Practice / Modèles continus pratiques pour données spatiotemporelles agrégées  

11:26-11:50 **Longhai Li** (University of Saskatchewan), **Xin (Cindy) Feng** (University of Saskatchewan), **Shi Qiu** (International Road Dynamics)
Estimating Cross-validatory Predictive P-values with Integrated Importance Sampling for Disease Mapping Models / Estimation de valeurs p prédictives par validation croisée avec échantillonnage préférentiel intégré pour les modèles de cartographie des maladies  

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 64) **E2 110 (EITC)**

Statistical Inference for Complex Dynamical Systems

Inférence statistique pour systèmes dynamiques complexes

Chair/Président: Jiguo Cao

Organizer/Responsable: Jiguo Cao

10:20-10:50	James O. Ramsay (McGill University), Michelle Carey (University College Dublin), Juan Li (McGill University) From Brain to Hand to Statistics with Dynamic Smoothing / Du cerveau à la main aux statistiques avec lissage dynamique  
10:50-11:20	Mike Dowd (Dalhousie University) Approximations to Facilitate Data Assimilation for High Dimensional Dynamical Systems / Approximations pour faciliter l'assimilation de données dans les systèmes dynamiques à haute dimension  
11:20-11:50	Oksana Chkrebtii (Ohio State University), Yury E. Garcia (Centro de Investigacion en Matematicas), Marcos A. Capistran (Centro de Investigacion en Matematicas) Challenges and Strategies for Inference on a Stochastic Multi-Pathogen Model of Disease Dynamics in San Luis Potosi, Mexico / Défis et stratégies d'inférence sur un modèle de dynamiques de maladie stochastique à pathogènes multiples à San Luis Potosi, Mexico  

10:20-11:50	Contributed / Communications libres (abstract/résumé 67)	E2 165 (EITC)
--------------------	---	----------------------











Causal Inference
Inférence causale

Chair/Président: Robert W. Platt

10:20-10:35	Di Shu (University of Waterloo), Grace Yi (University of Waterloo) Causal Inference with Measurement Error in Outcomes: Bias Analysis and Estimation Methods / Inférence causale comportant des erreurs de mesure dans les réponses : analyse du biais et méthodes d'estimation  
10:35-10:50	Shixiao Zhang (University of Waterloo), Peisong Han (University of Waterloo), Changbao Wu (University of Waterloo) A Unified Empirical Likelihood Approach to Testing MCAR and Subsequent Estimation / Approche unifiée de vraisemblance empirique pour des tests avec des données manquantes de façon complètement aléatoire et leur estimation subséquente  
10:50-11:05	Gabrielle Simoneau (McGill University), Erica E.M. Moodie (McGill University), Robert W. Platt (McGill University), Bibhas Chakraborty (Duke-NUS Medical School, National University of Singapore) Non-regular Inference for Dynamic Weighted Ordinary Least Squares: Understanding the Impact of Solid Food Intake in Infancy on Childhood Weight / Inférence non régulière pour l'approche dynamique pondérée des moindres carrés ordinaires : comprendre l'impact d'une diète d'aliments solides d'un nourrisson sur le poids durant l'enfance  
11:05-11:20	Yuying Xie (University of Waterloo), Yeying Zhu (University of Waterloo), Cecilia Cotton (University of Waterloo) Kernel-Based Causal Inference / Inférence causale basée sur le noyau  
11:20-11:35	Sean McGrath (McGill University), Russell Steele (McGill University), Andrea Benedetti (McGill University) Incorporating Medians in Meta-Analysis / Inclusion des médianes dans la méta-analyse  
11:35-11:50	Marc Simard (Institut national de santé publique du Québec), Nathalie Vandal (Institut National de Santé Publique du Québec), Meranda Nakhla (Research Institute of the McGill University Health Centre and McGill University), Elham Rahme (Research Institute of the McGill University Health Centre), Astrid Guttmann (Hospital for Sick Children, University of Toronto, Institute for Clinical Evaluative Sciences) Use of the Segmented Regression Analysis Approach to Adjust for Potential Confounders and Effect Modifiers when Evaluating Public Health Interventions / Utilisation de la régression segmentée pour tenir compte de la confusion et de l'effet modifiant dans l'évaluation d'une intervention en santé publique  

10:20-11:50 **Contributed / Communications libres** (abstract/résumé 71) **E2 160 (EITC)**
Probability and Estimation**Probabilité et estimation**

Chair/Président: Clarence Simard

- 10:20-10:35 **Clemonell Bilayi-Biakana** (University of Ottawa), **Rafal Kulik** (University of Ottawa), **Gail Ivanoff** (University of Ottawa)
Tail Empirical Processes for Stochastic Volatility Models / Processus empiriques de queue des modèles de volatilité stochastique  
- 10:35-10:50 **Hai Yan Liu** (University of Ottawa), **Ioana Schiopu-Kratina** (University of Ottawa), **Pierre-Jerome Bergeron** (University of Ottawa), **Mayer Alvo** (University of Ottawa)
Modeling Recurrent Gap times Through Conditional GEE / Modélisation des intervalles récurrents au moyen d'équations d'estimation généralisées  
- 10:50-11:05 **James Kim** (Royal Canadian Air Forces), **Mohan Chaudhry** (Royal Military College of Canada)
New approach to GIX/Geom/1 Queues involving Heavy-Tailed distributions / Nouvelle approche pour les files d'attente GIX/Geom/1 comportant des distributions à queue lourde  
- 11:05-11:20 **Yueleng Wang** (University of Windsor), **Sévérien Nkurunziza** (University of Windsor)
On Convergence of Sample Correlation Matrices in High-Dimensional Data / Convergence de matrice des corrélations de l'échantillon avec données à haute dimension  
- 11:20-11:35 **Yoshihiro Taniguchi** (University of Waterloo), **Christiane Lemieux** (University of Waterloo)
Importance Sampling Techniques for Single Index Models / Techniques d'échantillonnage préférentiel pour les modèles à un seul indice  

10:20-11:50 **Contributed / Communications libres** (abstract/résumé 74) **E2 150 (EITC)**
Regression Methods: Means, Modes, and Quantiles**Méthodes de régression : moyennes, modes et quantiles**

Chair/Président: Kevin E. Thorpe

- 10:20-10:35 **Zihang Lu** (University of Toronto)
A Comparison of Mixture Modeling Approaches for Clustering Longitudinal Binary Data / Comparaison d'approches de modélisation par mélange pour le regroupement de données binaires longitudinales  
- 10:35-10:50 **Amadou Diogo Barry** (Université du Québec à Montréal), **Karim Oualkacha** (Université du Québec à Montréal), **Arthur Charpentier** (Université de Rennes)
Quantile and Expectile Regression for Random Effects Models / La régression quantile et expectile appliquées au modèle avec effets aléatoires  
- 10:50-11:05 **Peijun Sang** (Simon Fraser University), **Jiguo Cao** (Simon Fraser University), **Richard Lockhart** (Simon Fraser University)
Component Selection and Estimation for Functional Semiparametric Additive Model / Sélection et estimation des composantes d'un modèle additif fonctionnel semiparamétrique  
- 11:05-11:20 **Bouchra Nasri** (HEC Montréal), **Bruno Rémillard** (HEC Montréal), **Taoufik Bouezmarni** (Université de Sherbrooke)
On Copula-Based Conditional Quantile Estimators / Estimateurs de quantiles conditionnels basés sur la copule  
- 11:20-11:35 **Menglu Che** (University of Waterloo), **Linglong Kong** (University of Alberta), **Rhonda Bell** (University of Alberta), **Yan Yuan** (University of Alberta)
Trajectory Modeling of Gestational Weight: a Functional Principal Component Analysis Approach / Modélisation de l'évolution du poids pendant la grossesse : analyse en composante principale fonctionnelle  

11:35-11:50 **Myrtha E. Reyna Vargas** (University of Toronto), **Nicholas Mitsakakis** (University of Toronto)
Ordinal Regression for the Analysis of Health Utility Data / Régression ordinaire pour l'analyse de données de cotation d'état de santé  

12:00-17:30 **Poster / Poster** **Atrium (EITC)**

Case Study 1: What is the Impact of Natural Disasters in Canada?

Étude de cas 1 : Quel est l'impact des catastrophes naturelles au Canada?

Chair/Président: Lisa M. Lix

12:00-17:30 **Sarah Watt** (University of Toronto), **David Canji** (University of Toronto), **Yunjing Li** (University of Toronto)
University of Toronto / University of Toronto  

12:00-17:30 **Scott S. White** (University of Winnipeg), **Justin Dyck** (University of Winnipeg), **Charanpal Singh** (University of Winnipeg), **Ian Hiebert** (University of Winnipeg)
University of Winnipeg / University of Winnipeg  

12:00-17:30 **Myrtha E. Reyna Vargas** (University of Toronto), **Ruixue Dai** (University of Toronto), **Shubham Sharma** (University of Toronto), **Mitchell Sutton** (University of Toronto), **Yeva Sahakyan** (University of Toronto), **Xin Zheng** (University of Toronto)
University of Toronto / University of Toronto  

12:00-17:30 **Melissa Van Bussel** (Trent University), **Carlone Scott** (Trent University), **Mark Weygang** (Trent University), **Skye Griffith** (Trent University)
Trent University / Trent University  

12:00-17:30 **Junchi Bin** (University of British Columbia - Okanagan), **Fang Shi** (University of British Columbia)
University of British Columbia / University of British Columbia  

12:00-17:30 **Zheng Jing Hu** (University of Toronto), **Rose Garrett** (University of Toronto), **Lusi Yang** (University of Toronto), **Siyu Wu** (University of Toronto), **Mehdi Rostamiforooshani** (University of Toronto)
University of Toronto / University of Toronto  

12:00-17:30 **Poster / Poster** **Atrium (EITC)**

Case Study 2: What Clinical and Genetic Characteristics Predict Chronic Disease?

Étude de cas 2 : Les données d'expression génique permettent-elles d'identifier les patients souffrant de maladies intestinales inflammatoires?

Chair/Président: Lisa M. Lix

Organizer/Responsable: Lisa M. Lix

12:00-17:30 **Jia Li** (University of Calgary), **Yunting Fu** (University of Calgary), **Levi James Mason** (University of Calgary)
University of Calgary / University of Calgary  

12:00-17:30 **Samantha Gorun** (University of Winnipeg), **Matthew Love** (University of Winnipeg), **Joshua Manalo** (University of Winnipeg), **Markus Bunge** (University of Winnipeg)
University of Winnipeg / University of Winnipeg  

12:00-17:30 **Thai Son Tang** (Dalla Lana School of Public Health), **Ting Zhang** (University of Toronto), **Lehang Zhong** (University of Toronto), **Yue Yin** (University of Toronto), **Yi Meng Chang** (University of Toronto), **Sudipta Saha** (University of Toronto)
University of Toronto / University of Toronto  

12:00-17:30 **Zixiang Guan** (University of Calgary), **Lingzhu Shen** (University of Calgary), **Yue Xu** (University of Calgary)
University of Calgary / University of Calgary  

12:00-17:30	Grace Guan Hsu (SFU), Lillian Lin (Simon Fraser University), Yue Ma (Simon Fraser University), Ran Wang (Simon Fraser University), Michelle Thiessen (Simon Fraser University) Simon Fraser University / Simon Fraser University	 
12:00-17:30	Kaiqiong Zhao (McGill University), Xiaoyu Wei (McGill University), Fatema Johara (McGill University) McGill University / McGill University	 
12:00-17:30	Janie Coulombe (McGill University), Sean McGrath (McGill University), Zhongya Wang (McGill University) McGill University / McGill University	 
12:00-17:30	Kelly Ramsay (University of Manitoba), Inesh Prabuddha (University of Manitoba), Kanika Grover (University of Manitoba), Shamsia Sobhan (University of Manitoba) University of Manitoba / University of Manitoba	 
12:00-17:30	Jiaying You (University of Manitoba), Ye Tian (University of Manitoba), Md. Mohaiminul Islam (University of Manitoba), Joynob Ara Siddiqua (University of Manitoba) University of Manitoba / University of Manitoba	 

13:30-15:00 **Invited / Sur invitation** (abstract/résumé 78) **E3 270 (EITC)**







Recent Developments in Biostatistics

Récents développements en biostatistique

Chair/Président: Xikui Wang

Organizer/Responsable: Xikui Wang

Sponsor/Commanditaires: Biostatistics Section / Groupe de biostatistique

13:30-14:00	Peter X. Song (University of Michigan), Ling Zhou (University of Michigan) Confidence Estimating Functions for Data Integration / Fonctions d'estimation des intervalles de confiance pour l'intégration de données	 
14:00-14:30	Mikelis Guntars Bickis (University of Saskatchewan), Naeima Ashleik (University of Saskatchewan), Juxin Liu (University of Saskatchewan) Inference from Survival Data using Imprecise Probabilities / Inférence à partir de données de survie au moyen de probabilités imprécises	 
14:30-15:00	J. Jack Lee (University of Texas MD Anderson Cancer Center) Bayesian Adaptive Designs for Efficient Drug Development and Precision Oncology / Plans d'expérience adaptatifs bayésiens pour le développement de médicaments efficaces et une oncologie de précision	 

13:30-15:00 **Invited / Sur invitation** (abstract/résumé 80) **E2 150 (EITC)**





Time Series Analysis with Applications in Brain Activity Analysis



Analyse de séries chronologiques avec applications à l'analyse de l'activité du cerveau

Chair/Président: Hao Yu

Organizer/Responsable: Hao Yu

Sponsor/Commanditaires: Probability Section / Groupe de probabilité

13:30-14:00	Jörn Diedrichsen (University of Western Ontario) Spatio-Temporal Noise Models for the Analysis of Functional Magnetic Resonance Imaging Data / Modèles de bruit spatiotemporels pour l'analyse des données d'imagerie par résonance magnétique fonctionnelle	 
14:00-14:30	Piotr Kokoszka (Colorado State University) Change Point Detection in Functional Time Series Models for Yield Curves / Détection des points de changement dans les modèles de séries chronologiques fonctionnelles pour les courbes de rendement	 

14:30-15:00 **Herold Dehling** (Ruhr University), **Alexander Schnurr** (University of Siegen)
 Change-Point Detection in Time Series using Ordinal Patterns / Détection des points de
 changement dans les séries chronologiques par motifs ordinaux  



13:30-15:00 **Invited / Sur invitation** (abstract/résumé 82) **E2 110 (EITC)**

Here Comes the Future? Data Science Programs for Undergraduate and Graduate Students
Voici l'avenir? Programmes de sciences des données au premier cycle et aux cycles supérieurs

Chair/Président: Bruce Dunham

Organizer/Responsable: Bruce Dunham

Sponsor/Commanditaires: Statistical Education Section / Groupe d'éducation en statistique

13:30-14:15 **Paul Gustafson** (University of British Columbia), **Jenny Bryan** (University of British Columbia), **Giuseppe Carenini** (University of British Columbia), **Vincenzo Coia** (University of British Columbia), **Giulio V. Dalla Riva** (University of British Columbia), **Michael A. Gelbart** (University of British Columbia), **Gail C. Murphy** (University of British Columbia), **Raymond T. Ng** (University of British Columbia), **Tiffany A. Timbers** (University of British Columbia)
 Designing and Launching a Professional Master's Program in Data Science: Tales from the UBC Experience / Conception et lancement d'un programme professionnel de maîtrise en science des données : anecdotes tirées de l'expérience à l'UBC  



14:15-15:00 **Stefan Steiner** (University of Waterloo)
 Data Science Programs at the University of Waterloo / Des programmes en science des données à l'Université de Waterloo  



13:30-15:00 **Invited / Sur invitation** (abstract/résumé 83) **E2 105 (EITC)**



Developments in Event History Analysis with Informative Censoring or Observation Schemes
Développements en analyse de l'historique des événements avec schémas de censure ou d'observation informative

Chair/Président: Hua Shen

Organizer/Responsable: Hua Shen

13:30-14:00 **Yayuan Zhu** (University of Texas MD Anderson Cancer Center), **Jerald F. Lawless** (University of Waterloo)
 Event History Analysis of Data from Non-Ignorable Observation Schemes / Analyse historique de données à partir de plans d'observation non ignorables  

14:00-14:30 **X. Joan Hu** (Simon Fraser University), **Huijing Wang** (Apple Inc.), **John Spinelli** (BC Cancer Agency)
 Risk Classification and Prediction with Administrative Health Data / Classification et prédiction des risques avec des données administratives sur la santé  







14:30-15:00 **Leilei Zeng** (University of Waterloo), **Nathalie Moon** (University of Waterloo), **Richard J. Cook** (University of Waterloo)
 Design and Relative Efficiency of Tracing Studies in Cohorts with Loss to Followup / Conception d'études de traçabilité et efficacité relative de celles-ci dans des cohortes avec pertes au suivi  

13:30-15:00 **Invited / Sur invitation** (abstract/résumé 85) **E2 125 (EITC)**

Recursive Partitioning Methods in Biostatistics
Méthodes de partitionnement récursif en biostatistique

Chair/Président: Liqun Diao

Organizer/Responsable: Liqun Diao

- 13:30-14:00 **Ruoqing Zhu** (University of North Carolina at Chapel Hill), **Yifan Cui** (University of North Carolina at Chapel Hill), **Mai Zhou** (University of Kentucky), **Michael Kosorok** (University of North Carolina at Chapel Hill)
Some Asymptotic Results on Tree-based Survival Models / Quelques résultats asymptotiques de modèles de survie basés sur les arbres  
- 14:00-14:30 **Annette M. Molinaro** (University of California, San Francisco), **Adam Olshen** (UCSF), **Robert Strawderman** (University of Rochester)
Expanding the Predictive Ability of an Interpretable Tree / Élargir la capacité prédictive des arbres d'interprétation  
- 14:30-15:00 **Danping Liu** (National Institutes of Health), **Jared C. Foster** (Eunice Kennedy Shriver National Institute of Child Health and Human Development, Mayo Clinic), **Paul S. Albert** (Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Cancer Institute), **Aiyi Liu** (Eunice Kennedy Shriver National Institute of Child Health and Human Development)
Tree-Based Approaches for Personalized Risk Prediction with Longitudinal Biomarkers: An Application to Fetal Growth / Approches basées sur les arbres pour la prédiction de risques personnalisés avec biomarqueurs longitudinaux : une application à la croissance fœtale  







13:30-15:00 **Invited / Sur invitation** (abstract/résumé 87) **E2 130 (EITC)**

Survey Sampling Theory and Practice – A Celebration of J.N.K. Rao's Research

Théorie et pratique de l'échantillonnage d'enquêtes – une célébration de la recherche de J.N.K. Rao

Chair/Président: Changbao Wu

Organizer/Responsable: Changbao Wu





- 13:30-14:00 **Mahmoud Torabi** (University of Manitoba), **Gauri Datta** (University of Georgia), **J.N.K. Rao** (Carleton University)
Small Area Estimation with Covariate Measurement Error / Estimation pour petits domaines avec erreurs de mesure des covariables  
- 14:00-14:30 **Wesley Yung** (Statistics Canada), **Mike Hidioglou** (Statistics Canada), **Elisabeth Neusy** (Statistics Canada)
Non-response Follow-up for Business Surveys / Suivi de la non-réponse dans les enquêtes auprès des entreprises  
- 14:30-15:00 **David Haziza** (Université de Montréal), **Jean-François Beaumont** (Statistics Canada), **Isabelle Lefebvre** (Université de Montréal)
Simplified Variance Estimation in Surveys / Estimation de la variance simplifiée dans les enquêtes  







13:30-15:00 **Contributed / Communications libres** (abstract/résumé 89) **E2 155 (EITC)**

Business and Actuarial Statistics

Statistique commerciale et actuarielle

Chair/Président: Sanjeena Dang

- 13:30-13:45 **Wenjun Jiang** (University of Western Ontario), **Jiandong Ren** (University of Western Ontario), **Hanping Hong** (University of Western Ontario)
Pareto-optimal Reinsurance with Two Constraints under Distortion Risk Measure / Réassurance optimale de Pareto avec deux contraintes sous une mesure du risque de distorsion  
- 13:45-14:00 **Danqiao Guo** (University of Waterloo), **Tony Wirjanto** (University of Waterloo), **Chengguo Weng** (University of Waterloo)
Improved Global Minimum Variance Portfolio via a Dominant Eigenvalue Shrinkage / Portefeuille amélioré de variance minimale globale à l'aide d'une diminution de la valeur propre dominante  

14:00-14:15	Yu Gao (University of Waterloo) Detecting Adverse Drug Effects from Pharmacovigilance Databases / Détection des effets secondaires des médicaments grâce aux bases de données de pharmacovigilance  
14:15-14:30	Vanessa Bierling (McMaster University), Paul D. McNicholas (McMaster University) A Latent Gaussian Mixture Model for Longitudinal Data / Un modèle de mélange gaussien latent pour des données longitudinales  
14:45-15:00	Nkumbuludzi Ndwapi (McMaster University), Paul D. McNicholas (McMaster University) Robust Mixture Discriminant Analysis / Analyse discriminante de mélanges robuste  

13:30-15:00 **Contributed / Communications libres** (abstract/résumé 93) **E2 160 (EITC)**

Longitudinal and Survival Data

Données longitudinales et de survie

Chair/Président: Tolulope T. Sajobi

13:30-13:45	Jia Li (University of Calgary), Alexander R. de Leon (University of Calgary), Haocheng Li (University of Calgary) A Longitudinal Linear Mixed Effect Model for Bivariate Responses with Measurement Errors and Misclassification in Covariates / Modèle linéaire longitudinal à effets mixtes pour réponses bidimensionnelles en présence d'erreurs de mesure et classification erronée dans les covariables  
13:45-14:00	Harris Quach (Pennsylvania State University) Accurate Confidence Intervals for Clustered Data and Small Samples / Intervalles de confiance adéquats pour données en grappes avec petits échantillons  
14:00-14:15	Yidan Shi (University of Waterloo), Leilei Zeng (University of Waterloo), Mary E. Thompson (University of Waterloo), Suzanne Tyas (University of Waterloo) Incorporating Auxiliary Information by Joint Modelling of Pseudo Data and Length Biased Data / Intégration d'informations auxiliaires par modélisation conjointe de pseudo-données et données biaisées en longueur  
14:15-14:30	Kunasekaran Nirmalkanna (Memorial University of Newfoundland), Candemir Cigsar (Memorial University of Newfoundland) Analysis of Recurrent Events with Dynamic Models for Dependent Gap Times: A Copula Approach / Analyse des événements récurrents à l'aide de modèles dynamiques pour les temps d'écart dépendants : une approche par copules  
14:30-14:45	Li-Pang Chen (University of Waterloo) Left-Truncated and Right-Censored Survival Data Analysis with Covariate Measurement Error / Analyse de données de survie tronquées à gauche et censurées à droite avec erreur de mesure des covariables  
14:45-15:00	Kaida Cai (University of Calgary), Hua Shen (University of Calgary), Xuwen Lu (University of Calgary) Group Variable Selection for Recurrent Event Data / Sélection de groupes de variables pour des données d'événements récurrents  

13:30-15:00 **Contributed / Communications libres** (abstract/résumé 98) **E2 165 (EITC)**

Statistical Genetics

Génétique statistique

Chair/Président: Lei Sun

- 13:30-13:45 **Christina M. Nieuwoudt** (Simon Fraser University), **Jinko Graham** (Simon Fraser University)
SimRVPedigree: An R Package to Simulate Pedigrees Ascertained for Multiple Relatives Affected by a Rare Disease / SimRVPedigree : Un paquet R pour simuler des pedigrees établis pour plusieurs personnes apparentées touchées par une maladie rare  
- 13:45-14:00 **Charith Bhagya Karunarathna** (Simon Fraser University), **Jinko Graham** (Simon Fraser University)
A Comparison of Association Methods for Fine-Mapping from Sequence Data in Diploid Populations using the Case-Control Study Design / Une comparaison des méthodes d'association pour la localisation détaillée à partir de données séquentielles dans des populations diploïdes utilisant le format de l'étude cas-témoin  
- 14:00-14:15 **Lin Zhang** (University of Toronto), **Lei Sun** (University of Toronto)
A Robust Allele-based Regression Framework for Testing Association and Hardy-Weinberg Equilibrium / Modèle de régression robuste basé sur les allèles pour tester l'association et l'équilibre Hardy-Weinberg  
- 14:15-14:30 **Elham Khodayari Moez** (University of Alberta), **Irina Dinu** (School of Public Health, University of Alberta)
Gene Set Enrichment Analysis in Longitudinal Studies / Analyse de l'enrichissement des ensembles de gènes dans les études longitudinales  
- 14:30-14:45 **Agnes Nessie Amu** (University of Manitoba), **Elif Fidan Acar** (University of Manitoba)
Factor Copula Analysis for Multivariate Ordinal Data / Analyse de copule avec facteurs pour données ordinales multivariées  
- 14:45-15:00 **Maxime Turgeon** (McGill University), **Stepan Grinek** (BC Cancer Agency), **Celia M.T. Greenwood** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University), **Aurélie Labbe** (HEC Montréal)
Reduced-Rank Singular Value Decomposition for Dimension Reduction with High-Dimensional Data / Décomposition en valeurs singulières à rang réduit appliquée aux méthodes de réduction dimensionnelle pour l'étude de données de grande dimension  

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 102) **E2 155 (EITC)**







Insurance Risk and Ruin Measures

Mesures de risque et de ruine en assurance

Chair/Président: Jose Garrido

Organizer/Responsable: Jose Garrido

Sponsor/Commanditaires: Actuarial Science Section / Groupe d'actuariat

- 15:30-16:00 **Ilie Radu Mitric** (Université Laval), **Julien Trufin** (Université Libre de Bruxelles), **Amine Mohamed Lkabous** (Université du Québec à Montréal)
Properties of Risk Measures Inspired from the Ruin Probability, the Deficit at Ruin, and the Time of Ruin / L'étude des propriétés de quelques mesures de risque dérivées de la probabilité de la ruine, le déficit à la ruine et le temps de ruine  
- 16:00-16:30 **Georgios Pitselis** (University of Piraeus)
Risk Measures Related to both the Information of an Individual Risk and Industry Risk / Les mesures du risque relatives à l'information en matière de risque individuel et de risque lié à l'industrie  
- 16:30-17:00 **Jose Garrido** (Concordia University), **Alejandro Balbàs** (University Carlos III de Madrid, Spain), **Ramin Okhrati** (United of Southampton, United Kingdom)
Good Deal Indices in Asset Pricing: Actuarial and Financial Implications / Le concept d'au-baine en évaluation d'actifs : implications actuarielles et financières  
-

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 104) **E2 125 (EITC)**





Using Randomization and Simulation in the Teaching of Statistics

Enseigner la statistique par la randomisation et la simulation

Chair/Président: Jim Stallard

Organizer/Responsable: Jim Stallard

Sponsor/Commanditaires: Statistical Education Section / Groupe d'éducation en statistique

- 15:30-16:15 **Nathan Tintle** (Dordt College)
Teaching the Statistical Investigation Process with Simulation-Based Inference to Improve Statistical Thinking / Enseignement du processus de recherche statistique avec inférence fondée sur la simulation en vue d'améliorer la pensée statistique  
- 16:15-17:00 **John Sheriff** (University of Lethbridge)
Incorporating Simulation-Based Inference in an Introductory Statistics Course / Intégration de l'inférence par simulation dans un cours de base en statistique  

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 105) **E2 110 (EITC)**







Design and Analysis of Validation Studies for Health Outcomes in Electronic Health Databases

Conception et analyse d'études de validation pour les résultats cliniques dans les bases de données électroniques en santé

Chair/Président: Lisa M. Lix

Organizer/Responsable: Lisa M. Lix

Sponsor/Commanditaires: CANSSI / INCASS

- 15:30-16:00 **Elham Rahme** (McGill University), **Jiayi Ni** (Research Institute of the McGill University Health Center), **Kaberi Dasgupta** (McGill University), **Denis Talbot** (Université Laval), **Geneviève Lefebvre** (Université du Québec à Montréal), **Lisa M. Lix** (University of Manitoba), **Lucie Blais** (Université de Montréal)
Comparing Bayesian External and Internal Validation Methods in Correcting Outcome Misclassification Bias in Logistic Regression / Comparaison de méthodes internes et externes de validation bayésienne pour corriger le biais de classification de l'issue dans la régression logistique  
- 16:00-16:30 **Eric Benchimol** (University of Ottawa)
Validation of Codes and Algorithms are an Essential Part of Research Using Routinely Collected Health Data / La validation des codes et des algorithmes est un élément essentiel de la recherche utilisant des données sur la santé recueillies systématiquement  
- 16:30-17:00 **Tyler Williamson** (University of Calgary), **Brendan Cord Lethebe** (University of Calgary), **Colin Weaver** (University of Calgary), **Tolulope T. Sajobi** (University of Calgary), **Hude Quan** (University of Calgary)
Feature Selection Methods for the Development of Case Definitions for Common Chronic Conditions in Primary Care Electronic Medical Record Data / Les méthodes de sélection de caractéristiques pour l'élaboration des définitions cliniques de cas des maladies chroniques courantes utilisant les dossiers médicaux électroniques de soins primaires  

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 107) **E2 130 (EITC)**

Design and Analysis Issues in Cluster Randomized Controlled Trials

Problèmes de conception et d'analyse dans les essais randomisés contrôlés par grappes

Chair/Président: Lehana Thabane

Organizer/Responsable: Jinhui Ma

15:30-16:00	Robert W. Platt (McGill University) Observational Analyses of Cluster-Randomized Trials / Analyses d'observation d'essais randomisés par grappes  
16:00-16:30	Eleanor M. Pullenayegum (The Hospital for Sick Children, Toronto) Analysis of Cluster-Randomized Trials with a Longitudinal Outcome Subject to Irregular Observation / Analyse d'essais randomisés par grappes avec variable dépendante longitudinale soumise à une observation irrégulière  
16:30-17:00	Jinhui Ma (Children's Hospital of Eastern Ontario Research Institute), Monica Taljaard (Ottawa Hospital Research Institute), Lisa Dolovich (McMaster University), Janusz Kaczorowski (University of Montreal), Larry Chambers (University of Ottawa), Lehana Thabane (McMaster University) Strategies to Handle Missing Binary Outcomes in Cluster Randomized Trials / Stratégies pour traiter les variables indépendantes binaires manquantes dans des essais randomisés par grappes  







15:30-17:00 **Invited / Sur invitation** (abstract/résumé 110) **E2 105 (EITC)**

Double Robustness and Dynamic Treatment Regimes

La double robustesse et les régimes de traitement dynamiques

Chair/Président: Michael Wallace

Organizer/Responsable: Michael Wallace

15:30-16:00	David A. Stephens (McGill University), Michael Wallace (University of Waterloo), Erica E.M. Moodie (McGill University) G-estimation and Model Selection for Dynamic Treatment Regimes / G-estimation et sélection de modèles pour régimes de traitement dynamiques  
16:00-16:30	Olli Saarela (University of Toronto) On the Bayesian Semi-Parametric Double Robustness Property / Double robustesse semi-paramétrique bayésienne  
16:30-17:00	Peisong Han (University of Waterloo) Achieving Higher Efficiency and Robustness in Longitudinal Studies with Dropout / Amélioration de l'efficacité et de la robustesse dans les études longitudinales avec abandons  

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 112) **E2 150 (EITC)**







Self-Normalized Processes: Dedicated to Miklos Csorgo on the Occasion of his 85th Birthday

Processus auto-normalisés : en l'honneur du 85e anniversaire de Miklos Csorgo

Chair/Président: Don McLeish

Organizer/Responsable: Barbara Szyszkowicz

Sponsor/Commanditaires: Probability Section / Groupe de probabilité







15:30-16:00	Miklos Csorgo (Carleton University) Self-Normalized Weak and Strong Functional Limit Laws in the Domain of Attraction of Stable Laws / Lois limites fonctionnelles faibles et fortes autonormalisées dans les domaines d'attraction des lois stables  
16:00-16:30	Yuliya V. Martynyuk (University of Manitoba) Functional Central Limit Theorems for Self-Normalized Processes with Statistical Applications / Théorèmes centraux limites fonctionnels pour des processus auto-normalisés, et applications statistiques  
16:30-17:00	Masoud Nasari (Carleton University) Self-Normalized Partial Sums of Randomly Weighted Data / Sommes partielles autonormalisées de données pondérées au hasard  

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 113) **E3 270 (EITC)**
Data Science in Biomedical Studies**Science des données dans les études biomédicales**

Chair/Président: X. Joan Hu and Lisa Le













Organizer/Responsable: X. Joan Hu and Lisa Le

Sponsor/Commanditaires: Committee on Women in Statistics / Comité sur les femmes en statistique

- 15:30-16:00 **Juxin Liu** (University of Saskatchewan), **Paul Gustafson** (University of British Columbia),
Dezheng Huo (University of Chicago)
Discordance in Hormone Receptor Status Between Two Primary Breast Cancers: The Impact
of Misclassification / Discordance dans l'état des récepteurs hormonaux entre deux cancers
primitifs du sein : l'incidence des erreurs de classification  
- 16:00-16:30 **Laura L.E. Cowen** (University of Victoria)
Estimating the Population Size of Greater Victoria's Injection Drug Users using Capture-
Recapture Methods / Estimation de la taille de la population des utilisateurs de drogues
injectables dans la région de Victoria au moyen des méthodes de « capture-recapture »  
- 16:30-17:00 **Wendy Lou** (University of Toronto)
Statistical Opportunities and Challenges in Multiple-Platform Biomedical Studies / Opportu-
nités et défis statistiques dans les études biomédicales multiplateformes  

15:30-17:00 **Contributed / Communications libres** (abstract/résumé 115) **E2 165 (EITC)**
Industrial Statistics**Statistique industrielle**

Chair/Président: Jean-François Plante

- 15:30-15:45 **Da Zhong (Dexen) Xi** (Western University), **Charmaine B. Dean** (Western University),
Steve Taylor (Pacific Forestry Centre)
Dependent Models for the Duration and Size of BC Fires / Modèles dépendants pour la durée
et la taille des feux en Colombie-Britannique  
- 15:45-16:00 **Yongho Lim** (Memorial University of Newfoundland), **Candemir Cigsar** (Memorial University
of Newfoundland)
Testing for Parallel Carryover Effects in Redundant Systems / Tester les effets de report
parallèle dans les systèmes redondants  
- 16:00-16:15 **Steven Wu** (Simon Fraser University)
Building a Basketball Analytics Startup as a Statistics Student / Développer une entreprise à
titre d'étudiant en statistiques à partir d'une analytique du basket-ball  
- 16:15-16:30 **Abdollah Safari** (Simon Fraser University), **Rachel MacKay Altman** (Simon Fraser Uni-
versity)
Efficient Conversion Probability Estimator for Display Advertising / Estimateur efficace de la
probabilité de conversion pour affichage publicitaire  
- 16:30-16:45 **Bryan Yates** (University of Waterloo), **Martin Lysy** (University of Waterloo), **Aleks Labuda**
(Asylum Research)
Parametric Spectral Density Estimation in Atomic Force Microscopy / Estimation paramé-
trique de la densité spectrale dans la microscopie à force atomique  
- 16:45-17:00 **Peter Tait** (McMaster University), **Paul D. McNicholas** (McMaster University)
Clustering of Functional Data using Mixture Model-Based Clustering / Regroupement de
données fonctionnelles à l'aide d'un regroupement basé sur un modèle de mélanges  
-

15:30-17:00 **Contributed / Communications libres** (abstract/résumé 119) **E2 160 (EITC)**
Spatial, Environmental, and Network Data**Données spatiales, environnementales et de réseau**



Chair/Président: Patrick E. Brown

- 15:30-15:45 **Dongmeng Liu** (Simon Fraser University), **Jinko Graham** (Simon Fraser University)
Simple Measures of Individual Cluster-Membership Certainty for Hard Partitional Clustering / Mesures simples de la certitude d'appartenance à un regroupement individuel pour le regroupement de partition dur.  
- 15:45-16:00 **Trevor Thomson** (Simon Fraser University), **X. Joan Hu** (Simon Fraser University), **John Braun** (University of British Columbia - Okanagan)
Statistical Modelling of Spot Fires in Forest Fire Control / Modélisation statistique des incendies de surface dans la lutte contre les incendies de forêt  
- 16:00-16:15 **Shabnam Balamchi** (University of Manitoba), **Shabnam Balamchi** (University of Manitoba)
Spatial Modeling of Repeated Events: An Application to Disease Mapping / Modélisation spatiale des événements répétés : une application à la cartographie des maladies  
- 16:15-16:30 **Kevin D.J. McGregor** (McGill University), **Aurélien Labbe** (HEC Montréal), **Celia M.T. Greenwood** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University)
A Penalized Approach to Estimating Network Changes in the Human Microbiome / Une approche pénalisée pour estimer les changements de réseau dans le microbiome humain  
- 16:30-16:45 **Fang He** (University of Western Ontario), **Duncan Murdoch** (University of Western Ontario), **Reg Kulperger** (University of Western Ontario)
Statistical Modelling of Carbon Dioxide Flux Data / Modélisation statistique de données de flux de dioxyde de carbone  
- 16:45-17:00 **Fahimeh Moradi** (University of Alberta), **Elham Khodayari Moez** (University of Alberta), **Irina Dinu** (University of Alberta)
Identifying Expression Quantitative Trait Loci in Genome-Wide Association Studies using Matrix eQTL / Identifier les loci de traits d'expression quantitatifs dans des études d'association pangénomiques par l'utilisation de Matrix eQTL  

Tuesday June 13**Mardi 13 juin****08:40-09:45****Invited / Sur invitation** (abstract/résumé 123)**UC 210 (UC)****SSC Gold Medal Address****Allocution du récipiendaire de la Médaille d'or de la SSC**

Chair/Président: John Petkau





Organizer/Responsable: Michael Evans

08:40-09:45**Harry Joe** (University of British Columbia)Classical and Modern Multivariate Statistical Analysis / Analyse statistique multivariée classique et moderne  **10:20-11:50****Invited / Sur invitation** (abstract/résumé 124)**E2 130 (EITC)****Statistical Modelling in Life Insurance****Modélisation statistique en assurance-vie**

Chair/Président: Johnny Li

Organizer/Responsable: Johnny Li







Sponsor/Commanditaires: Actuarial Science Section / Groupe d'actuariat

10:20-11:05**Rui Zhou** (University of Manitoba)Modelling Multi-Population Mortality Dependence with a Regime-Switching Copula / Modéliser la dépendance de la mortalité multi-populations avec une copule à changement de régimes  **11:05-11:50****Cary Chi-Liang Tsai** (Simon Fraser University), **Adelaide Di Wu** (Simon Fraser University)A Hierarchical Credibility Approach to Modeling Mortality Rates for Multiple Populations / Une approche de crédibilité hiérarchique pour modéliser les taux de mortalité de populations multiples  **10:20-11:50****Invited / Sur invitation** (abstract/résumé 126)**E2 110 (EITC)****Business Problems, Data Science Solutions****Problèmes commerciaux, solutions de sciences des données**

Chair/Président: Nathaniel T. Stevens

Organizer/Responsable: Nathaniel T. Stevens

Sponsor/Commanditaires: Business and Industrial Statistics Section / Groupe de statistique industrielle et de gestion

10:20-10:50**James David Wilson** (University of San Francisco), **Jon Mackay** (Waterloo Institute for Complexity Innovation)The Silicon Valley Wage Cartel: Understanding its Effects on the Regional Labour Market with Dynamic Networks / Le cartel des salaires de la Silicon Valley : Quels en sont les effets sur le marché du travail régional et ses réseaux dynamiques?  **10:50-11:20****Ali Ghodsi** (University of Waterloo)Generative Mixture of Networks / Mélange de réseaux génératif  **11:20-11:50****Nick Ross** (University of San Francisco)Practical Testing / Les tests pratiques  

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 128) **E3 270 (EITC)**



Beyond "Correlation is not Causation"



Au-delà de "Corrélation n'est pas causalité"

Chair/Président: Georges A. Monette

Organizer/Responsable: Georges A. Monette

Sponsor/Commanditaires: Statistical Education Section / Groupe d'éducation en statistique

10:20-11:05 **Alison L. Gibbs** (University of Toronto)
Is All That Coffee I'm Drinking Hurting or Harming Me? Understanding Causality / Une forte consommation de café : nuisible ou nocif? Comprendre la causalité  

11:05-11:50 **Georges A. Monette** (York University)
How to tell when Association might be Causation / Comment savoir quand une association peut être une causalité  

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 130) **E2 125 (EITC)**



Contemporary Statistical Methodology and Application



Méthodes statistiques contemporaines et leurs applications



Chair/Président: Liqun Wang

Organizer/Responsable: Liqun Wang

Sponsor/Commanditaires: ICSA-Canada Chapter / Section canadienne de l'ICSA

10:20-10:50 **Hsing-Ming Chang** (National Cheng Kung University, Taiwan)
Ergodicity and Approximation of the Conditional Causation Probability of Engineering Systems / Ergodicité et approximation de la probabilité conditionnelle de causalité des systèmes d'ingénierie  

10:50-11:20 **Su-Fen Yang** (National Chengchi University, Taiwan), **Chung-Ming Yang** (Ling Tung University, Taiwan), **Sin-Hong Wu** (National Chengchi University, Taiwan)
A Simple Variance Control Chart with Double Sampling Scheme / Graphique de contrôle simple pour la variance sous un plan d'échantillonnage double  

11:20-11:50 **Melody Ghahramani** (University of Winnipeg), **Scott S. White** (University of Winnipeg), **Alexander R. de Leon** (University of Calgary)
A Combined Estimating Function Approach to Count Data Regression Modeling / Approche de fonctions d'estimation combinées pour modéliser la régression des données de dénombrement  



10:20-11:50 **Invited / Sur invitation** (abstract/résumé 132) **E2 105 (EITC)**



Functional Data Analysis: Models and Applications



Analyse de données fonctionnelles : modèles et applications

Chair/Président: Haocheng Li

Organizer/Responsable: Haocheng Li

10:20-10:50 **Zhenhua Lin** (University of Toronto), **Fang Yao** (University of Toronto)
Adaptive Functional Regression with Manifold Structures / Régression fonctionnelle adaptative avec structures de variété  

10:50-11:20 **Jiguo Cao** (Simon Fraser University), **Peijun Sang** (Simon Fraser University), **Liangliang Wang** (Simon Fraser University)
Parametric Functional Principal Component Analysis / Analyse paramétrique des composantes principales fonctionnelles  



11:20-11:50 **Linglong Kong** (University of Alberta), **Zhengwu Zhang** (SAMSI), **Xiao Wang** (Purdue University), **Hongtu Zhu** (MD Anderson Cancer Centre)
Spatial Quantile Regression Models for High-Dimensional Imaging Data / Modèles de régression quantile spatiale pour données d'imagerie de grande dimension  



10:20-11:50 **Contributed / Communications libres** (abstract/résumé 134) **E2 165 (EITC)**



Methods for Censored and Recurrent Events



Méthodes pour les événements censurés et récurrents



Chair/Président: Karen Kopciuk

10:20-10:35 **Chien-Lin Mark Su** (McGill University), **Russell Steele** (McGill University), **Ian Shrier** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University)
Analysis of Recurrent Events Data Based on Trend-Renewal Process / Analyse de données d'événements récurrents fondée sur un processus de renouvellement de tendance  

10:35-10:50 **Taehan Bae** (University of Regina)
On the Mixtures of Length-biased Weibull Distributions / À propos des mélanges de distributions de Weibull avec biais de longueur  

10:50-11:05 **Defen Peng** (University of British Columbia, ICVHealth), **Kevin Burke** (University of Limerick), **Gilbert MacKenzie** (University of Limerick)
An Interval Censored MPR Survival Model / Modèle multiparamétrique de régression pour données de survie censurées par intervalle  

11:05-11:20 **Rachel MacKay Altman** (Simon Fraser University), **Andrew Henrey** (Simon Fraser University)
Practical Considerations when Analyzing Small Samples of Discrete Survival Times Using the Grouped Relative Risk Model / Considérations pratiques pour l'analyse de petits échantillons de données de temps de survie discrets à l'aide du modèle de risque relatif de groupe (GRRM)  



11:20-11:35 **Saima K. Khosa** (University of Saskatchewan), **Shahedul A. Khan** (University of Saskatchewan)
A Flexible Parametric Proportional Hazards Model for Time-to-Event Data / Un modèle de risques proportionnels paramétriques flexible pour les données de durées de vie  



10:20-11:50 **Contributed / Communications libres** (abstract/résumé 137) **E2 155 (EITC)**



New Research in Survey Methods







Nouvelles recherches en méthodes d'enquête

Chair/Président: Wesley Yung

10:20-10:35 **Puying Zhao** (University of Waterloo), **David Haziza** (Université de Montréal), **Changbao Wu** (University of Waterloo)
Empirical Likelihood and Semiparametric Inference with Survey Data / Vraisemblance empirique et inférence semi-paramétrique en présence de données d'enquête  

10:35-10:50 **Matei Mireuta** (Statistics Canada), **Jessica Andrews** (Statistics Canada, Business Surveys Methods Division (BSMD)), **Pierre Daoust** (Statistics Canada, Business Surveys Methods Division (BSMD))
The Relative Deviation from Predicted Values as a Tool to Prioritize Units for Failed Edit Follow-Up in IBSP / La différence relative par rapport aux valeurs prédites comme outil de priorisation du suivi des unités dans le cadre du PISE  

10:50-11:05 **Min Jiang** (Statistics Canada)
Job Vacancy and Wage Survey: Balancing Sampling and Operational Requirements using the Cube Method / Enquête sur les postes vacants et les salaires : équilibrer l'échantillonnage et les contraintes opérationnelles avec la méthode du Cube  

11:05-11:20	Amanda Halladay (Statistics Canada), Ming Jie Yang (Statistics Canada), François Brisebois (Statistics Canada) Segmenting the Canadian Population to Remedy Non-Response in Statistics Canada Household Surveys / Segmentation de la population canadienne afin de remédier à la non-réponse aux enquêtes auprès des ménages de Statistique Canada  
11:20-11:35	Jane Wang (Statistics Canada) The Collection Challenges of the 2016 Reverse Record Check Survey / Défis à la collecte de données dans la contre-vérification des dossiers 2016  
11:35-11:50	Gabrielle Poirier (Statistics Canada), Kevin Bosa (Statistics Canada), François Gagnon (Statistics Canada) Methods of Selecting a Person within a Household: Mailed Invitations for an Electronic Questionnaire / Méthodes de sélection d'un membre à l'intérieur d'un ménage : invitations postales pour un questionnaire électronique  

10:20-11:50 **Contributed / Communications libres** (abstract/résumé 141) **E2 160 (EITC)**

Statistical Computing and Complex Models

Calcul statistique et modèles complexes

Chair/Président: Russell Steele

10:20-10:35	Jiaxiu Li (University of New Brunswick), Guohua Yan (University of New Brunswick), Renjun Ma (University of New Brunswick) Joint Modelling of Count, Continuous and Semi-Continuous Longitudinal Data / Modélisation conjointe de données longitudinales de dénombrement, continues et semi-continues  
10:35-10:50	Kelly Ramsay (University of Manitoba), Stephane Durocher (University of Manitoba), Alexandre Leblanc (University of Manitoba) Efficiently Computing the Projection Median in R^3 / Calculer efficacement la médiane de projection dans R^3  
10:50-11:05	Yizhou Fang (University of Waterloo), Martin Lysy (University of Waterloo), Don Mcleish (University of Waterloo) Common-Factor Multivariate Stochastic Volatility Modeling with Observable Proxy / Modèle de volatilité stochastique multivarié à facteur commun avec un intermédiaire observable  
11:05-11:20	Yun Ling (University of Waterloo), Martin Lysy (University of Waterloo) Superfast Inference for Stationary Gaussian Processes / Inférence très rapide pour processus gaussiens stationnaires  
11:20-11:35	Sahir R. Bhatnagar (McGill), Yi Yang (McGill University), Alexia Jolicoeur-Martineau (Jewish General Hospital), Ashley Wazana (McGill University), Celia M.T. Greenwood (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University) Strong Heredity Penalized Regression Models for Non-Linear Gene-Environment Interactions / Modèles de régression pénalisés ayant la propriété d'hérédité forte pour les interactions entre facteurs génétiques et environnementaux  
11:35-11:50	Reza Ramezan (California State University, Fullerton) Computational Neuroscience: A Romance of Many Disciplines / La neuroscience computationnelle : un lien étroit entre plusieurs disciplines  


10:20-11:50 **Contributed / Communications libres** (abstract/résumé 146) **E2 150 (EITC)**

Statistical Methods for Infectious Diseases and Other Health Outcomes





Méthodes statistiques pour les maladies infectieuses et autres résultats cliniques

Chair/Président: Tyler Williamson

10:20-10:35	Zhaoyang Tian (University of Waterloo), Pengfei Li (University of Waterloo), Kun Liang (University of Waterloo) Nonparametric Estimation in a Compound Mixture Model with Application to a Malaria Study / Estimation non paramétrique dans un modèle de mélanges composé avec application à une étude sur la malaria  
10:35-10:50	Thuva Vanniyasingam (McMaster University), Caitlin Daly (McMaster University), Xuejing Jin (McMaster University), Yuan Zhang (McMaster University), Charles Cunningham (McMaster University), Gary Foster (McMaster University), Lehana Thabane (McMaster University) Investigating the Impact of Design Characteristics on Statistical Efficiency within Discrete Choice Experiments: A Systematic Survey / Étude d'impact des caractéristiques d'un plan d'étude sur l'efficacité statistique dans le cadre d'expériences de choix discrets : une étude systématique  
10:50-11:05	Md. Mahsin (University of Calgary), Rob Deardon (University of Calgary) Geo-Dependent Individual-Level Models for Infectious Diseases Transmission / Modèles géo-dépendants au niveau des individus pour la transmission des maladies infectieuses  
11:05-11:20	Claire Boteler (Queen's University), David Thomson (Queen's University), Troy Day (Queen's University) Seasonality in Influenza Mortality: Moving Beyond the Annual Cycle / Caractère saisonnier de la mortalité causée par la grippe : au-delà du cycle annuel  
11:20-11:35	Nicholas Mitsakakis (University of Toronto) Benefits of Transformations in the Analysis of Health Utility Data / Bénéfices des transformations dans l'analyse des données d'utilité  
11:35-11:50	Katherine Davies (University of Manitoba), William Volterman (Syracuse University) Progressively Type-II Censored Competing Risks Data from the Linear Exponential Distribution / Risques concurrents soumis à une censure progressive de type II à partir d'une distribution exponentielle linéaire  

12:00-17:30	Poster / Poster (abstract/résumé 150)	Atrium (EITC)
Poster Session Séance d'affichage		
12:00-17:30	Dola Pathak (University of Nebraska at Lincoln), Christopher Bilder (University of Nebraska at Lincoln) Estimating a Disease Prevalence with Group Testing Data and Imperfect Assays / Estimation de la prévalence d'une maladie à l'aide de données de tests de groupe et d'essais imparfaits  	
12:00-17:30	Sarah A. MacQueen (University of British Columbia Okanagan), Rebecca Tyson (University of British Columbia Okanagan) Modelling the Effects of Memory on Bumblebee Pollination Services / Modélisation des effets de la mémoire sur les services de pollinisation des bourdons  	
12:00-17:30	Hensley Hubert Mariathas (Memorial University of Newfoundland), Shabnam Asghari (Memorial University of Newfoundland), Alvin Simms (Memorial University of Newfoundland), Masoud Mahdavian (Memorial University of Newfoundland), Oliver Hurley (Memorial University of Newfoundland) Spatial Mixed Model for Binary Response with an Application to Single Low-Density Lipoprotein Dyslipidemia / Modèle spatial mixte pour réponse binaire avec une application à la dyslipidémie de lipoprotéines de basse densité  	

- 12:00-17:30 **Jingjia Chu** (University of Western Ontario), **Reg Kulperger** (University of Western Ontario), **Hao Yu** (University of Western Ontario)
A Time Series Approach for the Underlying Driven Process in Stocks and its Large Sample Theory / Une approche par séries chronologiques pour le processus sous-jacent des actions et sa théorie des grands échantillons  
- 12:00-17:30 **Kevin Matira** (McMaster University), **Paul D. McNicholas** (McMaster University)
Discriminant Analysis for Longitudinal Data / Analyse discriminante pour des données longitudinales  
- 12:00-17:00 **Bingrui (Cindy) Sun** (University of Calgary), **MohanaGowri Arumugam** (University of Calgary)
A Second Chance? - A SoTL Project on Implementing Midterm Corrections in a Large Introductory-Level Statistics Course / Une seconde chance? - Un projet SoTL sur la mise en œuvre des corrections à mi-parcours dans un grand cours d'introduction à la statistique  
- 12:00-17:30 **Rajib Dey** (Memorial University of Newfoundland), **Noel Cadigan** (Fisheries and Marine Institute of Memorial University of Newfoundland), **Taraneh Abarin** (Memorial University of Newfoundland)
Sensitivity to Model Misspecification of a Von Bertalanffy Growth Model with Measurement Error in Age / Sensibilité à l'erreur de spécification d'un modèle de croissance de Von Bertalanffy avec erreur de mesure sur l'âge  
- 12:00-17:30 **Sahar Arshadi** (Memorial University of Newfoundland), **Taraneh Abarin** (Memorial University of Newfoundland)
On Identifiability of Regression Models with Interactions / Identifiabilité des modèles de régression avec interactions  
- 12:00-17:30 **Tyler Roick** (McMaster University), **Dimitris Karlis** (Athens University), **Paul D. McNicholas** (McMaster University)
Clustering Discrete Valued Time Series / Regrouper des séries chronologiques à valeurs discrètes  
- 12:00-17:30 **Prageeth Manujaya Senadeera** (Memorial University of Newfoundland), **Noel Cadigan** (Marine Institute of Memorial University of Newfoundland)
Local Influence Analysis of a State-Space Fish Stock Assessment Model Compared to Conventional Modelling Approaches / Analyse d'influence locale d'un modèle d'évaluation des stocks de poissons d'espace-état par rapport aux approches de modélisation conventionnelles  
- 12:00-17:30 **Jiajia Yue** (Memorial University of Newfoundland), **Candemir Cigsar** (Memorial University of Newfoundland)
A Power Comparison of Robust Tests for Monotone Trend in Recurrent Event Processes / Comparaison de la puissance de tests robustes de la tendance monotone dans des processus d'événements récurrents  
- 12:00-17:30 **Sarah Ricciuti** (McMaster University), **Paul D. McNicholas** (McMaster University)
Mixture Model Averaging for Clustering / Moyenne de modèles de mélange pour du regroupement  
- 12:00-17:30 **Angelina Pesevski** (McMaster University), **Brian Franczak** (MacEwan University), **Paul D. McNicholas** (McMaster University)
High Dimensional Clustering: tHDDC / Regroupement de grande dimension : tRDGD  
- 12:00-17:30 **Mohsen Soltanifar** (University of Toronto), **Annie Dupuis** (Sickkids Hospital, Toronto), **Russell Schachar** (Sickkids Hospital, Toronto), **Michael Escobar** (University of Toronto)
Stop Signal Reaction Time: A New Weighted Frequentist Estimation Method / Temps de réaction du signal d'arrêt : une nouvelle méthode d'estimation fréquentiste pondérée  
- 12:00-17:30 **Junchi Bin** (University of British Columbia - Okanagan), **Bryan Gardiner** (Data Nerds), **Zheng Liu** (University of British Columbia (Okanagan))
Advanced Regression Models for Automated Valuation Model / Modèles de régression avancés pour le modèle de valorisation automatisée  

- 12:00-17:30 **Tayler Dawn Scory** (University of Calgary), **Karen Kopciuk** (Alberta Health Services)
Genetic Risk Prediction for Type X Colorectal Cancer Families / Prédiction du risque génétique
du cancer colorectal familial de type X  
- 12:00-17:30 **René Ferland** (Université du Québec à Montréal), **Sorana Froda** (Université du Québec à
Montréal), **Anthony Coache** (Université du Québec à Montréal)
Modeling and Treatment of Surveillance Flu Data / Modélisation et traitement des données
de la surveillance de la grippe  

13:30-15:00 **Invited / Sur invitation** (abstract/résumé 160) **E2 125 (EITC)**







Statistical Modelling in General Insurance

Modélisation statistique en assurance générale

Chair/Président: Taehan Bae

Organizer/Responsable: Taehan Bae

Sponsor/Commanditaires: Actuarial Science Section / Groupe d'actuariat

- 13:30-14:00 **Lysa Porth** (University of Manitoba), **Ken Seng Tan** (University of Waterloo), **Wenjun
Zhu** (Nankai University)
A Relational Model for Predicting Farm-Level Crop Yield Distributions in the Absence of
Farm-Level Data / Un modèle relationnel pour la prédiction des répartitions du rendement
des cultures au niveau des fermes en l'absence de données agricoles  
- 14:00-14:30 **Lei Hua**, **Jianxi Su** (Purdue University)
Full-Range Tail Dependence Copulas / Étendue complète de la dépendance des ailes des co-
pules  
- 14:30-15:00 **Di (Cindy) Xu** (University of Nebraska-Lincoln), **David Landriault** (University of Waterloo),
Bin Li (University of Waterloo)
Aggregate Claim Analysis in a Two-sided Exit Setting with Dependence / Analyse de récla-
mations agrégées dans un contexte de sortie bilatéral avec dépendance  

13:30-15:00 **Invited / Sur invitation** (abstract/résumé 162) **E2 130 (EITC)**





New Directions in Quantitative Finance

Nouvelles directions en finance quantitative

Chair/Président: Adam Metzler

Organizer/Responsable: Adam Metzler

Sponsor/Commanditaires: Probability Section / Groupe de probabilité

- 13:30-14:15 **Ralf Kellner** (University of Regensburg)
Improving Risk Assessment using Market Expectations and Information from Model Risks /
Améliorer l'évaluation du risque à l'aide des renseignements et des attentes du marché prove-
nant des modèles de risque  
- 14:15-15:00 **Adam Metzler** (Wilfrid Laurier University)
Stochastic Correlation and Regulatory Capital / Corrélation stochastique et capital réglemen-
taire  

13:30-15:00 **Invited / Sur invitation** (abstract/résumé 164) **E2 110 (EITC)**







Recent Developments in Small Area Estimation

Développements récents en estimation pour petits domaines

Chair/Président: Mahmoud Torabi

Organizer/Responsable: Mahmoud Torabi

Sponsor/Commanditaires: Survey Section / Groupe des méthodes d'enquête

13:30-14:00	Sanjoy Kumar Sinha (Carleton University) Robust Small Area Estimation in Generalized Linear Models / Estimation robuste sur petits domaines dans des modèles linéaires généralisés  
14:00-14:30	Snigdhanu Chatterjee , Kaibo Gong (University of Minnesota), Megan Heyman (Rose-Hulman Institute of Technology), Taps Maiti (Michigan State University) Small Area Model Selection / Sélection de modèles sur petits domaines  
14:30-15:00	Gauri Datta (U.S. Bureau of the Census), Abhyuday Mandal (University of Georgia), Adrijo Chakraborty (NORC) Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model / Estimation bayésienne hiérarchique robuste pour petites régions pour modèle de régression à erreur emboîtée  



13:30-15:00 **Invited / Sur invitation** (abstract/résumé 166) **E2 105 (EITC)**

Improving Statistics at CIHR

Comment améliorer la statistique aux IRSC

Chair/Président: Lawrence C. McCandless

Organizer/Responsable: Lawrence C. McCandless

13:30-15:00	Lawrence C. McCandless (Simon Fraser University), Erica E.M. Moodie (McGill University), Russell Steele (McGill University), David A. Stephens (McGill University), Nancy Reid (University of Toronto) Panel discussion / Améliorer les statistiques dans les Instituts de recherche en santé du Canada (discussion en groupe)  
-------------	--







13:30-15:00 **Invited / Sur invitation** (abstract/résumé 167) **E3 270 (EITC)**

Design and Methodological Issues in Registry-based Randomized Controlled Trials

Problèmes de conception et de méthodologie dans les essais randomisés contrôlés sur registres

Chair/Président: Robert W. Platt

Organizer/Responsable: Tolulope T. Sajobi



13:30-14:00	Lehana Thabane (McMaster University) The Randomized Registry Trial — Opportunities, Challenges and Solutions / Essais sur registres randomisés : possibilités, défis et solutions  
14:00-14:30	Kevin E. Thorpe (University of Toronto) Registry-Based Randomized Controlled Trials: The Case for Explanatory-Pragmatic Continuum / Essais randomisés sur registres : le continuum explicatif-pragmatique  
14:30-15:00	Tolulope T. Sajobi (University of Calgary), Oluwagbohunmi Awosoga (University of Lethbridge), Bijoy Menon (University of Calgary), Michael Hill (University of Calgary), Lehana Thabane (McMaster University) Statistical Methods for Synthesizing Evidence from Explanatory and Pragmatic trials / Méthodes statistiques pour la synthèse de résultats d'essais explicatifs et pragmatiques  











13:30-15:00 **Contributed / Communications libres** (abstract/résumé 169) **E2 165 (EITC)**

Bayesian Methods and Applications

Méthodes bayésiennes et leurs applications

Chair/Président: Rob Deardon

13:30-13:45	Haoxuan Zhou (Simon Fraser University), Dave Campbell (Simon Fraser University) A Bayesian Approach to Diagnosis of the Gaussian Random Field Assumption / Une approche bayésienne pour le diagnostic de l'hypothèse de champ aléatoire gaussien  
-------------	---

13:45-14:00	Thierry Chekouo Tekougang (University of Minnesota Duluth), Francesco Stingo (University of Florence), Kim-Anh Do (The University of Texas MD Anderson Cancer Center) A Bayesian Approach for Pattern Detection in Extremely Small Sample Sizes: Application to Proteomic Cell Line Data of Human Leukemia / Une approche bayésienne pour la détection de motifs dans des tailles d'échantillon extrêmement petites : application aux données de la lignée cellulaire protéomique de la leucémie humaine  
14:00-14:15	Tian Li (Simon Fraser University), Lawrence C. McCandless (Simon Fraser University), X. Joan Hu (Simon Fraser University), Julian M. Somers (Simon Fraser University) Bayesian Sensitivity Analysis for Non-ignorable Missing Data in Longitudinal Studies / Analyse de sensibilité bayésienne pour les données manquantes non ignorables dans les études longitudinales  
14:15-14:30	Vineetha K. V. Warriyar (University of Calgary), Rob Deardon (University of Calgary) Individual Level Modelling of Infectious Disease Data: EpiILM / Modélisation au niveau individuel des données sur les maladies infectieuses : EpiILM  
14:30-14:45	Shaun Zheng Sun (University of the Fraser Valley), Richard Lockhart (Simon Fraser University) Bayesian Optimality for Beran's Class of Tests of Uniformity around the Circle / Optimalité bayésienne pour la classe de tests d'uniformité sur le cercle de Beran  
14:45-15:00	Chel Hee Lee (University of Calgary), Martha E. Lyon (University of Saskatchewan), Krista Baerg (University of Saskatchewan) Development of Quality Assurance Program for Jaundice Meter Management / Élaboration d'un programme d'assurance de la qualité pour la gestion des icromètres  

13:30-15:00 **Contributed / Communications libres** (abstract/résumé 172) **E2 160 (EITC)**

Complex Regression and Inference









Régression complexe et inférence

Chair/Président: Mohamed Belalia

13:30-13:45	Jenny Tieu (Brock University), Ramona Rat (Brock University), Mei Ling Huang (Brock University) A New Weighted Quantile Regression / Une nouvelle régression pondérée sur les quantiles  
13:45-14:00	Shahedul A. Khan (University of Saskatchewan) Generalized Bent-Cable Regression for Change point Data / Régression généralisée «bent-cable» pour les données de points de rupture  
14:00-14:15	Jeffrey L. Andrews (University of British Columbia) Combating Overfitting in Finite Mixture Models / Combattre le surajustement dans les modèles de mélanges finis  
14:15-14:30	Olawale Fatai Ayilara (University of Manitoba), Mohammad Jafari Jozani (University of Manitoba) Quantile Regression with Nominated Samples / Régression par les quantiles avec des échantillons nommés  
14:30-14:45	Md. Shakhawat Hossain (University of Winnipeg), Mina Norouzirad (Shahrood University of Technology), Mohammad Arashi (Shahrood University of Technology) Estimation Strategies for Weighted Least Absolute Deviations Regression Models / Stratégies d'estimation pour les modèles de régression des moindres écarts absolus pondérés  
14:45-15:00	Jinhong You Estimation and Identification of Varying-Coefficient Additive Models for Locally Stationary Process / Estimation et identification de modèles additifs à coefficients variables pour des processus localement stationnaires  









13:30-15:00 **Contributed / Communications libres** (abstract/résumé 176) **E2 150 (EITC)**
Ecology and Geological Statistics**Écologie et statistique géologique**

Chair/Président: Connie Stewart

- 13:30-13:45 **Khurram Nadeem** (Natural Resources Canada), **Charmaine B. Dean** (Western University), **Douglas Woolford** (Sciences, Western University), **Steve Taylor** (Natural Resources Canada)
Predicting the Unpredictable: Severe Wildfire Risk Prediction in Canada / Prévoir l'imprévisible : prévision du risque de feux de friches dévastateurs au Canada  
- 13:45-14:00 **Jabed H. Tomal** (University of Toronto Scarborough), **Jan J.H. Ciborowski** (University of Windsor), **Karen Fung** (University of Windsor)
Piecewise Linear Quantile Regression for Estimation of Ecological Breakpoints / Régression quantile linéaire par parties pour l'estimation de points de rupture écologiques  
- 14:00-14:15 **Matias Salibian-Barrera** (University of British Columbia), **Daniel Dinsdale** (University of British Columbia)
Geostatistical Inference under Preferential Sampling / Inférence géostatistique sous échantillonnage préférentiel  
- 14:15-14:30 **Michelle Carey** (McGill University), **James O. Ramsay** (McGill University)
Parameter Estimation for PDEs Defined over Irregular Domains / Estimation des paramètres pour des équations aux dérivées partielles définies dans des domaines irréguliers  

13:30-15:00 **Contributed / Communications libres** (abstract/résumé 179) **E2 155 (EITC)**
Time Series and Longitudinal Data**Séries chronologiques et données longitudinales**

Chair/Président: Louis-Paul Rivest

- 13:30-13:45 **Gregory Rice** (University of Waterloo), **Lajos Horvath** (University of Utah)
Empirical Eigenvalue Based Testing for Structural Breaks in Linear Factor Models / Test empirique fondé sur la valeur propre des ruptures structurelles dans les modèles factoriels linéaires  
- 13:45-14:00 **Md. Erfanul Hoque** (University of Manitoba), **Mahmoud Torabi** (University of Manitoba)
Modeling the Random Effects Covariance Matrix for Longitudinal Data with Covariate Measurement Error / Modélisation de la matrice de covariance des effets aléatoires pour données longitudinales avec erreur de mesure des covariables  
- 14:00-14:15 **Lichen Chen** (University of Waterloo), **Adam Kolkiewicz** (University of Waterloo), **Tony Wirjanto** (University of Waterloo)
Empirical Likelihood Multiscale Inference of the Scaling Property of Time Series Models / Inférence multi-échelle de la vraisemblance empirique de la propriété d'échelle des modèles de séries chronologiques  
- 14:15-14:30 **Yuanhao Lai** (Western University), **A. Ian McLeod** (Western University)
A Nonparametric Test for Detecting Periodicity in Time Series / Un test non paramétrique pour détecter la périodicité dans les séries chronologiques  
-

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 182) **E2 110 (EITC)**




Recent Developments in Joint Modelling of Mixed Types of Outcomes

Développements récents dans la modélisation conjointe de types de résultats mixtes

Chair/Président: Xin (Cindy) Feng

Organizer/Responsable: Xin (Cindy) Feng

Sponsor/Commanditaires: Biostatistics Section / Groupe de biostatistique

- 15:30-16:00 **Charmaine B. Dean** (University of Western Ontario)
 Joint Analysis of Longitudinal Zero-heavy Panel Count Outcomes with Application to Understanding Desistance from Criminal Activity / Analyse conjointe de variables de dénombrement longitudinales à surreprésentation de zéros avec application à la compréhension du renoncement de la criminalité 
- 16:00-16:30 **Dongsheng Tu, Hui Song** (Dalian University of Technology), **Yingwei Peng** (Queen's University)
 Joint Modeling of Longitudinal Proportional Measurements and Survival Times with a Cure Fraction / Modélisation conjointe de mesures proportionnelles longitudinales et de temps de survie avec fraction de guérison 
- 16:30-17:00 **Elizabeth Juarez-Colunga** (University of Colorado Denver), **Brandie Wagner** (University of Colorado Denver), **Edith Zemanick** (University of Colorado Denver)
 Joint Analysis of a Longitudinal Binary Outcome and Recurrent Events with Application to Cystic Fibrosis Outcomes / Analyse conjointe d'une variable longitudinale binaire et d'événements récurrents avec une application sur la fibrose kystique 

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 184) **E2 130 (EITC)**




New Vistas in Applied Probability

Nouvelles perspectives en probabilité appliquée

Chair/Président: Gail Ivanoff

Organizer/Responsable: Gail Ivanoff

Sponsor/Commanditaires: Probability Section / Groupe de probabilité

- 15:30-16:00 **Yaniv Plan** (University of British Columbia), **Chris Liaw** (University of British Columbia), **Abbas Mehrabian** (University of British Columbia), **Roman Vershynin** (University of Michigan)
 A Simple Tool for Bounding the Deviation of Random Matrices on Geometric Sets, with Applications / Un outil simple pour borner la déviation de matrices aléatoires sur des ensembles géométriques; avec applications 
- 16:00-16:30 **Clarence Simard** (Université du Québec à Montréal)
 Approximation of Squared Integrable Random Variables with Nonlinear Stochastic Integrals / Approximation d'une variable aléatoire de carré intégrable par une intégrale stochastique non linéaire. 
- 16:30-17:00 **Yi Shen** (University of Waterloo)
 Probabilistic Symmetries and Random Locations / Symétries probabilistes et emplacements aléatoires 

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 186) **E2 125 (EITC)**

How to Stay Sane Pre-tenure: Addressing the Challenges of New Investigators

Comment rester sain d'esprit jusqu'à l'agrégation : les défis des nouveaux chercheurs

Chair/Président: Nathaniel T. Stevens

Organizer/Responsable: Nathaniel T. Stevens


Sponsor/Commanditaires: Committee on New Investigators / Comité des nouveaux chercheurs


15:30-17:00 **Nathaniel T. Stevens** (University of San Francisco), **Hua Shen** (University of Calgary), **Mireille E. Schnitzer** (Université de Montréal), **Jingjing Wu** (University of Calgary), **Erica E.M. Moodie** (McGill University)
Panel discussion / Discussion de groupe 


15:30-17:00 **Invited / Sur invitation** (abstract/résumé 187) **E3 270 (EITC)**

Statistical Methods for Omics Data
Méthodes statistiques pour les données “-omiques”

Chair/Président: Pingzhao Hu
Organizer/Responsable: Pingzhao Hu

15:30-16:00 **Wei Xu**
Statistical Methodology on Human Microbiome Data Analysis / Méthodologie statistique sur l'analyse de données sur le microbiome 


16:00-16:30 **Elif Fidan Acar** (University of Manitoba)
Aspects of Genetic Meta-Analysis under Data Uncertainty / Aspects de la meta-analyse de la génétique dans le cadre de l'incertitude entourant les données 


16:30-17:00 **David R. Bickel** (University of Ottawa)
Correcting False Discovery Rates for their Bias Toward False Positives / Correction des taux de fausses découvertes pour leur biais à l'égard des faux positifs 


15:30-17:00 **Invited / Sur invitation** (abstract/résumé 189) **E2 105 (EITC)**

Brewing Insight from Beer Data: In Honour of Gosset's Birthday
La sagesse est dans la bière et ses données : en l'honneur de l'anniversaire de Gosset

Chair/Président: Jason Loeppky and David Campbell
Organizer/Responsable: Jason Loeppky and David Campbell
Sponsor/Commanditaires: Accreditation Committee / Comité d'accréditation

15:30-16:00 **Pulindu Ratnasekera** (Simon Fraser University)
Geographic Topic Modelling to Predict Regional Beer Success / Modélisation géographique pour prévoir le succès d'une bière selon la région 


16:00-16:30 **Joanna Zhao** (Simon Fraser University), **Derek Qiu** (Simon Fraser University)
How to Brew your Beer Tour with Data Science / Préparer une tournée des brasseries grâce à la science des données 





16:30-17:00 **James A. Hanley** (McGill University)
Gosset at Guinness: How 'Student' Derived (and Checked) the Distribution of his z-statistic / W.S. Gosset, « Student », chez Guinness : comment il a obtenu (et vérifié) la distribution de sa statistique Z 

15:30-17:00 **Invited / Sur invitation** (abstract/résumé 191) **E2 150 (EITC)**

Applied Survey Research with Impact
Recherche appliquée sur les sondages avec un impact

Chair/Président: Matthias Schonlau
Organizer/Responsable: Matthias Schonlau
Sponsor/Commanditaires: Survey Section / Groupe des méthodes d'enquête

15:30-16:00 **Kristen Himelein** (World Bank)
Innovations in Sampling for Insecure and Data-Poor Environments / Innovations dans l'échantillonnage pour les environnements non sécuritaires et pauvres en données 

- 16:00-16:30 **Kim P. Huynh** (Bank of Canada)
Measuring Cash and Payment Card Usage at the Point-of-Sale for Consumers and Merchants / Mesure de l'utilisation des cartes de débit et de paiement au point de vente pour les consommateurs et les marchands  
- 16:30-17:00 **Robert Clark** (University of Wollongong, New Zealand), **Steven Johnston** (New Zealand Ministry of Health)
How the Design of the New Zealand Health Survey Contributes to Informed Decision Making in Public Health Policy / Comment la conception de l'Enquête sur la santé de Nouvelle-Zélande contribue à des décisions éclairées en matière de politique de santé publique  

15:30-17:00 **Contributed / Communications libres** (abstract/résumé 193) **E2 165 (EITC)**

Causality and Design
Causalité et conception





Chair/Président: Mireille E. Schnitzer

- 15:30-15:45 **Po Yang** (University of Manitoba), **William Li** (University of Minnesota)
Foldover Designs with Column Permutations / Plans de repliement avec permutation de colonnes  
- 15:45-16:00 **Katherine Dagnault** (University of Toronto), **Olli Saarela** (University of Toronto)
A Doubly Robust Estimator for Indirectly Standardized Mortality Ratios / Un estimateur doublement robuste pour les ratios de mortalité indirectement standardisés  
- 16:00-16:15 **Shomoita Alam** (McGill University), **Erica E.M. Moodie** (McGill University), **David A. Stephens** (McGill University)
Should a Propensity Score be Super? A Comparison of Estimation Methods / Un score de propension devrait-il être super? Comparaison de méthodes d'estimation  
- 16:15-16:30 **Lin Liu** (University of California, San Diego), **Loki Natarajan** (University of California, San Diego), **Karen Messer** (University of California, San Diego)
Estimation in Model Selection after Imputation / Estimation dans la sélection du modèle après imputation  
- 16:30-16:45 **Rui Hu** (MacEwan University), **Doug Wiens** (University of Alberta)
Robust Discrimination Designs over Hellinger Neighbourhoods / Plans de discrimination robuste sur les voisinages de Hellinger  
- 16:45-17:00 **Saumen Mandal** (University of Manitoba), **Monsur Chowdhury** (University of Winnipeg)
Optimal Designs for Estimating Parameters of the Bradley-Terry Model for Paired Comparisons / Plans optimaux pour l'estimation des paramètres du modèle de Bradley-Terry pour des comparaisons par paires  

15:30-17:00 **Contributed / Communications libres** (abstract/résumé 196) **E2 160 (EITC)**

New Methods in Classification
Nouvelles méthodes en classification

Chair/Président: Paul D. McNicholas

- 15:30-15:45 **Hyukjun (Jay) Gweon** (Waterloo University), **Matthias Schonlau** (University of Waterloo), **Stefan Steiner** (University of Waterloo)
An Overview of Existing and a Novel Approaches to Multi-label Classification / Aperçu d'une nouvelle approche et d'approches existantes de classification multi-étiquettes  
- 15:45-16:00 **Armin Hatefi** (University of Toronto), **Nancy Reid** (University of Toronto), **Mohammad Jafari Jozani** (University of Manitoba), **Omer Ozturk** (Ohio State University)
Estimation and Classification for Finite Mixture Models with Order Statistics / Estimation et classification pour les modèles de mélange fini avec des statistiques d'ordre  



- 16:00-16:15 **Wanhua Su** (MacEwan University)
Two Innovative Applications of Classification and Regression Trees / Deux applications innovantes des arbres de classification et de régression  
- 16:15-16:30 **Pingzhao Hu** (University of Manitoba), **Md. Mohaiminul Islam** (University of Manitoba), **Yang Wang** (University of Manitoba)
A deep learning-based integrative analysis framework for molecular classification of breast cancer / Analyse intégrative fondée sur le deep learning pour la classification moléculaire du cancer du sein  
- 16:30-16:45 **Jian Yang** (University of Calgary), **Jingjing Wu** (University of Calgary), **Haocheng Li** (Tom Baker Cancer Centre and University of Calgary), **Xiaolan Feng** (Tom Baker Cancer Centre and University of Calgary)
Minimum Profile Hellinger Distance Estimation for Two-Class Location Models / Estimation de la distance du profil minimal d'Hellinger pour les modèles d'emplacement à deux classes  
- 16:45-17:00 **Anita Brobbey** (University of Calgary), **Tyler Williamson** (University of Calgary), **Samuel Wiebe** (University of Calgary), **Lisa M. Lix** (University of Manitoba), **Tolulope T. Sajobi** (University of Calgary)
Comparison of Multivariate Classifiers in the Presence of Covariates in Repeated Measures Designs / Comparaison de classificateurs multivariés en présence de covariables dans les modèles de mesures répétées  

15:30-17:00 **Contributed / Communications libres** (abstract/résumé 199) **E2 155 (EITC)**

Insuring, Investing, and Paying Up

Assurer, investir et payer

Chair/Président: Cary Chi-Liang Tsai

- 15:30-15:45 **Jean-François Bégin** (Simon Fraser University)
Deflation Risk and Implications for Life Insurers / Risque de déflation et conséquences pour les assureurs-vie  
- 15:45-16:00 **Valéry Dongmo Jiongo** (Bank of Canada)
Outlier Robust Estimation of the Total Private Cost of Payment Methods for Large Businesses / Estimation robuste des valeurs aberrantes du coût total privé des méthodes de paiement pour les grandes entreprises  
- 16:00-16:15 **Suxin Wang** (Simon Fraser University), **Yi Lu** (Simon Fraser University), **Barbara Sanders** (Simon Fraser University)
Optimal Investment Strategies and Inter-Generational Risk-Sharing for Target Benefit Pension Plans / Stratégies d'investissement optimales et partage du risque entre générations pour régimes de retraite à prestations cibles  
- 16:15-16:30 **Francois Watier** (Université du Québec à Montréal), **Yassin El Kasmi** (Université du Québec à Montréal)
Constrained Mean-Variance Portfolio in a Market with Stochastic Volatility / Portefeuille à variance moyenne sous contraintes dans un marché à volatilité stochastique  
- 16:30-16:45 **Etienne Marceau** (Université de Laval), **Hélène Cossette** (Université Laval)
Portfolio of Exchangeable Risks: Aggregation with Partial Information on Dependence / Portefeuille de risques échangeables : agrégation avec information partielle sur la dépendance  
- 16:45-17:00 **Yi Lu** (Simon Fraser University), **Shuanming Li** (University of Melbourne)
On the Moments and the Distribution of Aggregate Discounted Claims in a Markovian Environment / Moments et répartition des déclarations de sinistres totales actualisés dans un environnement markovien  

Wednesday June 14

Mercredi 14 juin


08:40-09:45 **Invited / Sur invitation** (abstract/résumé 202) **E3 270 (EITC)**

CRM-SSC Prize in Statistics Address

Allocution de la récipiendaire du Prix CRM-SSC en statistique

Chair/Président: Hugh Chipman

Organizer/Responsable: Alejandro Murua

08:40-09:45 **Lei Sun** (University of Toronto)
 The Power of Simple Statistical Techniques in the Era of Big and Complex Data: Some Recent Examples from Genetic Association Studies / La puissance des techniques statistiques simples à l'ère des données volumineuses et complexes : quelques exemples récents d'études d'association génétique 

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 203) **E2 150 (EITC)**


Showcase: Graduate Student Research in Actuarial Sciences


Vitrine : Recherches d'étudiants gradués en sciences actuarielles


Chair/Président: Mathieu Boudreault


Organizer/Responsable: Mathieu Boudreault

Sponsor/Commanditaires: Actuarial Science Section / Groupe d'actuariat

10:20-10:42 **Hongcan Lin** (University of Waterloo), **David Saunders** (University of Waterloo), **Chengguo Weng** (University of Waterloo)
 Optimal Investment Strategies for Participating Contracts / Stratégies d'investissement optimales pour contrats avec participation 

10:42-11:04 **Ihsan Chaoubi** (Université Laval), **Hélène Cossette** (Université Laval), **Etienne Marceau** (Université Laval)
 Archimedean Copulas through Multivariate Gamma Distributions / Copules archimédiennes par distributions gamma multivariées 

11:04-11:26 **Lidan (Lucy) Zhang** (Simon Fraser University), **Yi Lu** (Simon Fraser University)
 A Multi-State Model for a Life Insurance Product with Integrated Health Rewards Program / Modèle multi-états pour produits d'assurance-vie avec programme de récompenses santé intégré 

11:26-11:50 **Ruixi Zhang** (University of Western Ontario), **Kristina Sendova** (University of Western Ontario)
 On Total Dividends under a Threshold Strategy / Total des dividendes sous une stratégie de seuil 

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 205) **E2 105 (EITC)**


Technology in the Classroom





La technologie en salle de classe

Chair/Président: Anne-Sophie Charest

Organizer/Responsable: Anne-Sophie Charest

Sponsor/Commanditaires: Statistical Education Section / Groupe d'éducation en statistique

10:20-10:50 **Jim Stallard** (University of Calgary), **Scott Robison** (University of Calgary)
 A More Dynamic Lab Experience: Implementation of Weekly Automated and Teaching Assistant Supported Lab Exercises to Increase Student Engagement through Small-Step Learning / Expérience de laboratoire plus dynamique : mise en œuvre d'exercices en laboratoire hebdomadaires automatisés et soutenus par un assistant à l'enseignement afin d'augmenter l'engagement des étudiants par l'apprentissage par petites étapes 

- 10:50-11:20 **Jenna G. Tichon** (University of Manitoba)
Educreations: Video Tools for Statistical Teaching / Educreations : outils vidéo pour l'enseignement des statistiques  
- 11:20-11:50 **Thierry Duchesne** (Université Laval)
Automated Homework Generation and Grading with R / Génération et correction automatisées de devoir avec R  

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 207) **E2 130 (EITC)**







Imputation of Multivariate Missing Data

Imputation de données manquantes multivariées

Chair/Président: Matthias Schonlau

Organizer/Responsable: Caren Hasler

Sponsor/Commanditaires: Survey Section / Groupe des méthodes d'enquête

- 10:20-10:50 **Jae Kwang Kim** (Iowa State University), **Shu Yang** (North Carolina State University)
Predictive Mean Matching Imputation in Survey Sampling / L'imputation par appariement d'après la moyenne prévisionnelle en échantillonnage  
- 10:50-11:20 **Louis-Paul Rivest** (Université Laval), **Caren Hasler** (University of Toronto), **Radu Craiu** (University of Toronto)
Vine Copulas for Imputation of Monotone Non-response / Les copules en vigne pour l'imputation d'une non-réponse monotone  
- 11:20-11:50 **Yves Tillé** (Université de Neuchâtel), **Audrey-Anne Vallée** (Université de Neuchâtel)
Balanced Imputation for Swiss Cheese Nonresponse / Imputation équilibrée pour la non-réponse en fromage suisse  







10:20-11:50 **Invited / Sur invitation** (abstract/résumé 209) **E2 125 (EITC)**

Optimal Design of Experiments and Inference

Conception optimale d'expériences et inférence

Chair/Président: Yanqing Yi

Organizer/Responsable: Yanqing Yi

- 10:20-10:50 **Michael Wallace** (University of Waterloo), **Erica E.M. Moodie** (McGill University), **David A. Stephens** (McGill University)
Dynamic Treatment Regimes via Reward Ignorant Modelling / Régimes de traitement dynamiques par modélisation aveugle au résultat  
- 10:50-11:20 **Nancy Flournoy** (University of Missouri)
Issues with Inference in Adaptive Optimal Design / Problèmes d'inférence concernant les plans adaptatifs optimaux  
- 11:20-11:50 **Xiaojuan Xu** (Brock University), **Sanjoy Kumar Sinha** (Carleton University)
Model-Robust Designs for Generalized Linear Mixed Models with Possible Misspecification / Plans robustes vis-à-vis du modèle pour modèles mixtes linéaires généralisés avec erreurs de spécification éventuelles  



10:20-11:50 **Invited / Sur invitation** (abstract/résumé 224) **E2 110 (EITC)**

Recent Advances in Longitudinal and Incomplete Data Analysis

Progrès récents en analyse de données longitudinales et incomplètes

Chair/Président: Haocheng Li and Hua Shen

Organizer/Responsable: Ying Yan

10:20-10:42	Michael A. McIsaac (Queen's University), Lauren Paul (Queen's University) Multiple Imputation of Incomplete Accelerometer Data / Imputation multiple des données incomplètes d'un accéléromètre  
10:42-11:04	Hua Shen (University of Calgary), Ying Yan (University of Calgary), Haocheng Li (University of Calgary) Analysis of Heterogeneous Lifetime Data with Measurement Error in Covariate / Analyse de données hétérogènes sur la durée de vie avec erreurs de mesure dans les covariables  
11:04-11:26	Haocheng Li (University of Calgary), Grace Yi (University of Waterloo) Missing Data Mechanisms for Incomplete Longitudinal Observations in both Responses and Covariates / Mécanismes de données manquantes pour des observations longitudinales incomplètes dans les réponses et les covariables  
11:26-11:50	Mireille E. Schnitzer (Université de Montréal), Steve Ferreira Guerra (Université de Montréal), Amélie Forget (Université de Montréal), Lucie Blais (Université de Montréal) A Self-Selecting Procedure for the Optimal Discretization of the Timeline for Longitudinal Causal Inference Methods with Electronic Health Data / Une procédure de sélection automatique pour la discrétisation optimale de la ligne de temps pour l'application de méthodes longitudinales en inférence causale avec données électroniques sur la santé  

10:20-11:50 **Invited / Sur invitation** (abstract/résumé 211) **E3 270 (EITC)**









Data Fusion, Network Meta-Analysis, and Causal Inference

Fusion de données, méta-analyse en réseau et inférence causale

Chair/Président: Mireille E. Schnitzer and Russell Steele

Organizer/Responsable: Mireille E. Schnitzer

Sponsor/Commanditaires: Biostatistics Section / Groupe de biostatistique

10:20-10:42	Elias Bareinboim (Purdue University) Causal Inference from Big Data: Theoretical Foundations and the Data-Fusion Problem / Inférence causale à partir de données volumineuses : les fondements théoriques et le problème de la fusion des données  
10:42-11:04	Russell Steele (McGill University), Mireille E. Schnitzer (Université de Montréal), Ian Shrier (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University) Breaking the Myth of Breaking Randomization: A Causal Examination of Arm-Based Meta-Analysis / Détruire le mythe de la détérioration la randomisation : un examen causal de la méta-analyse à volets  
11:04-11:26	Qinshu Lian (University of Minnesota-Twin Cities), Haitao Chu (University of Minnesota-Twin cities) A Bayesian HSROC Model for Network Meta-analysis of Diagnostic Tests / Modèle bayésien de résumé hiérarchique ROC pour la méta-analyse en réseau de tests diagnostiques  
11:26-11:50	Christopher Schmid (Brown University), Youdan Wang (Brown University) Hierarchical Models for Combining N-of-1 Trials / Modèles hiérarchiques pour combiner des essais individuels (N-de-1)  

10:20-11:50 **Contributed / Communications libres** (abstract/résumé 214) **E2 165 (EITC)**

Biological Data in Action

Données biologiques en action

Chair/Président: Dongsheng Tu









10:20-10:35	Farhad Shokoohi (McGill University), Celia M.T. Greenwood (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University), David A. Stephens (McGill University), Aur�lie Labbe (HEC Montr�al) A Novel Hidden Markov Model Approach for Differentially Methylated Region Identification in DNA Methylation Data / Une nouvelle approche de mod�le de Markov cach� pour l'identification de r�gions diff�remment m�thyl�es pour des donn�es de m�thylation de l'ADN  
10:35-10:50	Julien St-Pierre (Universit� du Qu�bec � Montr�al), Karim Oualkacha (Universit� du Qu�bec � Montr�al) A Multivariate Rare-Variant Association Test for Non-Normally Distributed Phenotypes / Test d'association multivari� entre variants rares et ph�notypes de distributions non normales  
10:50-11:05	Angelo J. Canty (McMaster University), Keelin Greenlaw (Lady Davis Institute, Jewish General Hospital, Montreal), Celia M.T. Greenwood (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University) Measurement and Association Testing of 5-hydroxymethylcytosine (5hmC) Methylation / Mesure et test d'association de la m�thylation 5-hydroxymethylcytosine (5hmC)  
11:05-11:20	Camila P.E. de Souza (University of British Columbia), Mirela Andronesu (University of British Columbia), Sohrab Shah (University of British Columbia) A Probabilistic-Based Approach for Clustering Sparse Single-Cell DNA Methylation Data / Approche probabiliste pour le regroupement de donn�es de m�thylation d'ADN d'une seule cellule  
11:20-11:35	Elaheh Torkashvand (University of Waterloo), Joel Dubin (University of Waterloo), Gregory Rice (University of Waterloo) Modeling T-cell Trajectory Data using Hidden Markov Models / Mod�lisation de donn�es de trajectoires de cellules T en utilisant des mod�les de Markov cach�s  
11:35-11:50	Karim Oualkacha (Universit� du Qu�bec � Montr�al), Mohamed Ouhourane (Universit� du Qu�bec � Montr�al), Yi Yang (McGill University) Block-Wise Descent Algorithms for Group Variable-selection in Quantile Regression / Algorithmes de descente par bloc pour la s�lection d'un groupe de variables dans la r�gression quantile  



10:20-11:50 **Contributed / Communications libres** (abstract/r sum  218) **E2 160 (EITC)**

Computational Methods

M thodes computationnelles

Chair/Pr sident: Lawrence C. McCandless

10:20-10:35	Fran�ois Alastair Marshall Mode Detection in a Non-Stationary Environment / D�tection de modes dans un environnement non stationnaire  
10:35-10:50	Samira Soleymani (Western University), A. Ian McLeod (Western University) Box-Cox Time Series / S�ries temporelles de Box-Cox  
10:50-11:05	Yilei Wu (University of Waterloo), Yingli Qin (University of Waterloo), Mu Zhu (University of Waterloo) An Extended Compound Symmetry Model for High-dimensional Covariance Matrix Estimation / Un mod�le �largi de sym�trie compos�e pour l'estimation d'une matrice de covariance � haute dimension  
11:05-11:20	Waleed Almutiry (Calgary University), Rob Deardon (Calgary University) Incorporating Contact Network Uncertainty in Individual-Level Models of Infectious Disease using Approximate Bayesian Computation / Int�grer l'incertitude d'un r�seau de contact dans les mod�les de maladies infectieuses au niveau des individus � l'aide d'un calcul bay�sien approximatif  



11:20-11:35 **Luc Villandré** (McGill University), **Aurélié Labbe** (HEC Montréal), **Bluma Brenner** (McGill AIDS Centre, Lady Davis Institute, Jewish General Hospital, Montreal), **Michel Roger** (Université de Montréal), **David A. Stephens** (McGill University)
 DM-PhyClus: A Bayesian Phylogenetic Clustering Algorithm Incorporating Dirichlet-Multinomial Priors / DM-PhyClus : Une approche bayésienne de groupement phylogénétique basée sur des a priori Dirichlet-multinomiaux  



10:20-11:50 **Contributed / Communications libres** (abstract/résumé 221) **E2 155 (EITC)**



Finance, Time Series, and Extremes



Finance, séries chronologiques et extrêmes



Chair/Président: René Ferland



10:20-10:35 **Kaiqiong Zhao** (McGill University), **Elif Fidan Acar** (University of Manitoba)
 Conditional Dependence Models under Covariate Measurement Error / Modèles de dépendance conditionnelle avec erreur de mesure des covariables  

10:35-10:50 **Janaki Thakshila Koralage** (Memorial University of Newfoundland), **Zhaozhi Fan** (Memorial University of Newfoundland)
 Localized Quantile Regression of Realized Volatility / Régression quantile localisée de la volatilité réalisée  

10:50-11:05 **Gary Sneddon** (Mount Saint Vincent University), **Tariqul Hasan** (University of New Brunswick), **Renjun Ma** (University of New Brunswick)
 Modelling Bivariate Count Time Series with Extra Zeros / Modélisation de séries chronologiques de dénombrement bivariées avec surreprésentation de zéros  

11:05-11:20 **Nathalie Akakpo** (Université Pierre et Marie Curie, Paris)
 A Convex Regression Approach to Pickands Dependence Function Estimation / Estimation de la fonction de dépendance de Pickands par une approche de type régression convexe  

11:20-11:35 **Debbie J. Dupuis** (HEC Montréal), **Marco Bee** (University of Trento), **Luca Trapin** (Scuola Normale Superiore)
 Realizing the Extremes: Estimation of Tail-Risk Measures from a High-Frequency Perspective / Réaliser les extrêmes : estimation des mesures de risque de queue à partir d'une perspective à haute fréquence  

11:35-11:50 **Jooyoung Lee** (University of Waterloo), **Richard J. Cook** (University of Waterloo)
 Copula Dependence Modeling for Multi-type Recurrent Events using Composite Likelihood / Modélisation de la dépendance des copules pour les différents types d'événements récurrents au moyen de la vraisemblance composite  

13:30-14:30 **Invited / Sur invitation** (abstract/résumé 227) **E2 110 (EITC)**



Isobel Loutit Lecture

Allocution invitée Isobel Loutit

Chair/Président: François Bellavance

Organizer/Responsable: François Bellavance

Sponsor/Commanditaires: Business and Industrial Statistics Section / Groupe de statistique industrielle et de gestion



13:30-14:30 **Shirley E. Mills** (Carleton University)
 Predictive Analytics in Business / Analytique prévisionnelle en affaires  

13:30-14:15 **Invited / Sur invitation** (abstract/résumé 228) **E3 270 (EITC)**

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

Chair/Président: Richard Lockhart

Organizer/Responsable: Richard Lockhart



13:30-14:15 **Bryan E. Shepherd** (Vanderbilt University School of Medicine), **Chun Li** (Case Western Reserve University), **Qi Liu** (Merck)
 Probability-Scale Residuals for Continuous, Discrete, and Censored Data / Résidu lié à une échelle de probabilités pour données continues, discrètes et tronquées  

14:15-15:00 **Invited / Sur invitation** (abstract/résumé 229) **E3 270 (EITC)**

Pierre Robillard Address**Allocution du récipiendaire du Prix Pierre-Robillard**

Chair/Président: Gordon Fick

Organizer/Responsable: Gordon Fick

14:15-15:00 **Andy Leung** (University of British Columbia), **Victor J. Yohai** (Universidad de Buenos Aires), **Ruben H. Zamar** (University of British Columbia)
 Robust Estimation under Cellwise and Casewise Contamination / Estimation robuste en cas de contamination « cellwise » ou « casewise »  



15:00-16:30 **Invited / Sur invitation** (abstract/résumé 230) **E3 270 (EITC)**



Recent Advances in Compositional Data Analysis and Applications**Progrès récents en analyse de données de composition et ses applications**

Chair/Président: Connie Stewart

Organizer/Responsable: Connie Stewart

15:00-15:30 **Holly N. Steeves** (Dalhousie University)
 Maximum Likelihood Estimation of Marine Predators' Diet Proportions / Estimation du maximum de vraisemblance des proportions alimentaires des prédateurs marins  

15:30-16:00 **Janice Lea Scealy** (Australian National University), **Alan Welsh** (Australian National University)
 Estimation and Inference in Directional Mixed Models for Compositional Data / Estimation et inférence dans les modèles mixtes directionnels pour données relatives à la composition  

16:00-16:30 **Michail Tsagris** (University of Crete), **Connie Stewart** (University of New Brunswick)
 A Dirichlet Regression Model for Compositional Data with Zeros / Un modèle de régression de Dirichlet pour des données compositionnelles avec zéros  



15:00-16:30 **Invited / Sur invitation** (abstract/résumé 233) **E2 105 (EITC)**





University of Manitoba Department of Statistics: 50th Anniversary Showcase**Département de statistique de l'Université du Manitoba : vitrine sur 50 ans d'histoire**

Chair/Président: Liquin Wang

Organizer/Responsable: Liquin Wang

Sponsor/Commanditaires: Business and Industrial Statistics Section / Groupe de statistique industrielle et de gestion

15:00-15:30 **Jerald F. Lawless** (University of Waterloo)
 Statistics, Science and Technology / Statistiques, science et technologie  

- 15:30-16:00 **John F. Brewster** (University of Manitoba)
Reflections on Interactions with Industry / Réflexions sur les interactions avec l'industrie  
- 16:00-16:30 **Brian Macpherson** (retired), **Smiley Cheng** (University of Manitoba)
Industrial Quality Improvement Workshops and Applications / Ateliers d'amélioration de la qualité et ses applications  



15:00-16:30 **Invited / Sur invitation** (abstract/résumé 235) **E2 110 (EITC)**

Waterloo at 50: Reflections of Former Students

Waterloo à 50 ans

Chair/Président: Mary Thompson

Organizer/Responsable: Mary Thompson

- 15:00-16:30 **Mary E. Thompson** (University of Waterloo), **Vern Farewell** (University of Cambridge), **Wayne Oldford** (University of Waterloo), **Karen Kopciuk** (University of Calgary), **Robert L. Brown** (University of Waterloo)
Panel discussion / Les 50 ans de l'Université de Waterloo : réflexions d'anciens étudiants (table ronde)  









15:00-16:30 **Invited / Sur invitation** (abstract/résumé 236) **E2 130 (EITC)**

Recent Developments in Copula Modelling: A Showcase Session of the CANSSI-CRT on 'Copula Dependence Modelling: Theory and Applications'

Développements récents en modélisation par les copules : vitrine sur le PRC de l'INCASS sur "Modélisation de la dépendance par les copules : théorie et applications"

Chair/Président: Elif Fidan Acar

Organizer/Responsable: Elif Fidan Acar

- 15:00-15:22 **Marie-Hélène Toupin** (Université Laval), **Jean-François Quessy** (Université du Québec à Trois-Rivières), **Louis-Paul Rivest** (Université Laval)
The Family of Chi-square Copulas and Applications in Spatial Statistics / La famille de copules khi-carré et son utilisation en statistique spatiale  
- 15:22-15:45 **Bo Chang** (University of British Columbia), **Harry Joe** (University of British Columbia)
Vine Copula Regression / Régression de copules en vigne  
- 15:45-16:07 **Samuel Perreault** (Université Laval), **Thierry Duchesne** (Université Laval), **Johanna Neslehova** (McGill University)
Improved Estimation and Identification of Block-Exchangeable Dependence Structures / Estimation améliorée et identification de structures de dépendance échangeables par blocs  
- 16:07-16:30 **Mohamed Belalia** (Université du Québec à Trois-Rivières), **Jean-François Quessy** (University of Quebec at Trois-Rivières)
Generalized Simulated Method-of-Moments for Copula Parameters of Arbitrary Dimension / Méthode des moments généralisée simulée pour des paramètres de copules de dimension arbitraire  







15:00-16:30 **Invited / Sur invitation** (abstract/résumé 239) **E2 125 (EITC)**

Recent Developments in the Analysis of Neuroimaging Data: A Showcase of the CANSSI-CRT on 'Joint Analysis of Neuroimaging Data'

Développements récents en analyse des données de neuroimagerie : vitrine sur le PRC de l'INCASS sur "l'analyse conjointe de données de la neuroimagerie"

Chair/Président: Farouk Nathoo

Organizer/Responsable: Farouk Nathoo

15:00-15:30	Dehan Kong (University of Toronto), Wei Hu (University of California Irvine), Weining Shen (University of California Irvine), Hua Zhou (University of California Los Angeles) Matrix Linear Discriminant Analysis / Analyse discriminante linéaire de matrices  
15:30-16:00	Cui Guo (University of Michigan), Timothy D. Johnson (University of Michigan) Scalar on Image Regression with Application to Multiple Sclerosis MRI Lesion Data. / Régression scalaire sur des images avec application à des données d'imagerie par résonance magnétique sur la sclérose en plaques  
16:00-16:30	Mina Gheiratmand (University of Alberta), Russell Greiner (University of Alberta), Matthew Brown (University of Alberta), Andrew Greenshaw (University of Alberta), Serdar Dursun (University of Alberta) Towards More Reliable Neuroimaging-Based Biomarkers in Mental Illness: The Case of Schizophrenia Discrimination using fMRI Data / Pour une plus grande fiabilité des marqueurs biologiques basés sur l'imagerie cérébrale en santé mentale : le cas de la discrimination en schizophrénie à l'aide de données IRM  

15:00-16:30 **Contributed / Communications libres** (abstract/résumé 241) **E2 165 (EITC)**

Measurement Error, Nonresponse, and Sample Design

Erreur de mesure, non-réponse et plan d'échantillonnage

Chair/Président: Zenaida Mateo



15:00-15:15	Edward Kroc (University of British Columbia), Bruno D. Zumbo (University of British Columbia) Calibration of Measurements: An Extension of Classic Measurement Error Theory / Étalonnage des mesures : une extension de la théorie classique de l'erreur de mesure  
15:15-15:30	Taraneh Abarin (Memorial University of Newfoundland) Non-identifiable Interaction Models with Misclassification / Modèles d'interaction non identifiables avec erreurs de classification  
15:30-15:45	Audrey-Anne Vallée (Université de Neuchâtel), Yves Tillé (Université de Neuchâtel) Revisiting Variance Decomposition when Independent Samples Intersect / Réexaminer la décomposition de la variance lorsque des échantillons indépendants s'entrecroisent  
15:45-16:00	Yilin Chen (Carleton University), Song Cai (Carleton University) Combining Multiple Samples through Density Ratio Models for More Efficient Estimation / Combinaison d'échantillons multiples via un modèle de rapport de densité pour une estimation plus efficace  
16:00-16:15	Ananthika Nirmalkanna (Memorial University of Newfoundland), Yildiz Yilmaz (Memorial University of Newfoundland) Multivariate Response-Dependent Two-Phase Sampling Designs / Plans d'échantillonnage à deux phases avec dépendance multivariée sur la réponse  
16:15-16:30	Jianhua Hu (Shanghai University of Finance and Economics, China), Jian Huang (University of Iowa), Feng Qiu (Zhejiang Agriculture and Forestry University, China) Response Best Subset Selection Model and Efficient Estimation in Multivariate Linear Regression / Modèle de sélection du meilleur sous-ensemble de réponse et estimation efficace dans une régression linéaire multivariée  









15:00-16:30 **Contributed / Communications libres** (abstract/résumé 244) **E2 160 (EITC)**

Teaching Statistics: New Perspectives

Enseignement de la statistique : nouvelles perspectives

Chair/Président: Alison Gibbs

15:00-15:15	Jeffrey D. Picka (University of New Brunswick) Motivating Introductory Statistics / Favoriser l'apprentissage des statistiques de base  
-------------	--

15:15-15:30	Jennifer Thornton (Mount Saint Vincent University) Flipping the Online Learning Environment / Inverser l'environnement d'apprentissage en ligne  
15:30-15:45	Anne Michele Millar (Mount Saint Vincent University), Andrea Perreault (Mount Saint Vincent University) Tall Tales vs Tall Tails: Rules of Thumb for Statistical Inference / Invraisemblance c. ailes relevées : règles pratiques d'inférence statistique  
15:45-16:00	Sohee Kang (University of Toronto Scarborough) Effectiveness of Gamification Feature in Introductory Statistics Course / Efficacité de la ludification dans les cours d'introduction à la statistique  
16:00-16:15	Sotirios Damouras (University of Toronto Scarborough), Sohee Kang (University of Toronto Scarborough) The Status of Statistics Curricula in Canada / L'état des programmes d'études en statistiques au Canada  

Abstracts • Résumés

SSC Presidential Invited Address
Allocution de l'invité du président de la SSC
Organizer/Responsable: Jack Gambino
Room/Salle: UC 210 (UC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 08:40-09:45]

Don A. Dillman (Washington State University)

The Challenge of Creating Data Collection Methods that are Neither Too Far Ahead nor Behind our Survey Respondents

Le défi de construire une méthode de collecte de données qui n'est ni en avance, ni en retard pour les répondants

The last ten years have witnessed enormous increases in computer memory, connectivity, storage and sensor capabilities, making it possible to obtain and send information from nearly anywhere to everywhere, and do it on pocket-size devices as easy to carry as one's wallet. The long-term implications of these developments for conducting surveys are likely to be enormous, but will happen at a more measured pace than the technology allows. In this presentation I will discuss the transition away from long established survey data collection methods towards newer ones, and the challenge data collection methodologists face in being neither too far ahead or too far behind where people are at this time of rapidly accelerating change. In particular, I will discuss the recent development of web-push data collection methods, i.e. contacting sample units by one mode, usually postal mail, to request a response over the Internet, while withholding other response options until later contacts.

Au cours des dix dernières années nous avons été témoins d'une augmentation fulgurante des capacités de mémoire et de sauvegarde, d'une connectivité accrue des appareils et d'une démocratisation des capteurs. Nous avons atteint un point où il est maintenant possible de transmettre de l'information de n'importe où vers n'importe où, et ce à l'aide de dispositifs format poche faciles à transporter. Les effets à long terme de ces changements sur la façon de mener une enquête risquent d'être substantiels, mais ils devraient arriver à un rythme moins effréné que l'évolution de la technologie ne le permet. Dans cet exposé, je décrirai la transition des méthodes traditionnelles de collecte de données d'enquête vers les plus récentes, ainsi que le défi pour les méthodologistes d'enquête afin que les données soient collectées avec des approches ni en avance, ni en retard sur la technologie utilisée par les gens, en ces temps où les changements surviennent de plus en plus rapidement. Je discuterai notamment de la nouvelle méthode de collecte de données promue par le Web (web-push) qui consiste à contacter une unité échantillonnale par un mode, habituellement la poste, en réclamant une réponse par internet tout en conservant les autres modes de réponse possibles pour les contacts ultérieurs.

Analysis of Complex Traits in Families and Populations

Analyse des caractères complexes dans les familles et populations

Chair/Président: Jinko Graham
Organizer/Responsable: Jinko Graham
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-10:50]

J. Concepcion LoredO-Osti (Memorial University of Newfoundland)

The Analysis of Longitudinal Multivariate (Discrete or Continuous) Traits under Irregular Time Measurements
Analyse longitudinale de traits multivariés (discrets ou continus) avec des temps de mesure irréguliers

There are well known methods to analyse longitudinal continuous variables with regularly spaced measurements and some of them (in particular, mixed models) have been adapted to deal with longitudinal quantitative traits. Nonetheless, the problem is far from being solved. The modelling and analysis of longitudinal multivariate discrete and or continuous traits poses various challenges to accommodate the family structure and irregular time measurements. In this presentation we discuss these challenges and propose some ways of addressing the problems.

Il existe des méthodes bien connues pour analyser des variables continues longitudinales dont les mesures sont prises régulièrement dans le temps. Certaines d'entre elles (notamment les modèles mixtes) ont été adaptées pour convenir aux traits quantitatifs. Le problème est toutefois loin d'être réglé. La modélisation et l'analyse de traits multivariés longitudinaux continus ou discrets pose différents défis lorsqu'il est question d'accommoder la structure de la famille et l'irrégularité des temps de mesure. Cet exposé présente ces défis et propose différentes façons d'aborder ces problèmes.

[Monday June 12/Lundi 12 juin, 10:50-11:20]

Kelly M. Burkett (University of Ottawa)

An Ancestral Tree-Based Approach to Detect Rare and Common Variants
Approche arborescente ancestrale pour détecter les variants rares et communs

For detecting genetic variants associated with a disease or trait, it is useful to consider the ancestral trees that gave rise to the sample's genetic variability. For both rare and common disease or trait influencing genetic variants, we expect to see haplotypes from individuals with similar values of the disease or trait clustered together in the ancestral tree corresponding to the genomic location of the variant. In this presentation, we describe how tree-based statistics can be used for detecting both rare and common genetic variants associated with either continuous or dichotomous outcomes. We summarize the performance of these statistics on simulated data having known and missing tree structures and we compare results to those obtained using conventional approaches to detect genetic association. Finally, application of the tree-based method to real data is also discussed.

Afin de détecter les variants génétiques associés à une maladie ou à un caractère, il est utile de considérer les arbres ancestraux qui ont donné lieu à la variabilité génétique des échantillons. Pour les variants génétiques rares et communs qui influencent les maladies ou certaines caractéristiques, nous nous attendons à voir des haplotypes qui se regroupent chez des individus qui ont des valeurs de maladie ou de caractéristique semblables, dans un arbre ancestral correspondant à la localisation génomique du variant. Dans cet exposé, nous décrivons comment les statistiques arborescentes peuvent être utilisées pour détecter les variants génétiques rares et communs avec des variables dépendantes soit continues ou dichotomiques. Nous montrons l'efficacité de ces statistiques avec des données simulées qui ont des structures arborescentes connues et manquantes. Nous comparons les résultats à ceux qui ont été obtenus avec des approches conventionnelles afin de détecter une association génétique. Enfin, nous présentons également l'application de la méthode arborescente sur des données réelles.

Analysis of Complex Traits in Families and Populations Analyse des caractères complexes dans les familles et populations

[Monday June 12/Lundi 12 juin, 11:20-11:50]

Fabrice Larribe (Université du Québec à Montréal)

Mapping Complex Traits, Rare Variants and Interaction via the Coalescent Process with Recombination

Cartographie de traits complexes, de variants rares et d'interaction par le processus de coalescence avec recombinaison

Genetic population data is the result of the evolution of chromosomes on this population; thus, it has been found natural to try to analyze such data by modeling the unknown history of this population. By combining an approach based on the coalescent process with recombination and importance sampling techniques, we can show how this stochastic process can be used to map genes influencing a disease. We address issues pertaining to complex diseases, whether this complexity is due to incomplete penetrance or phenocopy, to multiple rare variants, or to the interaction between genes. We introduce the progress made in building such histories and outline the remaining challenges.

Les données génétiques de population sont le résultat de l'évolution des chromosomes sur cette population ; il est alors apparu naturel de tenter d'analyser ces données en modélisant l'histoire inconnue de cette population. En combinant une approche fondée sur le processus de coalescence avec recombinaison et de l'échantillonnage préférentiel, on montre comment ce processus stochastique peut être utilisé pour cartographier des gènes influençant une maladie. Nous traitons des questions liées aux maladies complexes, que cette complexité soit due à de la pénétrance incomplète et de la phéno-copie, à de multiples variants rares, ou à une interaction entre gènes. Nous montrons les progrès réalisés pour construire de telles histoires et exposons les défis restant à résoudre.

Statistical Consulting: Bridging Innovation and Practice

Consultation statistique: entre innovation et pratique

Chair/Président: Sanjeena Dang
Organizer/Responsable: Sanjeena Dang
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-11:05]

Jeffrey A. Bakal (Alberta Health Services)

How Best to Bridge the Gap between Statistics and Medical Practice

Comment combler le fossé entre statistique et pratique médicale

One of the interesting questions in statistics is how to handle the duality of developing methodology that is sound and also easily interpretable in an applied space. Particularly in medical research there may be an understanding that a problem has been simplified for easy testing or power. These simplified solutions become the accepted standard of practice. As methodologies and capacity grow it is worth revisiting these simplifications to develop the methodology to better align with the goal. For the statistician there is an opportunity to either aim for a robust approach and present a new method, or new measurement, that addresses these simplifications or work to an incremental approach that steps the applied user closer to the goal over time. We will work through an example from the composite endpoints used in cardiovascular clinical trials. In these trials fatal and non-fatal events are combined into a single analysis for regulatory approval.

Une question intéressante se pose en statistique : comment gérer le double défi de mettre au point des méthodes à la fois appropriées et facilement interprétables dans un contexte appliqué. En recherche médicale tout particulièrement, il y a parfois lieu de penser qu'un problème a été simplifié pour en faciliter la vérification ou pour plus de puissance. Or ces solutions simplifiées deviennent ensuite la norme de pratique acceptée. Lorsque les méthodes et les capacités se développent ensuite, il est bon de revenir sur ces simplifications et d'adapter la méthodologie afin de mieux l'aligner sur l'objectif. Le statisticien peut alors soit viser une approche robuste et présenter une nouvelle méthode ou mesure qui remédie à ces simplifications, soit adopter une approche graduelle qui rapproche l'utilisateur appliqué de l'objectif au fil du temps. Nous étudierons un exemple d'issue composée utilisée dans les essais cliniques cardiovasculaires, où les événements fatals et non fatals sont combinés en une même analyse à des fins d'approbation réglementaire.

[Monday June 12/Lundi 12 juin, 11:05-11:50]

Paul D. McNicholas (McMaster University)

Some Experiences From an Industry Collaboration

Expériences tirées d'une collaboration en milieu industriel

A successful industry collaboration, which has spanned around eight years and received support from NSERC, OCE, and MITACS, is described. Details of many aspects of the collaboration are discussed, including funding opportunities that were exploited. Special attention is paid to some of research that took place as part of the collaboration and tangentially to it.

Description d'une collaboration menée avec succès en milieu industriel pendant près de huit ans avec le soutien du CRSNG, des OCE et de MITACS. Plusieurs aspects de cette collaboration sont abordés sous divers angles, y compris les possibilités de financement qui ont été mises à profit. Certains travaux de recherche menés de façon intégrante ou tangentielle dans le cadre de cette collaboration font l'objet d'une attention particulière.

Chair/Président: Karla Fox
Organizer/Responsable: Karla Fox
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-10:50]

Natalie Shlomo (University of Manchester)

Statistical Disclosure Control for Statistical Outputs: Where Do We Go From Here?

Contrôle de la divulgation de résultats statistiques : vers quoi allons-nous ?

I provide an overview of types of disclosure risks in traditional forms of data dissemination by statistical agencies, how the risk and information loss can be quantified and some recommended applications of disclosure control techniques. Statistical agencies, however, are making more use of restricting and licensing data as a statistical disclosure control approach given the difficulties in producing safe data. We look at new strategies for data dissemination and discuss their potential and limitations. These include web-based applications where the outputs are protected without the need for human intervention to check for disclosure risks, on-site data enclaves and its extension where researchers can access confidential data on their personal PCs through a remote server and the production and release of 'synthetic' data based on advanced statistical models. The link to the Computer Science definition of differential privacy will be discussed.

Un aperçu des types de risques liés à la divulgation relativement aux formes traditionnelles de diffusion des données par les agences de la statistique ainsi que du mode de quantification du risque et de la perte de données. Certaines recommandations en matière d'application de techniques de contrôle de la divulgation sont présentées. Cependant, les agences de la statistique recourent davantage à la restriction et à l'exploitation sous licence des données comme mode de contrôle de la divulgation de statistiques, compte tenu de la difficulté à produire des données sécurisées. Nous aborderons, notamment sous l'angle de leur potentiel et de leurs limites, les nouvelles stratégies de diffusion des données, y compris les applications sur le Web produisant des résultats protégés sans qu'aucune intervention humaine soit nécessaire pour vérifier les risques de divulgation, les enclaves de données et leur bloc d'étendue où les chercheurs peuvent accéder à des données confidentielles sur leur ordinateur personnel par l'entremise d'un serveur à distance et enfin la production et la publication de données « synthétiques » fondées sur des modèles statistiques avancés. Le lien vers la définition en science informatique du caractère privé différentiel sera aussi discuté.

[Monday June 12/Lundi 12 juin, 10:50-11:20]

Anne-Sophie Charest (Université Laval)

What is Synthetic about Synthetic Datasets?

Qu'y a-t-il de synthétique dans un jeu de données synthétiques ?

When synthetic datasets were proposed as a way to protect the confidentiality of respondents, the reasoning was that since they are not about anyone in the real dataset, they do not pose any risk. Our understanding of the risk of these datasets has however evolved, and we now know that they are not risk free. Nevertheless, synthetic datasets can be a useful tool for privacy protection and there is now a growing literature on the best techniques for creating and analysing synthetic datasets. But this has lead to the term synthetic dataset being used

À l'origine, on croyait que créer des jeux de données synthétiques pour protéger la confidentialité des répondants éliminerait complètement le risque de divulgation puisque ceux-ci ne contiendraient pas d'information à propos de vrais individus. Notre compréhension du risque de ces jeux de données a cependant évoluée, et nous savons maintenant qu'ils ne sont pas sans risque. Néanmoins, les jeux de données synthétiques sont utiles pour la protection de la confidentialité et on trouve de plus en plus d'articles sur les meilleures techniques pour les créer et les analyser. Mais le terme jeu de données synthétiques est utilisé dans des contextes variées, et il est

in a lot of different contexts, and it is not clear anymore what the essence of a synthetic dataset is. What really differentiates a synthetic dataset from a perturbed real dataset? This talk explores the main attributes of synthetic datasets and offers some thoughts for thinking and talking about them.

[Monday June 12/Lundi 12 juin, 11:20-11:50]

Mark Stinner (Statistics Canada)

Disclosure Control and Random Tabular Adjustment

Contrôle de la divulgation et ajustement tabulaire aléatoire

Statistical agencies are interested in publishing useful statistical data but doing so may lead to the disclosure of individuals' private data. This is a problem, as it leads to a trade-off between the utility of the published data and the risk of disclosure of confidential data. Disclosure control can be seen as the use of methods to deal with this problem by assessing and controlling the risk of disclosing confidential data while also providing researchers with useful statistical data. This paper describes a disclosure control model based largely on Bayesian decision theory. This model allows for the description of the concepts of disclosure control in terms of familiar statistical concepts such as expectation and variance. A method of disclosure control, called random tabular adjustment (RTA), is described. This method controls the risk of disclosure by randomly adjusting the data instead of suppressing cells. It fits naturally into the disclosure control model described.

maintenant difficile de savoir ce qui définit vraiment un jeu de données synthétique, et le différencie par exemple d'un jeu de données réelles modifié pour être confidentiel. Nous explorons ici les principales caractéristiques des jeux de données synthétiques et offrons des pistes de réflexion pour étudier ces objets.

Les agences de la statistique se préoccupent de publier des données statistiques utiles, ce qui pourrait cependant entraîner la divulgation de données personnelles à caractère privé. C'est un problème, car il en découle un choix entre l'utilité des données publiées et le risque de divulguer de l'information confidentielle. On peut voir le contrôle de la divulgation comme une possible solution à ce problème par l'utilisation de méthodes d'évaluation et de contrôle du risque de divulgation de renseignements confidentiels, tout en mettant à la disposition des chercheurs des données statistiques utiles. Il s'agit ici de décrire un modèle de contrôle de la divulgation largement fondé sur la théorie bayésienne de la décision. Ce modèle permet la description de concepts de contrôle de la divulgation selon des concepts statistiques familiers, comme l'espérance et la variance. Nous décrivons aussi une méthode de contrôle de la divulgation appelée ajustement tabulaire aléatoire. C'est une méthode de contrôle du risque de divulgation par l'ajustement aléatoire des données plutôt que par la suppression de cellules. Cette méthode s'intègre naturellement au modèle de contrôle de la divulgation dont il est question.

Chair/Président: Xin (Cindy) Feng
Organizer/Responsable: Xin (Cindy) Feng
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-10:42]

Farouk Nathoo (University of Victoria), **Yin Song** (University of Victoria)

A Potts Mixture Spatiotemporal Joint Model for Combined MEG and EEG Data

Modèle de mélange conjoint spatio-temporel de Potts pour données MEG et EEG combinées

We consider the ill-posed inverse problem arising when magnetoencephalography (MEG) and/or electroencephalography (EEG) are used to measure electromagnetic brain activity over an array of sensors at the scalp and it is of interest to map these data back to the sources of neural activity within the brain. We review existing approaches to solving this inverse problem and discuss the mesostate-space model (MSM) proposed by Daunizeau and Friston (Neuroimage, 2007). We then propose a new model that builds on the MSM and incorporates three major extensions: (i) We combine EEG and MEG data together and formulate a joint model for source reconstruction; (ii) we incorporate the Potts model to represent the spatial dependence in an allocation process that partitions the cortical surface into a small number of latent mesostates; (iii) we formulate the mesostate dynamics in a more flexible manner so that the model can characterize the functional connectivity between mesosources. We formulate the model, discuss computational implementation, and make comparisons to existing methods.

Nous considérons le problème inverse mal posé qui survient lorsque l'activité électromagnétique du cerveau est mesurée par magnétoencéphalographie (MEG) et/ou électroencéphalographie (EEG) via des capteurs fixés sur le cuir chevelu, et que l'on souhaite retracer le lien entre ces données et les sources d'activité des neurones dans le cerveau. Nous passons en revue les solutions existantes à ce problème et discutons du modèle méso-état-espace (mesostate-space model, ou MSM) proposé par Daunizeau et Friston (Neuroimage, 2007). Nous proposons ensuite un nouveau modèle qui s'appuie sur le modèle MSM tout en y ajoutant trois extensions : (i) nous combinons les données EEG et MEG et formulons un modèle conjoint permettant la reconstruction des sources ; (ii) nous intégrons le modèle de Potts afin de représenter la dépendance spatiale dans un processus de distribution qui divise la surface du cortex en un petit nombre de méso-états latents ; (iii) nous formulons la dynamique des méso-états de manière plus souple, afin que le modèle puisse caractériser la connectivité fonctionnelle entre méso-sources. Nous formulons enfin ce modèle et en discutons l'exécution computationnelle avant de le comparer aux méthodes existantes.

[Monday June 12/Lundi 12 juin, 10:42-11:04]

Andrew Lawson (Medical University of South Carolina), **Ray Boaz** (Medical University of South Carolina)

Spatially Structured Sparseness in Bayesian Spatial Health Modeling

Dispersion spatiale des données en modélisation bayésienne spatiale en santé

Often geospatial disease outcomes are characterized by sparseness when viewed as a count distribution. This is sometimes called zero-inflation. Classically the models for zero inflation are of two types: zero class modelled as 'structural' or 'Poisson' or as hurdle models where the zeroes are treated completely separately from the truncated Poisson positive counts. Often the idea of structural zeroes is not made specific in that no attempt is made to assess which zeroes are in that class. In this talk

Les effets géospatiaux des taux de maladies sont souvent caractérisés par une dispersion des données, lorsque celles-ci sont considérées comme des données de comptage. Ce phénomène est parfois appelé inflation des zéros. Généralement, les modèles utilisés à ces fins sont de deux types : i) les zéros sont modélisés de façon structurelle ou en supposant une distribution de Poisson ; ou ii) les zéros sont modélisés à l'aide d'un modèle « hurdle », dans lequel ils sont traités entièrement à part des valeurs mesurées positives tronquées de Poisson. Souvent, l'idée de zéros structurels n'est pas spéci-

Spatial and Spatiotemporal Statistics: Novel Methods and Applications

Statistique spatiale et spatiotemporelle : nouvelles méthodes et applications

I describe an approach which models the spatial structure of the structural zero class and provides estimates of the spatial distribution of the probability of being structural or not. This is applied in a spatio-temporal context. Different assumptions about the spatial prior distribution of the structural probability are compared using a Bayesian formulation. An example of spatio-temporal modeling of sudden infant death in the counties of Georgia USA will be presented.

ifique : aucune tentative n'est faite pour établir quels zéros sont dans cette classe. Dans cette présentation, je décris une approche qui modélise la structure spatiale de la classe de zéros structurels et qui offre des estimations de la distribution spatiale de la probabilité de leur état structurel ou non structurel. J'applique l'approche dans un contexte spatiotemporel. Je compare différentes hypothèses concernant la distribution spatiale de la probabilité structurelle à l'aide d'une formulation bayésienne. Je présente l'exemple d'un modèle spatiotemporel de la mort subite du nourrisson dans plusieurs comtés de l'État de Géorgie (États-Unis).

[Monday June 12/Lundi 12 juin, 11:04-11:26]

Patrick E. Brown (St. Michael's Hospital), **Ofir Hariri** (University of Toronto), **Alisha Albert-Green** (University of Toronto), **Jamie Stafford** (University of Toronto), **Paul Nguyen** (Queen's University), **Lennon Li** (Public Health Ontario)

Continuous Models for Aggregated Spatio-Temporal Data in Practice

Modèles continus pratiques pour données spatiotemporelles agrégées

Dr. Brown and his collaborators have presented the local-EM algorithm for smoothing aggregated data a number of times at past SSC conferences, although the methodology has been too computationally intensive (and labour intensive) to attract a following. This year's talk will change the situation with the presentation of an R package and a description of the methodological and computational improvements which have made the method useful in practice. Efficient use of R's spatial data structures and parallel processing has decreased to the computational burden of the method to a manageable level. Re-casting the algorithm as modelling a latent-Gaussian process rather than a kernel smoother has allowed for a sparse matrix Markov random field approximation for the covariance matrix. An R package includes several examples using real datasets.

Brown et ses collaborateurs ont à plusieurs reprises présenté un algorithme EM local pour lissage de données agrégées aux congrès de la SSC, mais la méthode demandait une trop grande capacité de calcul (et trop de travail) pour être adoptée par un large groupe d'utilisateurs. La présentation de cette année change cette perspective avec l'introduction d'un package R et une description des améliorations méthodologiques et computationnelles qui rendent cette méthode utile dans la pratique. Une utilisation efficace des structures de données spatiales et de la capacité de traitement parallèle qu'offre R a permis de réduire la charge computationnelle de la méthode à un niveau raisonnable. Par ailleurs, la reformulation de l'algorithme comme modèle à processus gaussiens latents plutôt qu'un lisseur de noyaux a permis une approximation des champs aléatoires de Markov à matrice creuse de la matrice de covariance. Nous incluons un package R avec plusieurs exemples utilisant des ensembles de données réels.

[Monday June 12/Lundi 12 juin, 11:26-11:50]

Longhai Li (University of Saskatchewan), **Xin (Cindy) Feng** (University of Saskatchewan), **Shi Qiu** (International Road Dynamics)

Estimating Cross-validatory Predictive P-values with Integrated Importance Sampling for Disease Mapping Models

Estimation de valeurs p prédictives par validation croisée avec échantillonnage préférentiel intégré pour les modèles de cartographie des maladies

An important statistical task in disease mapping problems is to identify divergent regions with unusually high or low risk of disease. Leave-one-out cross-validatory (LOOCV) model assessment is the gold standard for estimating predictive p-

L'une des principales tâches statistiques dans les problèmes de cartographie des maladies est l'identification des régions divergentes où le risque de la maladie est exceptionnellement élevé ou exceptionnellement faible. L'évaluation du modèle selon la méthode de validation croisée « leave-one-out »

values that can flag such divergent regions. However, LOOCV is time-consuming. We introduce a new method, called integrated importance sampling (iIS), for estimating LOOCV predictive p-values with only Markov chain samples drawn from the posterior based on a full data set. The formula used by iIS can be proved to be equivalent to the LOOCV predictive p-value. We compare iIS and three other existing methods in the literature with two disease mapping datasets. Our empirical results show that the predictive p-values estimated with iIS are almost identical to the predictive p-values estimated with actual LOOCV, and outperform those given by the existing three methods, namely, the posterior predictive checking, the ordinary importance sampling, and the ghosting method by Marshall and Spiegelhalter (2003).

(LOOCV) est la référence absolue pour estimer des valeurs p prédictives qui peuvent repérer de telles régions divergentes. Par contre, LOOCV demande beaucoup de temps. Nous présentons une nouvelle méthode, l'échantillonnage préférentiel intégré (iIS), pour l'estimation de valeurs p prédictives LOOCV utilisant uniquement des échantillons de chaînes de Markov provenant d'une distribution a posteriori basée sur un ensemble de données complet. Il est possible de démontrer que la formule utilisée par iIS est équivalente à la valeur p prédictive LOOCV. Nous comparons iIS et trois autres méthodes existantes de la littérature avec deux ensembles de données de cartographie de maladies. Nos résultats empiriques démontrent que les valeurs p prédictives estimées à l'aide de iIS sont presque identiques aux valeurs p prédictives estimées avec LOOCV, et qu'elles surpassent celles provenant des trois méthodes existantes, à savoir la vérification prédictive a posteriori, l'échantillonnage préférentiel ordinaire, et la méthode ghosting par Marshall et Spiegelhalter (2003).

Statistical Inference for Complex Dynamical Systems

Inférence statistique pour systèmes dynamiques complexes

Chair/Président: Jiguo Cao
Organizer/Responsable: Jiguo Cao
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-10:50]

James O. Ramsay (McGill University), **Michelle Carey** (University College Dublin), **Juan Li** (McGill University)

From Brain to Hand to Statistics with Dynamic Smoothing

Du cerveau à la main aux statistiques avec lissage dynamique

Systems of differential equations are often used to model buffering processes that modulate a non-smooth high-energy input so as to produce an output that is smooth and that distributes the energy load over time and space. Handwriting is buffered in this way. We show that the smooth complex script that spells “statistics” in Chinese can be represented as buffered version of a series of 46 equal-interval step inputs. The buffer consists of three undamped oscillating springs, one for each orthogonal coordinate. The periods of oscillation vary slightly over the three coordinate in a way that reflects the masses that are moved by muscle activations. We use the term “dynamic smoothing” for the estimation of a structured input functional object along with the buffer characteristics.

Des systèmes d'équations différentielles sont souvent utilisés pour la modélisation de processus de mise en mémoire tampon modulant une grande entrée d'énergie inégale pour produire une sortie lisse qui distribue la charge énergétique dans le temps et l'espace. L'écriture est tamponnée de cette façon. Nous démontrons que les lettres lisses et complexes qui épellent « statistiques » en chinois peuvent être représentées par des versions tamponnées d'une série de 46 entrées en échelon à intervalles égaux. Le tampon consiste en trois ressorts au mouvement oscillatoire sans amortissement, un pour chaque coordonnée orthogonale. Les périodes d'oscillation varient légèrement en fonction des trois coordonnées d'une façon qui reflète les masses qui sont bougées par l'activation musculaire. Nous utilisons le terme « lissage dynamique » pour l'estimation d'un objet fonctionnel d'entrée structuré ainsi que pour les caractéristiques du tampon.

[Monday June 12/Lundi 12 juin, 10:50-11:20]

Mike Dowd (Dalhousie University)

Approximations to Facilitate Data Assimilation for High Dimensional Dynamical Systems

Approximations pour faciliter l'assimilation de données dans les systèmes dynamiques à haute dimension

A major challenge for inference in spatio-temporal dynamical systems is the high dimension of the simulation models used; model integration, or forward simulation, is extremely computationally costly. Ensemble-based data assimilation has emerged as a method of choice for state estimation for environmental prediction based on numerical models of the ocean, climate, and atmosphere. There is also interest in parameter estimation and uncertainty quantification (UQ) for these problems. In this talk, I will discuss approximate approaches to facilitate state and parameter estimation, as well as UQ, for these computationally costly systems. This overview will include ensemble Kalman filters, and various approximations for particle filters. The use of emulators is also discussed. The approaches

Un grand défi pour l'inférence dans les systèmes dynamiques spatio-temporels est la dimension élevée des modèles de simulation utilisés ; l'intégration du modèle, ou simulation vers l'avant, est très coûteuse en ressources informatiques. L'assimilation de données basée sur un ensemble est devenue une méthode de choix dans l'estimation de l'état pour les prédictions environnementales basées sur des modèles numériques du climat, de l'océan et de l'atmosphère. Ces problèmes génèrent aussi un intérêt particulier pour l'estimation de paramètres et la quantification de l'incertitude (UQ). Au cours de cette présentation, je discuterai des méthodes approximatives qui facilitent l'estimation d'état et de paramètres ainsi que l'UQ pour ces systèmes coûteux en ressources informatiques. Cet aperçu inclura des filtres de Kalman pour ensembles et différentes approximations pour des filtres particuliers. L'utilisation d'émulateurs est aussi discutée. Les

are illustrated using data assimilation problems in oceanography.

approches sont illustrées à l'aide de problèmes d'assimilation de données en océanographie.

Statistical Inference for Complex Dynamical Systems

Inférence statistique pour systèmes dynamiques complexes

[Monday June 12/Lundi 12 juin, 11:20-11:50]

Oksana Chkrebtii (Ohio State University), **Yury E. Garcia** (Centro de Investigacion en Matematicas), **Marcos A. Capistran** (Centro de Investigacion en Matematicas)

Challenges and Strategies for Inference on a Stochastic Multi-Pathogen Model of Disease Dynamics in San Luis Potosi, Mexico

Défis et stratégies d'inférence sur un modèle de dynamiques de maladie stochastique à pathogènes multiples à San Luis Potosi, Mexico

Acute respiratory diseases (ARI) are a public health concern associated with morbidity and mortality, especially in children and the elderly. The nature of individual interactions mean that disease dynamics in a large population may be described by a stochastic process model governed by a system of stochastic differential equations (SDE). When multiple ARI are in circulation, the dynamics are further complicated, and the corresponding SDE cannot be solved analytically. In this analysis we examine the challenges of inference for models of multiple pathogen dynamics, including the inability to distinguish different pathogens based on symptoms and the partial symmetries in the model. We discuss how to overcome resulting identifiability problems by utilizing additional data sources and modeling data from multiple outbreak seasons simultaneously.

Les maladies respiratoires graves (ARI) sont un problème de santé publique associé à la morbidité et à la mortalité, particulièrement chez les enfants et les personnes âgées. La nature des interactions individuelles signifie que les dynamiques de la maladie dans une grande population peuvent être représentées par un modèle de processus stochastiques régi par un système d'équations différentielles stochastiques (SDE). Lorsque plusieurs ARI sont en circulation, les dynamiques sont plus compliquées et les SDE correspondants ne peuvent pas être résolus analytiquement. Cette analyse examine les défis de l'inférence pour des modèles de dynamiques à pathogènes multiples dont l'incapacité de distinguer différents pathogènes à partir des symptômes et les symétries partielles dans le modèle. Nous abordons la façon de surmonter les problèmes d'identifiabilité qui en découlent en utilisant des sources de données additionnelles et en modélisant simultanément les données provenant de différentes saisons d'épidémies.

Chair/Président: Robert W. Platt
Room/Salle: E2 165 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-10:35]

Di Shu (University of Waterloo), **Grace Yi** (University of Waterloo)

Causal Inference with Measurement Error in Outcomes: Bias Analysis and Estimation Methods

Inférence causale comportant des erreurs de mesure dans les réponses : analyse du biais et méthodes d'estimation

The average treatment effect (ATE) is often of primary interest in causal inference, and it is commonly estimated based on the inverse probability weighting (IPW) method. Its validity, however, is challenged by the presence of error-prone outcomes. In this paper we investigate various issues concerning IPW estimation of ATE with mismeasured outcome variables. We study the impact of measurement error and reveal important consequences of the naive analysis which ignores measurement error. When a continuous outcome variable is mismeasured under an additive measurement error model, the naive analysis still yields a consistent estimator; when the outcome is binary, we derive the asymptotic bias in a closed-form. To conduct valid inference, we develop estimation procedures for practical scenarios where either validation data or replicates are available. Furthermore, with validation data we propose an efficient method which substantially outperforms usual methods of using validation data.

L'effet moyen du traitement suscite souvent un grand intérêt dans l'inférence causale, et il est couramment estimé d'après la méthode de pondération de la probabilité inverse. Cependant, sa validité est remise en question par la présence de réponses sujettes à l'erreur. Dans cet exposé, nous examinons les différents problèmes relatifs à l'estimation de l'effet moyen du traitement par la pondération de la probabilité inverse avec des variables de réponse mesurées incorrectement. Nous étudions les répercussions de l'erreur de mesure et dévoilons les lourdes conséquences de l'analyse naïve qui ne tient pas compte de l'erreur de mesure. Lorsqu'une variable réponse continue n'est pas mesurée correctement dans un cadre d'un modèle additif d'erreur de mesure, l'analyse naïve produit toujours un estimateur convergent ; lorsque le résultat est binaire, nous obtenons un biais asymptotique sous une forme analytique fermée. Afin d'effectuer une inférence valide, nous mettons au point des procédures d'estimation de scénarios pratiques dans lesquels soit les données de validation ou les répétitions sont disponibles. De plus, nous proposons une méthode efficace qui est nettement meilleure que les méthodes habituelles utilisant des données de validation.

[Monday June 12/Lundi 12 juin, 10:35-10:50]

Shixiao Zhang (University of Waterloo), **Peisong Han** (University of Waterloo), **Changbao Wu** (University of Waterloo)

A Unified Empirical Likelihood Approach to Testing MCAR and Subsequent Estimation

Approche unifiée de vraisemblance empirique pour des tests avec des données manquantes de façon complètement aléatoire et leur estimation subséquente

For estimation with missing data, a crucial step is to determine if the data are missing completely at random (MCAR), in which case a complete-case analysis would suffice. Most existing tests for MCAR do not provide a method for subsequent estimation once the MCAR is rejected. In the setting of estimating the means of some response variables that are subject to missingness, we propose a unified approach to testing MCAR and the subsequent estimation. Upon rejecting MCAR, the same set of weights used for testing can then be used for estima-

Pour une estimation avec des données manquantes, il est essentiel de déterminer si les données sont manquantes de façon complètement aléatoire (MFCA) et dans ce cas, une analyse de cas complète devrait suffire. La plupart des tests existants avec données MFCA ne fournissent aucune méthode d'estimation subséquente en cas de rejet de l'hypothèse MCFA. Pour l'estimation de la moyenne de certaines variables de réponse avec données manquantes, nous proposons une approche unifiée pour tester les données MCFA et l'estimation subséquente. S'il y a rejet de données MCFA, le même ensemble de poids utilisé pour le test peut aussi servir pour

tion. The resulting estimators are consistent if the missingness of each response variable depends on a set of fully observed auxiliary variables only and the true outcome regression model is among the user-specified functions for deriving the weights. The proposed procedure is based on the calibration idea from survey sampling literature and the empirical likelihood theory. This is joint work with Peisong Han and Changbao Wu.

l'estimation. Les estimateurs qui en résultent sont convergents si l'état manquant de chaque variable de réponse dépend seulement d'un ensemble de variables auxiliaires complètement observées et que le véritable modèle de régression de chaque variable réponse est parmi les fonctions spécifiées par l'utilisateur pour dériver les poids. La procédure proposée se fonde sur le principe d'étalonnage tiré de documents d'échantillonnage de sondages et de la théorie empirique de la vraisemblance. En collaboration avec Peisong Han et Changbao Wu.

[Monday June 12/Lundi 12 juin, 10:50-11:05]

Gabrielle Simoneau (McGill University), **Erica E.M. Moodie** (McGill University), **Robert W. Platt** (McGill University), **Bibhas Chakraborty** (Duke-NUS Medical School, National University of Singapore)

Non-regular Inference for Dynamic Weighted Ordinary Least Squares: Understanding the Impact of Solid Food Intake in Infancy on Childhood Weight

Inférence non régulière pour l'approche dynamique pondérée des moindres carrés ordinaires : comprendre l'impact d'une diète d'aliments solides d'un nourrisson sur le poids durant l'enfance

A dynamic treatment regime (DTR) is a set of decision rules to be applied across multiple stages of treatments. The decisions are tailored to individuals, inputting an individual's observed characteristics and outputting a treatment decision at each stage for that individual. Dynamic weighted ordinary least squares (dWOLS) is a theoretically robust and easily implemented method for estimating an optimal DTR. Like many DTR estimators, the dWOLS estimators can be non-regular when true treatment effects are zero or very small, resulting in invalid Wald-type or standard bootstrap confidence intervals. Inspired by an analysis of the effect of diet in infancy on weight and body size in childhood, we investigate the use of the m-out-of-n bootstrap with dWOLS for valid inferences of optimal DTR. We provide an extensive simulation study to compare the performance of different choices of resample size m in situations where the treatment effects are non-regular. We illustrate the methodology to study the effect of solid food intake in infancy on long-term health outcomes.

Un plan dynamique de traitement (PDT) est un ensemble de fonctions de décisions appliquées à plusieurs étapes de traitement. Les fonctions décisionnelles sont adaptées aux caractéristiques individuelles des patients, utilisant des variables dépendantes observées sur le patient et retournant une décision de traitement à chaque étape pour ce patient. L'approche dynamique pondérée des moindres carrés ordinaires (dWOLS) est une méthode théoriquement robuste et facile à implémenter pour estimer un PDT optimal. Comme la plupart des estimateurs de PDT, les estimateurs de dWOLS peuvent être non réguliers lorsque les effets du traitement sont nuls ou négligeables, résultant en des intervalles de confiance de type Wald ou bootstrap invalides. Inspiré par une analyse de l'effet de la diète d'un nourrisson sur son poids durant son enfance, nous étudions l'utilisation du «m-out-of-n bootstrap» avec dWOLS pour une inférence valide d'un PDT optimal. Nous conduisons une étude de simulation pour comparer la performance du «m-out-of-n bootstrap» pour différent choix de m lorsque les estimateurs dWOLS sont non réguliers. Nous illustrons la méthodologie pour étudier l'effet d'une diète d'aliments solides chez le nourrisson sur des indices corporels mesurés à long terme.

[Monday June 12/Lundi 12 juin, 11:05-11:20]

Yuying Xie (University of Waterloo), **Yeying Zhu** (University of Waterloo), **Cecilia Cotton** (University of Waterloo)

Kernel-Based Causal Inference

Inférence causale basée sur le noyau

An important goal in estimating the causal effect is to achieve balance in the covariates. We propose

Un objectif important dans l'estimation de l'effet causal est d'atteindre un équilibre dans les covariables. Nous propo-

using kernel distance to measure balance among different treatment groups and propose a new propensity score estimator by setting the kernel distance to be zero. The estimating equations are solved by generalized method of moments. Simulation studies are conducted across different scenarios varying in the degree of nonlinearity. The simulation study shows that the proposed approaches produce smaller mean squared errors in estimating causal treatment effects than many existing approaches including the well-known covariate balance propensity score (CBPS) approach. Further, it is shown that the kernel distance is one of the best bias indicators in estimating the causal effect compared to other balance measures, such as absolute standardized mean difference (ASMD) and KS statistic.

sons d'utiliser la distance du noyau pour mesurer l'équilibre entre les différents groupes de traitement et proposer un nouvel estimateur du score de propension en fixant la distance du noyau à zéro. Les équations d'estimation sont résolues par la méthode des moments généralisés. Des études de simulation sont menées sous différents scénarios où le degré de non-linéarité varie. L'étude de simulation montre que les approches proposées produisent des erreurs quadratiques moyennes plus petites dans l'estimation des effets causaux du traitement que de nombreuses approches existantes, y compris l'approche bien connue du score de propension équilibrant les covariables (SPEC). De plus, il est démontré que la distance du noyau est l'un des meilleurs indicateurs de biais pour estimer l'effet causal par rapport aux autres mesures d'équilibre, telles que la différence moyenne standardisée en valeur absolue (DMSA) et la statistique KS.

[Monday June 12/Lundi 12 juin, 11:20-11:35]

Sean McGrath (McGill University), **Russell Steele** (McGill University), **Andrea Benedetti** (McGill University)

Incorporating Medians in Meta-Analysis

Inclusion des médianes dans la méta-analyse

In meta-analysis of a continuous outcome, usually studies that report medians are excluded or the median and spread are used to estimate a mean and standard deviation in order to estimate a pooled effect across studies. In our previous work, we proposed several methods to directly meta-analyze medians. We showed via simulation that these methods outperform methods that estimate means and standard deviations from studies reporting medians, and we provided guidelines for data analysts to improve the performance of meta-analyzing studies that report medians. We apply these novel methods to recently published meta-analyses and evaluate the extent to which the pooled estimates and conclusions drawn from these meta-analyses change when using these methods.

Lors de la méta-analyse d'une issue continue, les études qui présentent des médianes sont généralement exclues, ou bien on utilise la médiane et l'écart pour estimer une moyenne et un écart type et estimer ainsi un effet groupé entre études. Dans nos travaux précédents, nous avons proposé plusieurs méthodes pour la méta-analyse directe des médianes. Nous avons montré par simulation que ces méthodes surpassent celles qui estiment les moyennes et les écarts types des études incluant des médianes; nous avons proposé aux analystes de données des directives pour les aider à améliorer la performance de ce type d'études de méta-analyse. Nous appliquons ces nouvelles méthodes à des méta-analyses récemment publiées et évaluons la mesure dans laquelle les estimations groupées et les conclusions de ces études changent si on applique nos méthodes.

[Monday June 12/Lundi 12 juin, 11:35-11:50]

Marc Simard (Institut national de santé publique du Québec), **Nathalie Vandal** (Institut National de Santé Publique du Québec), **Meranda Nakhla** (Research Institute of the McGill University Health Centre and McGill University), **Elham Rahme** (Research Institute of the McGill University Health Centre), **Astrid Guttmann** (Hospital for Sick Children, University of Toronto, Institute of Health Policy, Management and Evaluation, and Institute for Clinical Evaluative Sciences)

Use of the Segmented Regression Analysis Approach to Adjust for Potential Confounders and Effect Modifiers when Evaluating Public Health Interventions

Utilisation de la régression segmentée pour tenir compte de la confusion et de l'effet modifiant dans l'évaluation d'une

intervention en santé publique

The segmented regression analysis approach is a useful method that may be used to evaluate effects of public health interventions. This approach may be superior to standard time-series analyses because it allows for short and long-term effects evaluation, while adjusting for potential confounders and assessing study subject characteristics (like socioeconomic status [SES]) that may influence these effects. We used the segmented regression approach to study if implementation of a pediatric diabetes network in Ontario decreased SES-disparities in emergency (ED) visits. We used population-based aggregate data of annual crude rates for each level of SES and potential confounders: geographic location, sex and age group. We determined population average adjusted ED-visit rate ratios by accounting for the number of individuals in each aggregate using generalized estimating equation with a Poisson link. We observed that, after network implementation, ED-visits decreased significantly in the lowest SES quintiles. We concluded that the segmented regression analysis approach may be used to analyse aggregate ecological data while adjusting for potential confounders and effect modifiers.

La régression segmentée est une technique à envisager pour évaluer l'efficacité d'intervention en santé publique. Cette technique, par rapport aux séries chronologiques classiques, permet d'évaluer l'effet à court et long terme d'interventions selon les caractéristiques des sujets à l'étude (ex : statut socioéconomique [SSE]) tout en ajustant pour des facteurs confondants. À l'aide de la régression segmentée, nous avons évalué si l'implantation d'un réseau de diabète pédiatrique en Ontario réduit les disparités selon le SSE en ce qui a trait aux admissions à l'urgence (AU). Nous avons utilisé des données populationnelles agrégées pour chacun des niveaux de SSE et des variables potentiellement confondantes : zone géographique, sexe et groupe d'âge. Nous avons estimé les taux populationnels moyens ajustés d'AU en considérant le nombre d'individus dans chaque agrégat avec un modèle d'équations d'estimation généralisées avec un lien de Poisson. Nous avons observé, après l'implantation du réseau, une diminution plus importante dans les niveaux de SSE les plus faibles. Nous concluons que la régression segmentée est une technique qui peut être utilisée pour analyser les données écologiques tout en ajustant pour des variables potentiellement confondantes ou ayant un effet modifiant.

Probability and Estimation Probabilité et estimation

Chair/Président: Clarence Simard
Room/Salle: E2 160 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-10:35]

Clemonell Bilayi-Biakana (University of Ottawa), **Rafal Kulik** (University of Ottawa), **Gail Ivanoff** (University of Ottawa)

Tail Empirical Processes for Stochastic Volatility Models

Processus empiriques de queue des modèles de volatilité stochastique

We consider the tail empirical process (TEP) related to a distribution with a regularly varying tail. This is an important tool used in nonparametric estimation of extremal quantities, like the Hill estimator of the index of regular variation, or various risk measures. In this talk, we consider a long memory stochastic volatility model of interest in finance. We first start by investigating some probabilistic properties of this model. We establish central and non-central limit theorems for the TEP and apply these results to investigate the asymptotic behaviour of the aforementioned extremal quantities. Our theoretical results are illustrated by simulation studies.

Nous considérons le processus empirique se rapportant à la queue d'une distribution variant régulièrement. Ceci est outil essentiel utilisé dans l'estimation non paramétrique des quantités extrêmes comme l'estimateur Hill de l'indice de variation régulière, ou diverses mesures de risque. Dans cet exposé, nous considérons un modèle de volatilité stochastique à mémoire longue trouvant ses applications en finance. Nous commençons par étudier certaines propriétés probabilistes de ce modèle. Nous établissons des théorèmes de limite centrale et non centrale pour ce processus empirique et nous appliquons ces résultats à l'étude du comportement asymptotique des quantités extrêmes mentionnées en amont. Nos résultats théoriques sont illustrés par des simulations.

[Monday June 12/Lundi 12 juin, 10:35-10:50]

Hai Yan Liu (University of Ottawa), **Ioana Schiopu-Kratina** (University of Ottawa), **Pierre-Jerome Bergeron** (University of Ottawa), **Mayer Alvo** (University of Ottawa)

Modeling Recurrent Gap times Through Conditional GEE

Modélisation des intervalles récurrents au moyen d'équations d'estimation généralisées

In this paper, we present the analysis of recurrent event data subject to censoring using generalized estimating equations for the conditional means and variances of the gap times. Censoring is dealt with by imposing conditional independence assumptions on the censored gap times, the covariates and the censoring times. Simulation results demonstrate the relative robustness of parametric estimates even when this assumption is incorrect. We actually estimate the censored gap time using only the observed data, by solving estimating equations that are asymptotically unbiased. We prove the strong consistency using modern analytical techniques and illustrate our results through simulations.

Nous présentons l'analyse de données d'événements récurrents sujets à la censure au moyen d'équations d'estimation généralisées permettant de déterminer les moyennes et les variances conditionnelles des intervalles de temps. La censure est réglée en imposant des hypothèses d'indépendance conditionnelle sur les intervalles de temps censurés, les covariables et les temps de censure. Les résultats de simulations démontrent la robustesse relative des estimations paramétriques même si cette hypothèse est erronée. En fait, nous estimons les intervalles de temps censurés seulement au moyen de données observées, en résolvant des équations d'estimation qui sont asymptotiquement sans biais. Nous démontrons la convergence forte au moyen de techniques d'analyse modernes et illustrons nos résultats par des simulations.

[Monday June 12/Lundi 12 juin, 10:50-11:05]

James Kim (Royal Canadian Air Forces), **Mohan Chaudhry** (Royal Military College of Canada)

New approach to GIX/Geom/1 Queues involving Heavy-Tailed distributions

Nouvelle approche pour les files d'attente GIX/Geom/1 comportant des distributions à queue lourde

We provide an analytically simple and computationally efficient solution to the discrete-time queueing model GIX /Geom/1. Using the imbedded Markov chain method, the analysis has been carried out for the late-arrival system and the results for the early-arrival system have been derived from those of the late-arrival system. The probability distributions of the numbers in the system at various epochs are all found in terms of roots of the underlying characteristic equation. Numerical results are also discussed.

Nous fournissons une solution analytiquement simple et informatiquement efficace au modèle de files d'attente à temps discret GIX/Geom/1. En utilisant la méthode des chaînes de Markov intégrées, l'analyse a été effectuée sur le système d'arrivée tardive et les résultats pour le système d'arrivée précoce ont été dérivés de ceux du système d'arrivée tardive. Nous trouvons les distributions de probabilité des nombres dans le système à différentes époques en termes de racines de l'équation caractéristique sous-jacente. Nous discutons aussi des résultats numériques.

[Monday June 12/Lundi 12 juin, 11:05-11:20]

Yueleng Wang (University of Windsor), **Sévérien Nkurunziza** (University of Windsor)

On Convergence of Sample Correlation Matrices in High-Dimensional Data

Convergence de matrice des corrélations de l'échantillon avec données à haute dimension

In this talk, we consider an estimation problem concerning the matrix of correlation coefficient in context of high dimensional data settings. In particular, we revisited some results in Li and Rolsalsky [Li, D. and Rolsalsky, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices, The Annals of Applied Probability, 16, 1, 423-447]. We extend one of the findings in Li and Rosalsky (2006) and, we simplify remarkably the proof of one of the result of the quoted paper. Further, we generalize a theorem which is useful in deriving the existence of the p th moment as well as in studying the convergence rates in law of large numbers.

Dans cet exposé, nous abordons un problème d'estimation concernant la matrice des coefficients de corrélation dans un contexte de données à haute dimension. En particulier, nous revisitons quelques résultats de Li et Rolsalsky [Li, D. and Rolsalsky, A. (2006), Some strong limit theorems for the largest entries of sample correlation matrices, The Annals of Applied Probability, 16, 1, 423-447]. Nous élargissons un des résultats de Li et Rolsalsky (2006) et nous simplifions remarquablement la preuve d'un des résultats de l'article cité. De plus, nous généralisons un théorème qui est utile pour dériver l'existence du p ième moment ainsi que pour étudier les taux de convergence dans la loi des grands nombres.

[Monday June 12/Lundi 12 juin, 11:20-11:35]

Yoshihiro Taniguchi (University of Waterloo), **Christiane Lemieux** (University of Waterloo)

Importance Sampling Techniques for Single Index Models

Techniques d'échantillonnage préférentiel pour les modèles à un seul indice

In rare-event simulation, a plain Monte Carlo estimator is bound to be imprecise since the samples from the underlying distribution has a low chance of landing on the important region. Importance sampling (IS) is a popular variance reduction technique in these situations. IS draws samples from a proposal distribution which is constricted in such a way that it places heavier weights on the important region. The design of the effective proposal distribution, however, is generally hard when many variables are involved as it becomes harder to characterize which part of the domain corresponds to the important region. Fortunately, any high-dimensional

Dans le cadre d'une simulation d'événements rares, une estimation simple selon la méthode de Monte-Carlo ne pourrait être qu'imprécise puisque les échantillons issus de la distribution sous-jacente ont peu de chances d'atterrir sur la région importante. L'échantillonnage préférentiel est une technique populaire de réduction de variance dans ces situations. L'échantillonnage préférentiel extrait simplement des échantillons à partir d'une distribution proposée qui est restreinte de telle manière qu'elle place les pondérations les plus lourdes sur la région importante. Cependant, il est en général difficile de concevoir efficacement la distribution proposée lorsque plusieurs variables sont utilisées puisqu'il devient plus difficile de caractériser la partie du domaine qui correspond

problems have a low-dimensional structures. In many cases, one-dimensional projection (single index) of the input vector captures a large majority of overall variance. Our IS technique exploits such a low-dimensional structure of those problems and constructs an efficient proposal distribution.

à la région importante. Heureusement, tous les problèmes de grandes dimensions ont des structures de faible dimension. Souvent, la projection unidimensionnelle (à un seul indice) du vecteur d'entrée capture la majorité de l'écart total. Notre technique d'échantillonnage préférentiel exploite une telle structure de faible dimension de ces problèmes et permet de construire une distribution proposée efficace.

Chair/Président: Kevin E. Thorpe
Room/Salle: E2 150 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 10:20-10:35]

Zihang Lu (University of Toronto)

A Comparison of Mixture Modeling Approaches for Clustering Longitudinal Binary Data

Comparaison d'approches de modélisation par mélange pour le regroupement de données binaires longitudinales

Mixture modelling has been increasingly used in clinical and epidemiological research for clustering patients into different developmental trajectories. Many previous studies have been concerned with continuous data, whereas only a limited number of studies focus on evaluating appropriate approaches for clustering binary data. Using simulated data, this study aims to investigate several mixture modeling approaches, namely the standard latent class model, the growth mixture model, the K-means mixture model, and a Bayesian latent class model. Various scenarios with differing sample sizes, number of classes, class sizes, and development trajectories shapes are considered. We focus on evaluating the performance of these approaches in recovering the true class membership and development trajectories for longitudinal binary responses. For illustration, we also present a case study using real data.

La modélisation par mélange est de plus en plus utilisée en recherche clinique et épidémiologique pour regrouper les patients dans différentes trajectoires de développement. De nombreuses études antérieures se sont intéressées aux données continues, alors qu'un nombre limité d'études se concentrent sur l'évaluation des approches appropriées pour le regroupement des données binaires. En utilisant des données simulées, cette étude cherche à étudier plusieurs approches de modélisation par mélange, à savoir le modèle de classes latentes standard, le modèle de mélange de croissance, le modèle de mélange K-means et un modèle bayésien de classes latentes. Nous considérons différents scénarios avec différentes tailles d'échantillons, nombre de classes, taille des classes et formes des trajectoires de développement. Nous nous concentrons sur l'évaluation de la performance de ces approches pour récupérer la véritable classe d'appartenance et les trajectoires de développement pour les réponses binaires longitudinales. À titre illustratif, nous présentons une étude de cas utilisant des données réelles.

[Monday June 12/Lundi 12 juin, 10:35-10:50]

Amadou Diogo Barry (Université du Québec à Montréal), **Karim Oualkacha** (Université du Québec à Montréal), **Arthur Charpentier** (Université de Rennes)

Quantile and Expectile Regression for Random Effects Models

La régression quantile et expectile appliquées au modèle avec effets aléatoires

Quantile and expectile regression models pertain to the estimation of unknown quantiles/expectiles of the cumulative distribution function of a dependent variable as a function of a set of covariates and a vector of regression coefficients. Both approaches make no assumption on the shape of the distribution of the response variable, allowing for investigation of a comprehensive class of covariate effects. We develop methods for panel data within a random effects framework. We provide asymptotic properties of the underlying model parameter estimators and suggest appropriate estimators of their variances-covariances matrices. The performance of the suggested estimators is evaluated through sim-

Les régressions quantile et expectile appartiennent à la famille des méthodes semi paramétriques. Elles permettent l'estimation de l'impact des variables explicatives sur les quantiles et les expectiles de la variable dépendante. Nous avons développé les deux méthodes pour un modèle avec effets aléatoires. Nous avons étudié les propriétés asymptotiques des estimateurs et avons proposé des estimateurs de leur matrice de variance covariance sous l'hypothèse d'indépendance et de distribution identique (iid) des résidus. Les deux méthodes sont appliquées à des données simulées et leur performance est illustrée avec des données réelles. Les résultats simulés montrent que les expectiles sont facilement estimables et ont des qualités statistiques comparables aux quantiles. En conclusion, comme les expectiles sont à la moyenne ce

Regression Methods: Means, Modes, and Quantiles

Méthodes de régression : moyennes, modes et quantiles

ulation studies and the methodology is illustrated using real data. The simulations show that expectile regression is comparable to quantile regression, easily computable and has relevant statistical properties. In conclusion, expectiles are to the mean what quantiles are to the median, and should be used and interpreted as quantilized mean.

que les quantiles sont à la médiane, les expectiles devraient être utilisés davantage et interprétés comme des moyennes quantilisées (quantilized mean).

[Monday June 12/Lundi 12 juin, 10:50-11:05]

Peijun Sang (Simon Fraser University), **Jiguo Cao** (Simon Fraser University), **Richard Lockhart** (Simon Fraser University)

Component Selection and Estimation for Functional Semiparametric Additive Model

Sélection et estimation des composantes d'un modèle additif fonctionnel semiparamétrique

We propose a functional semiparametric additive model which incorporates the effects of functional covariates and scalar covariates. More specifically, the effect of a functional covariate is modeled non-parametrically while a linear form is adopted to model the effects of the scalar covariates. This strategy can enhance flexibility in modelling the effect of the functional covariate while maintaining interpretability for the effects of scalar covariates. Additionally, an estimation method is developed to smooth and select non-vanishing components when estimating the nonparametric components. The coefficients for the scalar covariates can be obtained simultaneously by our proposed method. The finite sample performance of our proposed model and the corresponding estimation method are compared with various methods using simulation studies. Our proposed method is demonstrated by a real application of predicting the protein content in meat samples using Tecator absorbance spectra and fat and water contents of the meat samples.

Nous proposons un modèle fonctionnel semiparamétrique additif qui intègre les effets des covariables fonctionnelles et des covariables scalaires. Plus précisément, l'effet d'une covariable fonctionnelle est modélisé de manière non paramétrique tandis qu'une forme linéaire est adoptée pour modéliser les effets des covariables scalaires. Cette stratégie peut accroître la souplesse dans la modélisation de l'effet de la covariable fonctionnelle tout en maintenant l'interprétation des effets des covariables scalaires. De plus, une méthode d'estimation est développée pour lisser et sélectionner les composantes qui demeurent lors de l'estimation des composantes non paramétriques. Les coefficients des covariables scalaires peuvent être obtenus simultanément à l'aide de la méthode proposée. Nous comparons la performance sous échantillons finis du modèle proposé et la méthode d'estimation correspondante à diverses méthodes à l'aide d'études de simulation. Nous illustrons notre méthode à l'aide d'une application réelle d'estimation de la teneur en protéines dans des échantillons de viande en utilisant des spectres d'absorbance de Tecator et d'estimation des teneurs en graisse et en eau des échantillons de viande.

[Monday June 12/Lundi 12 juin, 11:05-11:20]

Bouchra Nasri (HEC Montreal), **Bruno Rémillard** (HEC Montréal), **Taoufik Bouezmarni** (Université de Sherbrooke)

On Copula-Based Conditional Quantile Estimators

Estimateurs de quantiles conditionnels basés sur la copule

Recently, two different copula-based approaches have been proposed to estimate the conditional quantile function of a variable Y with respect to a vector of covariates X : the first estimator is related to quantile regression weighted by the conditional copula density, while the second estimator is based on the inverse of the conditional distribution function written in terms of margins and the copula.

Récemment, deux approches différentes basées sur la fonction copule ont été proposées pour estimer la fonction des quantiles conditionnels d'une variable Y par rapport à un vecteur de covariables X : le premier estimateur est lié au modèle de régression des quantiles pondéré par la densité de la copule conditionnelle, tandis que le second estimateur est basé sur l'inverse de la distribution conditionnelle écrite en termes des marges et de la copule. En s'appuyant sur théo-

Using empirical processes, we show that even if the two estimators look quite different, they converge to the same limit. We also propose a bootstrap procedure for the limiting process in order to be able to construct uniform confidence bands around the conditional quantile function. A case study based on hydro-climatic data illustrates the proposed methodology.

rie des processus empiriques, nous montrons que les deux estimateurs, même s'ils semblent différents, convergent en fait vers la même limite. Nous proposons aussi une méthode de ré-échantillonnage permettant la construction d'une bande de confiance uniforme autour de la fonction des quantiles conditionnels. Une étude de cas utilisant des données hydro-climatiques sert à illustrer la méthodologie proposée.

[Monday June 12/Lundi 12 juin, 11:20-11:35]

Menglu Che (University of Waterloo), **Linglong Kong** (University of Alberta), **Rhonda Bell** (University of Alberta), **Yan Yuan** (University of Alberta)

Trajectory Modeling of Gestational Weight: a Functional Principal Component Analysis Approach

Modélisation de l'évolution du poids pendant la grossesse : analyse en composante principale fonctionnelle

Suboptimal gestational weight gain (GWG), which is linked to increased risk of adverse outcomes for a pregnant woman and her infant, is prevalent. In a large cohort study of Canadian pregnant women, our goals are to estimate the individual weight growth trajectory using sparsely collected data, and to identify the factors affecting the total GWG. The first goal was achieved through functional principal component analysis (FPCA) by conditional expectation. For the second goal, we used linear regression with the total weight gain as the response variable. The trajectory modeling through FPCA had significantly smaller mean square error and improved adaptability than the traditional nonlinear mixed-effect models, demonstrating a novel tool to facilitate real-time monitoring and interventions of GWG. Our regression analysis showed that prepregnancy body mass index (BMI) had a high predictive value for the total GWG, which agrees with the published weight gain guideline.

Une prise de poids sous-optimale pendant la grossesse (GWG), qui est associé à un risque accru de répercussions néfastes pour la femme enceinte ainsi que pour son enfant, est fréquente. Dans une vaste étude sur des femmes enceintes canadiennes, nos objectifs sont d'estimer l'évolution croissante du poids individuel en utilisant des données recueillies de façon clairsemée et d'identifier les facteurs influant sur le GWG total. Le premier objectif a été atteint par une analyse en composante principale fonctionnelle (FPCA) par espérance conditionnelle. Pour le deuxième objectif, nous avons utilisé une régression linéaire avec la prise de poids totale comme variable dépendante. La modélisation de l'évolution par la FPCA avait une erreur quadratique moyenne considérablement plus petites et une adaptabilité améliorée, comparé aux modèles traditionnels à effet mélangés non linéaires, illustrant un nouvel outil pour faciliter la surveillance en temps réel et les interventions sur le GWG. Notre analyse de régression démontre que l'indice de masse corporel (IMC) mesuré avant la grossesse est une forte valeur prédictive pour le GWG total, ce qui est en accord avec la ligne directrice publiée sur la prise de poids.

[Monday June 12/Lundi 12 juin, 11:35-11:50]

Myrtha E. Reyna Vargas (University of Toronto), **Nicholas Mitsakakis** (University of Toronto)

Ordinal Regression for the Analysis of Health Utility Data

Régression ordinale pour l'analyse de données de cotation d'état de santé

Health utility (HU) data are negatively skewed and bounded by 1, with most observations lying close to that bound and some with extremely low levels. HU has been analyzed by linear regression (OLS) which assumes normally distributed errors and lacks a restriction at the upper limit; Ordinal Regression (OR) is not limited by such assumptions and it produces estimates bounded at 1. This study

Les données de cotation d'état de santé (Health Utility, HU) sont négativement asymétriques et bornées à 1, tandis que la plupart des observations sont proches de cette limite et certaines avec de très faibles niveaux. Les données HU ont été analysées avec une régression linéaire ordinale (OLS) qui suppose des erreurs normalement distribuées et l'absence d'une restriction à la limite supérieure. La régression ordinale (OR) n'est pas limitée par de telles assertions et elle produit des

compares the performance of OR to OLS for the prediction of HU given covariates, using simulated and real data of prostate cancer patients. Using various sample sizes and three different distributions, HU data were simulated and analyzed by OLS and OR. Models were evaluated by the bias and coverage probability of the estimated mean, and by using the Root Mean Square Error (RMSE) comparing simulated and predicted HU values. Results show that OR provides more accurate estimates regardless of sample size, while OLS bias increases as HU approaches zero. Coverage probability is similar in both methods.

estimations bornées à 1. Cette étude sert à comparer la performance de OR par rapport à OLS pour la prédiction de HU à partir de covariables, à l'aide de données réelles et simulées relatives à des patients atteints de cancer de la prostate. En utilisant des échantillons de taille variée et trois distributions différentes, des données HU ont été simulées et analysées avec une régression OLS ou OR. Les modèles ont été évalués en fonction du biais et du taux de couverture de la moyenne estimée, ainsi que la par la racine de l'écart quadratique moyen par comparaison des données HU simulées et prédites. Les résultats indiquent que OR fournit des estimations plus précises, quelle que soit la taille de l'échantillon, tandis que le biais de OLS augmente à mesure que les données HU approchent la valeur nulle. Les deux méthodes ont une probabilité de couverture similaire.

Recent Developments in Biostatistics Récents développements en biostatistique

Chair/Président: Xikui Wang
Organizer/Responsable: Xikui Wang
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-14:00]

Peter X. Song (University of Michigan), **Ling Zhou** (University of Michigan)

Confidence Estimating Functions for Data Integration

Fonctions d'estimation des intervalles de confiance pour l'intégration de données

The theory of statistical inference along with the strategy of divide-and-combine for large-scale data analysis has recently attracted considerable interest due to the popularity of the MapReduce scheme. The key to the development of statistical inference lies in the method of combining results yielded from separately mapped data batches. We consider a general inferential methodology based on estimating functions, which allows us to perform regression analyses of massive complex data via the MapReduce scheme, such as longitudinal data, survival data and quantile regression, which cannot be done using the maximum likelihood method. The proposed statistical inference inherits many key large-sample properties of estimating functions. Also we show that the proposed method is closely connected to the generalized method of moments (GMM). Our method provides a unified framework for many kinds of statistical models and data types, which is illustrated via numerical examples in both simulation studies and real-world data analyses.

La théorie de l'inférence statistique ainsi que la stratégie d'analyse de données à grande échelle de division et de combinaison a récemment suscité un grand intérêt en raison de la popularité du schéma MapReduce. La clé du développement d'une inférence statistique réside dans la méthode de combinaison des résultats produits à partir de lots de données cartographiés séparément. Nous nous penchons sur une méthodologie inférentielle générale qui s'appuie sur des fonctions d'estimation. Ces fonctions nous permettent d'effectuer, au moyen de MapReduce, des analyses de régression de grandes quantités de données complexes, comme des données longitudinales, des données de survie et une régression quantile qui ne peuvent pas être effectuées avec la méthode de maximum de vraisemblance. L'inférence statistique proposée hérite de nombreuses propriétés essentielles sous grands échantillons des fonctions d'estimation. Nous montrons également que la méthode proposée est étroitement liée à la méthode généralisée des moments. Nous illustrons notre méthode, qui offre un cadre unifié pour de nombreux genres de modèles statistiques et de données, par des exemples numériques d'études et d'analyses de données réelles et de simulation.

[Monday June 12/Lundi 12 juin, 14:00-14:30]

Mikelis Guntars Bickis (University of Saskatchewan), **Naeima Ashleik** (University of Saskatchewan), **Juxin Liu** (University of Saskatchewan)

Inference from Survival Data using Imprecise Probabilities

Inférence à partir de données de survie au moyen de probabilités imprécises

Imprecise probability (IP) is a generalization of Finetti's approach to probability, formalized by Williams (1975) and extensively discussed in a monograph by Walley (1991), who presented these ideas in a JRSS discussion paper (1996). The methodology can be interpreted as Bayesian sensitivity analysis using a set of priors. In his 1996 presentation, Walley introduced a practical form for this inference on discrete data by using a family of Dirichlet priors with a fixed concentration parameter. He proposed that it could also be used for

La probabilité imprécise est une généralisation de l'approche probabiliste de de Finetti que Williams a formalisée (1975) et que Walley a examinée en profondeur dans une monographie (1991) avant de présenter ses idées dans un article de discussion dans le Journal of the Royal Statistical Society (1996). La méthodologie peut être interprétée comme une analyse de sensibilité de Bayes au moyen d'un ensemble de lois a priori. Dans cet article de 1996, Walley a présenté une forme pratique d'inférence sur des données discrètes au moyen d'une famille de lois a priori de Dirichlet et d'un paramètre de concentration fixe. Il a également proposé d'utiliser

Recent Developments in Biostatistics Récents développements en biostatistique

survival data by discretization. Coolen (1997) extended Walley's model to allow for censoring, and Coolen and Yan (2004) proposed a nonparametric predictive inference paradigm which also led to IP conclusions. Bickis (2009) used IP concepts to estimate the hazard function. After reviewing the methodology, we will be presenting an IP version of the log-rank test, which will produce a sequence of imprecise posteriors updated at the time of each death.

liser cette forme d'inférence pour des données de survie au moyen d'une discrétisation. Coolen (1997) a étendu le modèle de Walley et Yan (2004) pour permettre une censure, et Coolen et Yan (2004) ont proposé un paradigme d'inférence prédictive non paramétrique qui a également mené à des conclusions de probabilité imprécise. Bickis (2009) a utilisé des concepts de probabilité imprécise pour estimer la fonction de risque. Après avoir passé examiné la méthodologie, nous présenterons une version de probabilité imprécise du test logarithmique par rangs qui produira une séquence de données a posteriori imprécise au moment de chaque décès.

[Monday June 12/Lundi 12 juin, 14:30-15:00]

J. Jack Lee (University of Texas MD Anderson Cancer Center)

Bayesian Adaptive Designs for Efficient Drug Development and Precision Oncology

Plans d'expérience adaptatifs bayésiens pour le développement de médicaments efficaces et une oncologie de précision

Clinical trial is a prescribed learning process. Bayesian methods take the “learn as we go” approach and are uniquely suitable for such learning. In recent years, rapid advancements in medicine demand innovative methods to identify better therapies and the most appropriate population in a timely and efficient way. I will first illustrate the concept of Bayesian update and Bayesian inference, then, give an overview of Bayesian adaptive designs in dose finding, predictive probability, multi-arm platform design, and outcome adaptive randomization, etc. Applications including BATTLE trials in lung cancer and I-SPY trials in breast cancer will be given. Bayesian adaptive designs increase the study efficiency, allow more flexible trial conduct, and treat more patients with more effective treatments in the trial but also possess desirable frequentist properties. Perspectives will be given on future development in trial design, conduct, and evaluation to streamline and speed up drug approval.

L'essai clinique est un processus d'apprentissage prescrit. Les méthodes bayésiennes adoptent l'approche « d'apprentissage au fil du temps » et sont singulièrement adaptées à cet apprentissage. Au cours des dernières années, les progrès considérables réalisés dans le domaine médical ont nécessité des méthodes innovantes afin de trouver le plus rapidement et le plus efficacement possible de meilleures thérapies, ainsi que la population la plus appropriée. Je vais d'abord illustrer le concept de la mise à jour et de l'inférence bayésiennes, puis je donnerai un aperçu des plans d'expérience adaptatifs bayésiens en ce qui concerne l'établissement de la posologie, la probabilité prédictive, le plan par plateforme multi-bras et la randomisation adaptative à la réponse, etc. Nous présenterons des applications, notamment les essais BATTLE sur le cancer du poumon et les essais I-SPY sur le cancer du sein. Des plans d'expérience adaptatifs bayésiens permettent non seulement d'augmenter l'efficacité de l'étude, de mener plus facilement des essais et de traiter davantage de patients au moyen de traitements plus efficaces dans l'essai, mais ils possèdent également des propriétés fréquentistes souhaitables. Nous présenterons les perspectives en ce qui concerne les progrès à venir dans la conception, la conduite et l'évaluation d'essais pour faciliter et accélérer l'approbation des médicaments.

Time Series Analysis with Applications in Brain Activity Analysis

Analyse de séries chronologiques avec applications à l'analyse de l'activité du cerveau

Chair/Président: Hao Yu
Organizer/Responsable: Hao Yu
Room/Salle: E2 150 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-14:00]

Jörn Diedrichsen (University of Western Ontario)

Spatio-Temporal Noise Models for the Analysis of Functional Magnetic Resonance Imaging Data

Modèles de bruit spatiotemporels pour l'analyse des données d'imagerie par résonance magnétique fonctionnelle

Modern functional magnetic resonance imaging (fMRI) techniques provide measurements of ongoing activity in the whole human brain at an unprecedented spatial (1mm) and temporal (1s) resolution. Unfortunately, the measurements are corrupted by substantial noise, arising both from intermittent artifacts such as subject motion, as well as from non-neural physiological processes. Latter processes lead to characteristic low-frequency drifts in the times-series data. Noise processes also show a characteristic spatial structure that depends substantially on the underlying neuroanatomy. It has become clear that many current noise models implemented in commonly used fMRI analysis packages lead to suboptimal estimation of model parameters and do not provide adequate statistical inference. I will show some of the developments in our lab to improve spatio-temporal noise models and will demonstrate their impact on single-voxel analysis, multi-variate pattern analysis, and whole-brain functional connectivity analysis.

Les techniques modernes d'imagerie par résonance magnétique fonctionnelle (IRMf) fournissent des mesures de l'activité cérébrale dans son ensemble à une résolution spatiale (1 mm) et temporelle (1 s) sans précédent. Malheureusement, ces mesures sont altérées par un bruit de fond considérable, découlant tant d'artéfacts intermittents comme un mouvement du sujet que de processus physiologiques non neuraux. Ces derniers conduisent notamment à des effets de dérive de basse fréquence caractéristiques dans les données chronologiques. Les processus de bruit présentent aussi une structure spatiale caractéristique qui dépend largement de la neuroanatomie sous-jacente. Il est devenu évident que de nombreux modèles de bruit actuellement utilisés dans les logiciels d'analyse IRMF courants donnent une estimation sous-optimale des paramètres du modèle et ne permettent pas une inférence statistique adéquate. Je présente certains travaux de notre laboratoire visant à améliorer les modèles de bruit spatiotemporel et en montre l'impact sur l'analyse à voxel unique, l'analyse de structures multivariées et l'analyse de la connectivité fonctionnelle du cerveau.

[Monday June 12/Lundi 12 juin, 14:00-14:30]

Piotr Kokoszka (Colorado State University)

Change Point Detection in Functional Time Series Models for Yield Curves

Détection des points de changement dans les modèles de séries chronologiques fonctionnelles pour les courbes de rendement

Yield curves are functions defined on time to maturity with corresponding values equal to yield (interest) on a bond, typically a standardized government issued instrument. Yield curves are commonly used to predict future states of the economy on the basis of the interest investors demand for government debt of various maturities. These curves form a time series of functions, one function per day. The talk will discuss methods of detecting a change point in the mean function of such a functional time series. After reviewing recent research,

Les courbes de rendement sont des fonctions définies selon la durée de vie restante et dont les valeurs sont égales au rendement (intérêt) d'une obligation, généralement un instrument standardisé émis par le gouvernement. Ces courbes de rendement sont couramment utilisées pour prédire les états futurs de l'économie sur la base des taux d'intérêt exigés par les investisseurs pour les dettes gouvernementales à diverses échéances. Ces courbes forment une série chronologique de fonctions, avec une fonction par jour. Dans cette présentation, nous étudions diverses méthodes permettant de détecter un point de changement dans la fonction

Time Series Analysis with Applications in Brain Activity Analysis

Analyse de séries chronologiques avec applications à l'analyse de l'activité du cerveau

we will present two methods: one which uses a factor representation of the yield curves, the other a fully nonparametric method. Both methods permit the second order structure to change independently of the changes in the mean structure. Based on the asymptotic theory, two numerical approaches to the implementation of the tests will be presented and compared. Application to US Federal Reserve yield curves will be presented.

moyenne d'une telle série chronologique fonctionnelle. Après avoir passé en revue les dernières recherches, nous présentons deux méthodes : l'une utilisant une représentation factorielle des courbes de rendement, l'autre une méthode non paramétrique. Les deux méthodes permettent une modification de la structure d'ordre 2 indépendamment des modifications de la structure moyenne. Sur la base de la théorie asymptotique, nous présentons et comparons deux approches numériques de la mise en œuvre de ces tests. Nous présentons une application aux courbes de rendement de la Réserve fédérale américaine.

[Monday June 12/Lundi 12 juin, 14:30-15:00]

Herold Dehling (Ruhr University), **Alexander Schnurr** (University of Siegen)

Change-Point Detection in Time Series using Ordinal Patterns

Détection des points de changement dans les séries chronologiques par motifs ordinaux

We propose new concepts in order to analyse and model the dependence structure between two time series. Our methods rely exclusively on the order structure of the data points. Hence, the methods are stable under monotone transformations of the time series and robust against small perturbations or measurement errors. Ordinal pattern dependence can be characterised by four parameters. We propose estimators for these parameters, and we calculate their asymptotic distributions. Furthermore, we derive a test for structural breaks within the dependence structure. All results are supplemented by simulation studies and empirical examples.

Nous proposons de nouveaux concepts pour analyser et modéliser la structure de dépendance entre deux séries chronologiques. Nos méthodes dépendent uniquement de la structure d'ordre des points de données. Elles sont donc stables pour des transformations monotones de la série chronologique et robuste aux petites perturbations ou erreurs de mesure. La dépendance des motifs ordinaux peut se caractériser par quatre paramètres. Nous proposons des estimateurs pour ces paramètres et en calculons les distributions asymptotiques. Ensuite, nous déterminons un test pour les ruptures structurelles dans la structure de dépendance. Tous les résultats sont complétés par des études de simulation et des exemples empiriques.

Here Comes the Future? Data Science Programs for Undergraduate and Graduate Students
Voici l'avenir ? Programmes de sciences des données au premier cycle et aux cycles supérieurs

Chair/Président: Bruce Dunham
Organizer/Responsable: Bruce Dunham
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-14:15]

Paul Gustafson (University of British Columbia), **Jenny Bryan** (University of British Columbia), **Giuseppe Carenini** (University of British Columbia), **Vincenzo Coia** (University of British Columbia), **Giulio V. Dalla Riva** (University of British Columbia), **Michael A. Gelbart** (University of British Columbia), **Gail C. Murphy** (University of British Columbia), **Raymond T. Ng** (University of British Columbia), **Tiffany A. Timbers** (University of British Columbia)

Designing and Launching a Professional Master's Program in Data Science: Tales from the UBC Experience
Conception et lancement d'un programme professionnel de maîtrise en science des données : anecdotes tirées de l'expérience à l'UBC

The University of British Columbia recently launched a professional master's program in Data Science. The first cohort of students began their studies in September 2016. In this talk I will describe this program, and some of the challenges and successes encountered thus far. Points of emphasis will include our thinking on the choice of topics to include, the choice of prerequisites, and the choice of delivery methods. Since our students will be undertaking capstone projects throughout May and June 2017, I also plan to give a "live update" on how this is working.

L'Université de la Colombie-Britannique a lancé récemment un programme professionnel de maîtrise en science des données. Un premier groupe d'étudiants a entrepris ce programme en septembre 2016. Nous décrivons le programme et en révélons certaines des difficultés et réussites jusqu'à maintenant. Nous insisterons notamment sur les points suivants : nos réflexions sur le choix des sujets d'étude, des prérequis à l'admission et des méthodes d'enseignement. Comme les étudiants entreprendront des projets de fin d'études pendant les mois de mai et juin 2017, il est prévu de donner une mise à jour en direct de leur progression.

[Monday June 12/Lundi 12 juin, 14:15-15:00]

Stefan Steiner (University of Waterloo)

Data Science Programs at the University of Waterloo
Des programmes en science des données à l'Université de Waterloo

In this talk I will describe the new undergraduate and Masters' Data Science programs at the University of Waterloo. I will discuss both programs' genesis, motivation, target audience as well as challenges encountered along the way. This session will conclude with discussion of Data Science generally and the issues and questions in starting a Data Science program.

Dans le cadre de cette allocution, nous décrirons les nouveaux programmes de premier cycle et de maîtrise en science des données à l'Université de Waterloo. Il sera question de l'origine et de la raison d'être des programmes, du public cible ainsi que des difficultés pendant le processus d'élaboration. Nous conclurons avec un aperçu général de la science des données, abordant aussi les difficultés et questions liées à l'instauration d'un programme dans cette discipline.

Developments in Event History Analysis with Informative Censoring or Observation Schemes

Développements en analyse de l'historique des événements avec schémas de censure ou d'observation informative

Chair/Président: Hua Shen
Organizer/Responsable: Hua Shen
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-14:00]

Yayuan Zhu (University of Texas MD Anderson Cancer Center), **Jerald F. Lawless** (University of Waterloo)

Event History Analysis of Data from Non-Ignorable Observation Schemes

Analyse historique de données à partir de plans d'observation non ignorables

Observational databases on events related to health and disease processes are widely used to study human health, but can pose challenges for rigorous analysis. These are due in part to selection effects related to inclusion in a database and to missing information on important factors. In addition, the times at which data are observed on an individual may be related to outcomes and covariates under study. Such selection and observation processes usually include elements that are non-ignorable, and failure to adjust for them can produce biased analysis. This talk will discuss likelihood and weighted estimating function methods for dealing with non-ignorable observation processes. Applications to the analysis of data from a psoriatic arthritis cohort will be considered.

Des bases de données d'observation sur des événements liés à des processus morbides et à la santé sont largement utilisées, mais elles peuvent rendre les analyses rigoureuses difficiles. Ces difficultés sont en partie causées par les effets de sélection liés à l'inclusion dans une base de données et au manque d'information quant à des facteurs importants. De plus, les moments auxquels les données sont observées chez un individu peuvent être reliés à des résultats et à des covariables qui font l'objet d'une étude. De tels processus d'observation et de sélection englobent généralement des éléments non ignorables, et l'absence de corrections pour en tenir compte risque de biaiser l'analyse. Dans cet exposé, nous nous intéresserons aux méthodes utilisant des fonctions d'estimation pondérées et fondées sur la vraisemblance afin de prendre en compte les processus d'observation non ignorables. Nous examinerons les applications de cette analyse de données à partir d'une cohorte de polyarthrite psoriasique.

[Monday June 12/Lundi 12 juin, 14:00-14:30]

X. Joan Hu (Simon Fraser University), **Huijing Wang** (Apple Inc.), **John Spinelli** (BC Cancer Agency)

Risk Classification and Prediction with Administrative Health Data

Classification et prédiction des risques avec des données administratives sur la santé

The population of cancer survivors has been increasing rapidly as a result of advances in cancer treatment. Consequently there has been greatly increased interest in evaluating late mortality and morbidity of survivors and their health care utilization. This talk presents strategies for risk classification and prediction with the counts of physician visits and the associated costs of cancer survivors from administrative health databases. We illustrate the methodology using the physician claims associated with the subjects in the CAYACS survivor cohort with the BCCA (Cancer Agency of British Columbia, Canada).

La population des survivants du cancer a rapidement augmenté en raison des avancées dans le traitement du cancer. Par conséquent, on s'est de plus en plus intéressé à l'évaluation de la mortalité et de la morbidité tardives chez les survivants, et de leur utilisation des services de soins de santé. Cet exposé présente les stratégies de classification et de prédiction des risques au moyen des nombres de visites des survivants du cancer chez le médecin et des coûts connexes à partir de bases de données administratives sur la santé. Nous illustrons la méthodologie au moyen de demandes de remboursement de médecins associées aux sujets dans la cohorte de survivants du programme de recherche CAYACS (Childhood, Adolescent, Young Cancer Survivorship – programme de recherche de la Colombie-Britannique sur les survivants du cancer enfants, adolescents et jeunes adultes) avec la Cancer Agency of British Columbia, Canada.

Developments in Event History Analysis with Informative Censoring or Observation Schemes

Développements en analyse de l'historique des événements avec schémas de censure ou d'observation informative

[Monday June 12/Lundi 12 juin, 14:30-15:00]

Leilei Zeng (University of Waterloo), **Nathalie Moon** (University of Waterloo), **Richard J. Cook** (University of Waterloo)

Design and Relative Efficiency of Tracing Studies in Cohorts with Loss to Followup

Conception d'études de traçabilité et efficacité relative de celles-ci dans des cohortes avec pertes au suivi

Considerable investments are being made to conduct large-scale cohort studies which involve routine followup of participants. Loss-to-followup is common particularly when the study duration is long. Standard likelihood methods ignoring the missing data yield consistent but less efficient estimators under the Sequential Missing at Random (SMAR), whereas bias arises when this assumption is violated unless the missing data are modelled correctly. Tracing the disease status of those lost to followup help to address the missing data problem and lead to potential efficiency gain, however relatively little attention has been given to the design of such studies. This work proposes selection models for tracing based on the history of the disease process, and such models are optimized to achieve the maximum efficiency gains under a fixed sample size or budget. An extension to develop an adaptive two-phase tracing design will also be discussed for the loss-to-followup under non-SMAR.

Des investissements considérables ont été réalisés pour mener des études de cohortes à grande échelle incluant des suivis de routine des participants. Les pertes au suivi sont courantes, notamment lorsque la durée de l'étude est longue. Les méthodes classiques fondées sur la vraisemblance, qui ignorent les données manquantes, produisent des estimateurs convergents, mais qui sont moins efficaces dans le cadre de données manquantes séquentielles au hasard, tandis qu'un biais survient lorsque cette hypothèse est violée, à moins que les données manquantes soient modélisées correctement. La récolte d'information concernant l'état de santé des participants perdus au suivi contribue à régler le problème des données manquantes et a mené à un gain d'efficacité potentiel, mais on a accordé peu d'attention à ces études. Cet exposé propose des modèles de sélection pour la récolte d'information concernant l'état de santé en fonction de l'historique du processus de la maladie, et de tels modèles sont optimisés pour réaliser le maximum de gains d'efficacité dans le cadre d'une taille fixe d'échantillon ou de budget. On examinera également une extension pour concevoir un plan d'expérience adaptatif en deux phases pour la récolte d'information concernant l'état de santé en présence de pertes au suivi qui ne sont pas des données manquantes séquentielles au hasard.

Recursive Partitioning Methods in Biostatistics

Méthodes de partitionnement récursif en biostatistique

Chair/Président: Liqun Diao
Organizer/Responsable: Liqun Diao
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-14:00]

Ruoqing Zhu (University of North Carolina at Chapel Hill), **Yifan Cui** (University of North Carolina at Chapel Hill), **Mai Zhou** (University of Kentucky), **Michael Kosorok** (University of North Carolina at Chapel Hill)

Some Asymptotic Results on Tree-based Survival Models

Quelques résultats asymptotiques de modèles de survie basés sur les arbres

There are only a handful of results available for tree-based survival models. We investigate the method from the perspective of splitting rules, where the log-rank test statistics is calculated and compared. The splitting rule is essentially treating both resulting child nodes as being identically distributed. However, we demonstrate that this approach is affected by censoring, which may lead to inconsistency of the method. Based on this observation, we develop an adaptive concentration bound in the sense that for each terminal node, the estimation centers around the true within-node average of the underlying survival function, while this quantity may again be affected by the censoring distribution. We then show that the consistency of the method by can be achieved if the splitting rule is modified to satisfy certain restrictions.

Il n'y a seulement que quelques résultats disponibles pour les modèles de survie basés sur les arbres. Nous étudions la méthode du point de vue des règles de fractionnement où la statistique de test log-rang est calculée et comparée. La règle de fractionnement traite les nœuds enfants comme étant distribués de façon identique. Nous démontrons que cette approche est affectée par la censure, ce qui peut mener à une incohérence de la méthode. Partant de ce constat, nous développons une limite de concentration adaptative, en ce sens que pour chaque nœud terminal l'estimation est centrée autour de la vraie moyenne intérieure du nœud de la fonction de survie sous-jacente, tout en permettant à cette quantité d'être de nouveau affectée par la loi de censure. Nous démontrons ensuite que la cohérence de la méthode peut être atteinte si la règle de fractionnement est modifiée pour satisfaire certaines restrictions.

[Monday June 12/Lundi 12 juin, 14:00-14:30]

Annette M. Molinaro (University of California, San Francisco), **Adam Olshen** (UCSF), **Robert Strawderman** (University of Rochester)

Expanding the Predictive Ability of an Interpretable Tree

Élargir la capacité prédictive des arbres d'interprétation

We recently developed partDSA, a multivariate method that, similarly to CART, utilizes loss functions to select and partition predictor variables to build a tree-like regression model for a given outcome. However, unlike CART, partDSA permits both 'and' and 'or' conjunctions of predictors, elucidating interactions between variables as well as their independent contributions. partDSA thus permits tremendous flexibility and provides an ideal foundation for developing a clinician-friendly tool for accurate risk prediction and stratification. An interesting extension of partDSA is as an aggregate learner (Olshen et al. (2017)). In this talk, we describe an approach to combining the predictive abil-

Nous avons récemment développé partDSA, une méthode multivariée qui, tout comme CART, utilise des fonctions de perte pour sélectionner et partitionner des variables prédictives pour construire un modèle de régression en arbre pour une variable dépendante donnée. Cependant, contrairement à CART, partDSA permet les conjonctions « et » et « ou » des prédicteurs, raffinant ainsi les interactions entre les variables ainsi que leurs contributions indépendantes. partDSA permet donc une grande flexibilité et offre une base idéale pour le développement d'un outil convivial pour les cliniciens pour la prédiction précise du risque ainsi que pour la stratification. Une extension intéressante de partDSA est en tant qu'apprenant global (Olshen et al. (2017)). Dans cet exposé, nous décrivons une approche qui combine la capacité prédic-

Recursive Partitioning Methods in Biostatistics

Méthodes de partitionnement récursif en biostatistique

ity of an aggregate learner with the clinical utility of a single tree. The method is illustrated via a data analysis for cancer patients based on various clinical and biomarker covariates.

tive d'un apprenant global avec l'utilité clinique d'un arbre unique. La méthode est illustrée par une analyse de données de patients cancéreux fondée sur différentes covariables cliniques et biomarqueurs.

[Monday June 12/Lundi 12 juin, 14:30-15:00]

Danping Liu (National Institutes of Health), **Jared C. Foster** (Eunice Kennedy Shriver National Institute of Child Health and Human Development and Mayo Clinic), **Paul S. Albert** (Eunice Kennedy Shriver National Institute of Child Health and Human Development and National Cancer Institute), **Aiyi Liu** (Eunice Kennedy Shriver National Institute of Child Health and Human Development)

Tree-Based Approaches for Personalized Risk Prediction with Longitudinal Biomarkers: An Application to Fetal Growth

Approches basées sur les arbres pour la prédiction de risques personnalisés avec biomarqueurs longitudinaux : une application à la croissance fœtale

Longitudinal monitoring of biomarkers is often helpful for predicting disease or a poor clinical outcome. Much research on longitudinal biomarkers has been on the overall accuracy in predicting an outcome. However, while good predictive accuracy may be found only in a subgroup of the population, identification of such subgroups can guide the design of follow-up schemes in future studies. In the Fetal Growth Study, we consider the prediction of both large and small for gestational age births by using longitudinal ultrasound measurements, and we attempt to identify subgroups of women for whom prediction is more (or less) accurate, should they exist. We propose tree-based approaches to identifying such subgroups, and a pruning algorithm that explicitly incorporates a desired type I error rate, allowing us to control the risk of false discovery of subgroups. The methods proposed are applied to data from the Scandinavian Fetal Growth Study and are evaluated via simulations.

L'observation longitudinale des biomarqueurs est souvent utile pour prédire une maladie ou de mauvais résultats cliniques. Plusieurs recherches sur les biomarqueurs longitudinaux portent sur l'exactitude globale de la prédiction d'un résultat. Cependant, bien qu'une bonne validité prédictive puisse se retrouver seulement dans un sous-groupe de la population, l'identification de tels sous-groupes peut guider la conception de projets ultérieurs lors d'études futures. Dans l'Étude de croissance fœtale, nous examinons la prédiction à la fois grande et petite pour les âges gestationnels à la naissance en utilisant des mesures échographiques longitudinales et nous tentons d'identifier des sous-groupes de femmes pour lesquels la prédiction est plus (ou moins) précise, si de tels sous-groupes existent. Nous proposons des approches basées sur des arbres pour l'identification de ces sous-groupes et un algorithme d'élagage qui incorpore explicitement un taux d'erreur de type I désiré, nous permettant de contrôler le risque de fausses découvertes de sous-groupes. Les méthodes proposées sont appliquées à des données de l'Étude de croissance fœtale scandinave et sont évaluées par simulation.

Survey Sampling Theory and Practice – A Celebration of J.N.K. Rao’s Research
Théorie et pratique de l’échantillonnage d’enquêtes – une célébration de la recherche
de J.N.K. Rao

Chair/Président: Changbao Wu
Organizer/Responsable: Changbao Wu
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-14:00]

Mahmoud Torabi (University of Manitoba), **Gauri Datta** (University of Georgia), **J.N.K. Rao** (Carleton University)

Small Area Estimation with Covariate Measurement Error

Estimation pour petits domaines avec erreurs de mesure des covariables

Small area estimation (SAE) has become a very active area of research in statistics. Many models studied in SAE focus on one or few variables of interest from a single survey without paying close attention to the treatment of covariates. It is useful to utilize the idea of “borrowing strength” from covariates to build a model which combines two (or multiple) surveys. In many real applications, there are also covariates measured with errors. In this talk, we study a nested error linear regression model with combining two surveys which has multiple unit level error-free covariates and multiple area level covariates subject to measurement error. In particular, we derive an empirical best predictor of a small area mean with corresponding mean squared prediction error estimation. The performance of the proposed approach is studied through simulation studies and also by a real application.

L’estimation pour petits domaines (SAE) est devenue un secteur de recherche statistique très dynamique. Mais de nombreux modèles en SAE mettent l’accent sur un ou quelques variables d’intérêt dans une enquête sans tenir compte du traitement des covariables. Il est utile « d’exploiter la force » des covariables pour construire un modèle qui combine deux (ou plusieurs) enquêtes. Dans bon nombre d’applications, certaines covariables présentent des erreurs de mesure. Dans cette présentation, nous étudions un modèle de régression linéaire à erreurs emboîtées pour combiner deux enquêtes qui contiennent de multiples covariables sans erreur au niveau unitaire et de multiples covariables au niveau du domaine qui sont sujettes à erreur de mesure. Plus particulièrement, nous dérivons un meilleur prédicteur empirique de la moyenne du petit domaine, avec une estimation correspondante de l’erreur quadratique moyenne de prédiction. Nous étudions la performance de l’approche proposée par des études de simulation et sur une application réelle.

[Monday June 12/Lundi 12 juin, 14:00-14:30]

Wesley Yung (Statistics Canada), **Mike Hidiroglou** (Statistics Canada), **Elisabeth Neusy** (Statistics Canada)

Non-response Follow-up for Business Surveys

Suivi de la non-réponse dans les enquêtes auprès des entreprises

Follow-up of nonrespondents in business surveys is a time and resource intensive activity. Given the decline in response rates, non-response follow-up takes on more and more importance to ensure continued quality of estimates produced. Given a fixed budget to follow-up non-responding units, what is the best way to select units for non-response follow-up in business surveys? Should all non-respondents be followed up or just a sample of them? If a sample is followed-up, how should it be selected? Should the sample be selected using simple random sampling (SRS), stratified SRS or probability proportional to size (PPS) sampling? These questions were addressed in two ways. Firstly, a simulation study

Le suivi des non-répondants aux enquêtes auprès des entreprises est une activité qui consomme énormément de temps et de ressources. Or, compte tenu de la baisse des taux de réponse, le suivi de la non-réponse s’impose de plus en plus si l’on veut maintenir la qualité des estimations produites. Pour un budget de suivi des unités non répondantes fixe, quelle est la meilleure façon de choisir les unités à suivre dans les enquêtes auprès des entreprises? Faut-il faire le suivi de tous les non-répondants ou seulement auprès d’un échantillon d’entre eux? Dans ce cas, comment sélectionner l’échantillon? Par sondage aléatoire simple (SAS), SAS stratifié ou probabilité proportionnelle de la taille? Nous répondons à ces questions de deux manières. D’abord, nous comparons par étude de simulation les biais de Monte-Carlo et l’erreur quadratique

Survey Sampling Theory and Practice – A Celebration of J.N.K. Rao’s Research

Théorie et pratique de l’échantillonnage d’enquêtes – une célébration de la recherche de J.N.K. Rao

compared the Monte Carlo biases and mean square error using data from an existing survey. Secondly, the exact follow-up sample size for a follow-up using simple random sampling and assuming uniform response rates was developed.

moyenne à l’aide de données tirées d’une enquête existante. Ensuite, nous définissons la taille précise de l’échantillon de suivi par sondage aléatoire simple, en considérant que les taux de réponse sont uniformes.

[Monday June 12/Lundi 12 juin, 14:30-15:00]

David Haziza (Université de Montréal), **Jean-François Beaumont** (Statistics Canada), **Isabelle Lefebvre** (Université de Montréal)

Simplified Variance Estimation in Surveys

Estimation de la variance simplifiée dans les enquêtes

Variance estimation presents challenges in the context of complex survey designs. Conventional variance estimators rely on second-order inclusion probabilities that can be difficult to compute for some sampling designs. Based on a procedure proposed by Ohlsson (1998) in the context of sequential Poisson sampling, we suggest a simplified variance estimator that requires only the first-order inclusion probabilities. The idea is to approximate an estimation strategy (which consists of a sampling design and a point estimator) by an equivalent strategy based on Poisson sampling. Proportional-to-size sampling designs and cluster designs will be discussed. Results of a simulation study will be presented.

L’estimation de la variance présente des défis dans les sondages à plan complexe. Les estimateurs de la variance traditionnels requièrent des probabilités d’inclusion d’ordre deux qui peuvent être difficiles à calculer pour certains plans d’échantillonnage. En nous inspirant d’une procédure proposée par Ohlsson (1998) dans le contexte de l’échantillonnage de Poisson séquentiel, nous suggérons un estimateur de la variance simplifié qui ne requiert que les probabilités d’inclusion d’ordre un. L’idée est d’approximer une stratégie d’estimation (qui consiste en un plan d’échantillonnage et un estimateur ponctuel) par une stratégie équivalente fondée sur l’échantillonnage de Poisson. Nous discutons des plans d’échantillonnage proportionnels à la taille d’échantillon et en grappe. Enfin, nous présentons les résultats d’une étude par simulation.

Chair/Président: Sanjeena Dang
Room/Salle: E2 155 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-13:45]

Wenjun Jiang (University of Western Ontario), **Jiandong Ren** (University of Western Ontario), **Hanping Hong** (University of Western Ontario)

Pareto-optimal Reinsurance with Two Constraints under Distortion Risk Measure

Réassurance optimale de Pareto avec deux contraintes sous une mesure du risque de distorsion

We study the Pareto-optimal reinsurance policies, where both the insurer's and the reinsurer's risks and returns are considered. We assume that the insurer and the reinsurer measure their risks by some distortion risk measures with possibly different distortion operators. In addition, the reinsurance premium is determined by a distortion risk measure. In our analysis, we suppose that a reinsurance policy is feasible only if the resulting risk of each party is below some pre-determined values. Methodologically, we show that the generalized Neyman-Pearson method, the Lagrange multiplier method, and the dynamic control methods can be utilized to solve the optimization problem with constraints. Special cases when both parties' risks are measured by Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) are studied in great details. An numerical example is provided to illustrate practical implications of the results.

Nous étudions les politiques de réassurance optimale de Pareto, qui tiennent compte des risques et des rendements de l'assureur et du réassureur. Nous supposons que l'assureur et le réassureur mesurent leurs risques par certaines mesures du risque de distorsion avec possiblement différents opérateurs de distorsion. En outre, la prime de réassurance est déterminée par une mesure du risque de distorsion. Dans notre analyse, nous supposons qu'une politique de réassurance n'est réalisable que si le risque résultant de chaque partie est inférieur à certaines valeurs prédéterminées. Méthodologiquement, nous montrons que la méthode de Neyman-Pearson généralisée, la méthode du multiplicateur de Lagrange et les méthodes de contrôle dynamique peuvent être utilisées pour résoudre le problème d'optimisation avec des contraintes. Les cas particuliers où les risques des deux parties sont mesurés en fonction de la valeur exposée au risque (VaR) et de la valeur exposée au risque extrême (TVaR) sont étudiés en détail. Les implications pratiques des résultats sont illustrées à l'aide d'un exemple numérique.

[Monday June 12/Lundi 12 juin, 13:45-14:00]

Danqiao Guo (University of Waterloo), **Tony Wirjanto** (University of Waterloo), **Chengguo Weng** (University of Waterloo)

Improved Global Minimum Variance Portfolio via a Dominant Eigenvalue Shrinkage

Portefeuille amélioré de variance minimale globale à l'aide d'une diminution de la valeur propre dominante

In this talk, we first discuss the conventional two-step method for constructing a Global Minimum Variance Portfolio (GMVP). It is noted that in a vast portfolio allocation problem, the conventional method fails to give us a portfolio with a minimum true risk due to a severe estimation error in the sample covariance matrix. Then we show that under a set of conditions for the true covariance matrix, it is possible to reduce the true portfolio risk by replacing the sample covariance matrix with a modified covariance matrix constructed from a shrunk dominant eigenvalue in the first step. Simulation and empirical results are used to illustrate the effect of

Dans cet exposé, nous abordons d'abord la méthode conventionnelle en deux étapes pour construire un portefeuille de variance minimale globale (PVMG). Il convient de noter que dans un vaste problème d'allocation de portefeuille, la méthode classique ne nous donne pas un portefeuille avec un risque réel minimal en raison d'une grave erreur d'estimation dans la matrice de covariance échantillonnale. Nous montrons ensuite que sous un ensemble de conditions pour la vraie matrice de covariance, il est possible de réduire le risque réel du portefeuille en remplaçant la matrice de covariance échantillonnale par une matrice de covariance modifiée construite en diminuant d'abord la valeur propre dominante. L'effet de la méthode de diminution est illustré à l'aide de résultats

Business and Actuarial Statistics Statistique commerciale et actuarielle

this shrinkage method.

simulés et empiriques.

[Monday June 12/Lundi 12 juin, 14:00-14:15]

Yu Gao (University of Waterloo)

Detecting Adverse Drug Effects from Pharmacovigilance Databases

Détection des effets secondaires des médicaments grâce aux bases de données de pharmacovigilance

The World Health Organization and many countries have built pharmacovigilance databases to detect potential adverse effects due to marketed drugs. Although a number of methods have been developed for early detection of adverse drug effects, the vast majority of them do not consider the multiplicity arising from testing thousands of drug and adverse event combinations. We first derive the optimal statistic to maximize the power of detection while maintaining the desired error rate. We then propose a nonparametric empirical Bayes method to estimate the optimal statistic and demonstrate its superior performance through simulation. Finally, the proposed method is applied to the pharmacovigilance database in the United Kingdom.

L'Organisation mondiale de la santé et de nombreux pays ont créé des bases de données de pharmacovigilance pour détecter les éventuels effets secondaires des médicaments commercialisés. Bien que diverses méthodes existent pour la détection précoce des effets secondaires des médicaments, la plupart ne tiennent pas compte de la multiplicité de possibilités découlant de milliers de médicaments et de combinaisons d'événements indésirables. Nous commençons par dériver la statistique optimale qui permet de maximiser la puissance de détection tout en maintenant le taux d'erreur voulu. Nous proposons ensuite une méthode bayésienne empirique non paramétrique qui permet d'estimer la statistique optimale et en montrons la performance supérieure par simulation. Enfin, nous appliquons la méthode proposée à la base de données de pharmacovigilance du Royaume-Uni.

[Monday June 12/Lundi 12 juin, 14:15-14:30]

Vanessa Bierling (McMaster University), **Paul D. McNicholas** (McMaster University)

A Latent Gaussian Mixture Model for Longitudinal Data

Un modèle de mélange gaussien latent pour des données longitudinales

We introduce a mixture model for clustering high-dimensional longitudinal data; that is, data containing measurements taken at a large number of time points. The model uses an extension of the mixture of common factor analyzers model to allow the variability in such measurements to be explained by measurements taken at a smaller number of time points. A modified Cholesky decomposition on the latent covariance matrix is utilized to take into account the longitudinal nature of the data.

Nous présentons un modèle de mélange pour le classement de données longitudinales de grandes dimensions, autrement dit, des données comportant des mesures relevées à plusieurs moments. Le modèle se veut une extension du modèle de mélange d'analyseurs de facteurs communs afin de permettre que la variabilité de ces mesures puisse s'expliquer par des mesures prises à des moments moins fréquents. Nous utilisons une décomposition Cholesky modifiée de la matrice des covariances latentes pour tenir compte de la nature longitudinales des données.

[Monday June 12/Lundi 12 juin, 14:45-15:00]

Nkumbuludzi Ndwapu (McMaster University), **Paul D. McNicholas** (McMaster University)

Robust Mixture Discriminant Analysis

Analyse discriminante de mélanges robuste

The purpose of this work is to develop robust discriminant analysis procedures making use of mixture models. This work is motivated by recent model-based clustering work. This includes work on mixtures of multivariate t-distributions, mixtures of multivariate power exponential distributions, and mixtures of contaminated Gaussian distributions. Parameter estimation is carried out us-

L'objectif de cette étude est de mettre au point des procédures d'analyse discriminante robustes à l'aide de modèles de mélanges. L'étude est motivée par de récents travaux sur le groupage par modèle, notamment sur les mélanges de distributions t multivariées, les mélanges de distributions de puissance exponentielles multivariées et les mélanges de distributions gaussiennes contaminées. Nous estimons les paramètres à l'aide de l'algorithme EM et sélectionnons le nombre de

ing the expectation-maximization algorithm, and the number of components per class is selected using the Bayesian information criterion. The proposed approaches are illustrated via simulated as well as real data.

composants par classe à l'aide du critère d'information bayésien. Nous illustrons les approches proposées via des données simulées et réelles.

Longitudinal and Survival Data Données longitudinales et de survie

Chair/Président: Tolulope T. Sajobi
Room/Salle: E2 160 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-13:45]

Jia Li (University of Calgary), **Alexander R. de Leon** (University of Calgary), **Haocheng Li** (University of Calgary)

A Longitudinal Linear Mixed Effect Model for Bivariate Responses with Measurement Errors and Misclassification in Covariates

Modèle linéaire longitudinal à effets mixtes pour réponses bidimensionnelles en présence d'erreurs de mesure et classification erronée dans les covariables

This talk concerns estimation of a longitudinal linear mixed effects model with bivariate responses mixed with continuous and binary variables in the presence of measurement errors in the covariates. We consider both continuous covariates as well as categorical covariates with misclassification up to three categories. Through simulation studies, we investigate the impact of measurement errors on the covariates on the naive estimator of each parameter in the model when the other parameters are fixed.

Cet exposé porte sur le modèle linéaire longitudinal à effets mixtes avec des réponses bidimensionnelles mélangées à des variables continues et binaires en présence d'erreurs de mesure dans les covariables. Nous considérons les covariables continues, ainsi que les covariables catégoriques avec une classification erronée jusqu'à trois catégories. Au moyen d'études de simulation, nous étudions les répercussions des erreurs de mesure des covariables sur l'estimateur naïf de chaque paramètre du modèle lorsque les autres paramètres sont fixés.

[Monday June 12/Lundi 12 juin, 13:45-14:00]

Harris Quach (Pennsylvania State University)

Accurate Confidence Intervals for Clustered Data and Small Samples

Intervalle de confiance adéquats pour données en grappes avec petits échantillons

The presence of clustering in data sets can impede statistical inference, especially for too few or small clusters of observations. To address the complications of clustered data, research has focused on adjustments to conventional distribution-free approaches in lieu of likelihood-based inference, which has been known to be sensitive for small samples. However, advances in likelihood theory offer an alternative solution to the clustering problem via higher-order approximations that are accurate, and can be attractive when working with small samples. I demonstrate that higher-order likelihood methods perform comparably well to some current distribution-free approaches in terms of attaining the appropriate coverage and can be preferable in terms of power. Some illustrative examples using the Tennessee Student-Teacher Achievement Ratio (STAR) dataset are provided as a context for comparison of higher-order likelihood and distribution-free approaches.

La présence de grappes dans les jeux de données peut compromettre l'inférence statistique, notamment pour les regroupements des observations trop peu nombreux ou trop petits. Afin de résoudre les problèmes de données en grappes, les recherches ont mis l'accent sur des ajustements à des approches non paramétriques classiques au lieu de l'inférence fondée sur la vraisemblance, qui était reconnue pour être sensible aux petits échantillons. Cependant, les progrès réalisés dans la théorie de vraisemblance offrent une autre solution au problème de mise en grappes au moyen d'approximations d'ordre supérieur qui sont précises et qui peuvent être intéressantes lorsqu'on travaille avec des petits échantillons. Je démontre que les méthodes de vraisemblance d'ordre supérieur fonctionnent relativement bien par rapport aux approches non paramétriques actuelles en ce qui concerne l'obtention d'une couverture appropriée et qu'elles peuvent être préférables en ce qui concerne la puissance. Je présenterai quelques exemples d'illustration au moyen du jeu de données provenant du projet Student Teacher Achievement Ratio (taux de réussite des élèves et des enseignants) de l'État du Tennessee que nous utiliserons comme contexte de comparaison d'ap-

Longitudinal and Survival Data Données longitudinales et de survie

proches de vraisemblance d'ordre supérieur et d'approches non paramétriques.

[Monday June 12/Lundi 12 juin, 14:00-14:15]

Yidan Shi (University of Waterloo), **Leilei Zeng** (University of Waterloo), **Mary E. Thompson** (University of Waterloo), **Suzanne Tyas** (University of Waterloo)

Incorporating Auxiliary Information by Joint Modelling of Pseudo Data and Length Biased Data

Intégration d'informations auxiliaires par modélisation conjointe de pseudo-données et données biaisées en longueur

Survival data are often limited in sample size, subject to length-biased sampling, or limited to a specific time/age range. Information for a sample from the same or similar cohorts, having more observations or a longer observation time, may be obtained from the literature. However, we may not be able to access the source data, or may have trouble combining the two samples. We introduce a way to incorporate the information from the auxiliary sample by creating pseudo datasets which match the form and conditions of the observed sample, eliminating the length-biased sampling problem and improving efficiency. We will illustrate the method with a real dataset.

Les données de survie sont souvent limitées par la taille d'échantillon, par un échantillonnage biaisé en longueur ou encore par une plage de temps/âges spécifique. On peut parfois obtenir des informations de la littérature pour un échantillon de la même cohorte ou d'une cohorte similaire qui contient plus d'observations ou une durée d'observation plus longue. Cependant, il n'est pas toujours possible d'accéder aux données sources ou de combiner les deux échantillons. Nous présentons une façon d'intégrer les informations de l'échantillon auxiliaire en créant des pseudo-ensembles de données qui correspondent à la forme et aux conditions de l'échantillon observé, éliminant ainsi le problème d'échantillonnage biaisé en longueur et améliorant l'efficacité. Nous illustrons la méthode sur un ensemble de données réelles.

[Monday June 12/Lundi 12 juin, 14:15-14:30]

Kunasekaran Nirmalkanna (Memorial University of Newfoundland), **Candemir Cigsar** (Memorial University of Newfoundland)

Analysis of Recurrent Events with Dynamic Models for Dependent Gap Times: A Copula Approach

Analyse des événements récurrents à l'aide de modèles dynamiques pour les temps d'écart dépendants : une approche par copules

The analysis of past developments of processes through dynamic covariates is useful to understand the present and the future of processes generating recurrent events. In this study, we consider two important features of such processes through dynamic models. These features are related to monotonic trends and clustering of events in recurrent event data and are common in medical studies. We discuss certain issues in the estimation of these features through dynamic models based on event counts when unexplained heterogeneity is present in the data. Furthermore, we show that the violation of the strong assumption of independent gap times may introduce substantial bias in the estimation of these features with models based on event counts. To address these issues, we therefore apply a copula-based estimation method for the gap time models. This approach does not rely on the

L'analyse des développements passés des processus par des covariables dynamiques est utile pour comprendre le présent et l'avenir des processus générant des événements récurrents. Dans cette étude, nous considérons deux caractéristiques importantes de ces processus à l'aide de modèles dynamiques. Ces caractéristiques sont liées aux tendances monotones et aux regroupements d'événements dans les données d'événements récurrents et sont fréquentes dans les études médicales. Nous discutons de certaines questions dans l'estimation de ces caractéristiques à l'aide de modèles dynamiques basés sur le nombre d'événements en présence d'hétérogénéité inexpliquée dans les données. De plus, nous montrons que la violation de l'hypothèse forte des temps d'écarts indépendants peut introduire un biais important dans l'estimation de ces caractéristiques avec des modèles basés sur le nombre d'événements. Pour résoudre ces problèmes, nous appliquons donc une méthode d'estimation fondée sur les copules pour les modèles de temps d'écart. Cette approche ne repose pas

Longitudinal and Survival Data Données longitudinales et de survie

strong independent gap time assumption and provides valid estimation of model parameters.

sur l'hypothèse forte d'un écart indépendant important et permet une estimation valable des paramètres du modèle.

Longitudinal and Survival Data Données longitudinales et de survie

[Monday June 12/Lundi 12 juin, 14:30-14:45]

Li-Pang Chen (University of Waterloo)

Left-Truncated and Right-Censored Survival Data Analysis with Covariate Measurement Error

Analyse de données de survie tronquées à gauche et censurées à droite avec erreur de mesure des covariables

Analysis of left-truncated and right-censored (LTRC) survival data has received extensive interest. Many inference methods have been developed for various survival models, including the Cox model and the additive hazards model. These methods, however, break down for many applications where survival data are error-contaminated. Although error-prone survival data commonly arise in practice, little work has been available in the literature for handling left-truncated and right-censored survival data with measurement error. In this paper, we explore this important problem under the Cox model. We develop valid inference methods and derive kernel smoothing estimators for the survival model parameters. We establish asymptotic results for the proposed estimators and assess the performance of our proposed methods using simulation studies. Our methods have a broad scope of application, including handling length-biased sampling data.

L'analyse des données de survie tronquées à gauche et censurées à droite (TGCD) a suscité un intérêt considérable. De nombreuses méthodes d'inférence ont été développées pour de nombreux modèles de survie, y compris le modèle de Cox et le modèle de risques additifs. Cependant, ces méthodes ne fonctionnent pas dans de nombreuses applications où les données de survie sont contaminées par des erreurs. Bien que la présence de données de survie sujettes à l'erreur soit courante en pratique, peu de travaux sont disponibles dans la littérature pour traiter les données de survie tronquées à gauche et censurées à droite avec erreur de mesure. Dans cet article, nous explorons ce problème important sous le modèle de Cox. Nous développons des méthodes d'inférence valides et dérivons des estimateurs de lissage par noyau pour les paramètres du modèle de survie. Nous établissons des résultats asymptotiques pour les estimateurs proposés et évaluons la performance des méthodes proposées à l'aide d'études de simulation. La portée d'application de nos méthodes est étendue, y compris aux données d'échantillonnage biaisées sur la longueur.

[Monday June 12/Lundi 12 juin, 14:45-15:00]

Kaida Cai (University of Calgary), **Hua Shen** (University of Calgary), **Xuwen Lu** (University of Calgary)

Group Variable Selection for Recurrent Event Data

Sélection de groupes de variables pour des données d'événements récurrents

In many scientific applications, such as biological studies, the predictors or covariates are usually high dimensional and naturally grouped. In this research, we consider the Andersen-Gill regression model for the analysis of recurrent event data with high dimensional group covariates. In order to study the effects of the covariates on the occurrence of recurrent events, a hierarchically penalized group selection method is introduced to address group selection problem under the Andersen-Gill model. We also consider an adaptive hierarchically penalized method for selecting covariates more efficiently, especially for identifying the important covariates in important groups. The asymptotic oracle properties of these methods are investigated. Our simulation studies show that the proposed methods perform well in selecting important groups and impor-

Dans plusieurs applications scientifiques, telles que les études en biologie, les prédicteurs ou les covariables sont habituellement de haute dimension et groupés naturellement. Dans cette étude, nous considérons le modèle de régression Andersen-Gill pour l'analyse de données d'événements récurrents avec covariables groupées de haute dimension. Pour étudier les effets des covariables sur la fréquence des événements récurrents, une méthode de sélection de groupes pénalisés hiérarchiquement est introduite pour s'attaquer au problème de la sélection du groupe du modèle Andersen-Gill. Nous considérons aussi une méthode pénalisée hiérarchiquement adaptative pour sélectionner les covariables plus efficacement, particulièrement pour identifier les covariables importantes dans les groupes importants. Les propriétés oracles asymptotiques de ces méthodes sont étudiées. Nos simulations démontrent que les méthodes proposées performant bien dans la sélection des groupes importants et des covariables indi-

Longitudinal and Survival Data Données longitudinales et de survie

tant individual covariates in these groups simultaneously. We illustrate these methods using some real life data sets from medicine.

viduelles importantes dans ces groupes simultanément. Nous illustrons ces méthodes à l'aide de données réelles en médecine.

Statistical Genetics Génétique statistique

Chair/Président: Lei Sun
Room/Salle: E2 165 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 13:30-13:45]

Christina M. Nieuwoudt (Simon Fraser University), **Jinko Graham** (Simon Fraser University)

SimRVPedigree: An R Package to Simulate Pedigrees Ascertained for Multiple Relatives Affected by a Rare Disease
SimRVPedigree : Un paquet R pour simuler des pedigrees établis pour plusieurs personnes apparentées touchées par une maladie rare

Family-based studies are receiving renewed attention because of their ability to identify genetic susceptibility factors associated with rare diseases. These studies have more power to detect rare variants, require smaller sample sizes, and can more accurately detect sequencing errors than case-control studies. However, garnering enough families for analysis of a rare disease could require years of effort, making these studies difficult to replicate. To address this shortcoming we have created an R package, SimRVPedigree, to randomly simulate families ascertained to contain multiple relatives affected by a rare disease. The package aims to mimic the process of family development, while allowing users to incorporate multiple facets of family ascertainment. We illustrate how approximate Bayesian computation with SimRVPedigree may be used to estimate the relative risk of disease for genetic cases in a sample of ascertained families.

Les études familiales reçoivent actuellement une attention renouvelée en raison de leur capacité à identifier les facteurs de prédisposition génétique associés aux maladies rares. Ces études ont plus de puissance pour détecter des variants rares, nécessitent des tailles d'échantillon plus petites et peuvent détecter plus précisément les erreurs de séquençage que les études cas-témoins. Cependant, le fait de réunir suffisamment de familles pour l'analyse d'une maladie rare pourrait nécessiter des années d'effort, ce qui rend ces études difficiles à reproduire. Pour pallier à ce problème, nous avons créé un paquet R, SimRVPedigree, afin de simuler aléatoirement les familles qui ont été identifiées comme ayant plusieurs personnes apparentées touchées par une maladie rare. Le paquet vise à imiter le processus de développement familial, tout en permettant aux utilisateurs d'intégrer de multiples aspects d'échantillonnage des familles. Nous démontrons comment le calcul bayésien approché avec SimRVPedigree peut être utilisé pour estimer le risque relatif de maladie pour des cas génétiques dans un échantillon de familles déterminées.

[Monday June 12/Lundi 12 juin, 13:45-14:00]

Charith Bhagya Karunaratna (Simon Fraser University), **Jinko Graham** (Simon Fraser University)

A Comparison of Association Methods for Fine-Mapping from Sequence Data in Diploid Populations using the Case-Control Study Design

Une comparaison des méthodes d'association pour la localisation détaillée à partir de données séquentielles dans des populations diploïdes utilisant le format de l'étude cas-témoin

Many methods have been proposed to detect disease association with sequence variants in candidate genomic regions. However, the literature lacks a comparison of these methods in terms of their ability to localize or fine-map the causal risk variants lying within the candidate region. We extend a previous comparison of the detection abilities of these methods to a comparison of their localization abilities. In contrast to previous work, cases and controls are sampled from a diploid (i.e., two-parent) rather than a haploid (one-parent) population. We simulated 200 sequencing datasets of

De nombreuses méthodes ont été proposées pour détecter l'association de la maladie avec des variantes de séquences dans des régions génomiques candidates. Or, on manque une comparaison de ces méthodes en termes de leur capacité à localiser les variantes de risque causales qui se trouvent dans la région candidate. Nous étendons une comparaison précédente des capacités de détection de ces méthodes à une comparaison de leurs capacités de localisation. Contrairement aux travaux antérieurs, les cas et les témoins sont pris d'une population diploïde (c'est-à-dire biparentale) plutôt que haploïde (un parent). Nous avons simulé 200 ensembles de données de séquençage d'une région génomique candidate de 2 millions

a 2-million base-pair candidate genomic region for 50 cases and 50 controls. Risk variants were in a middle subregion. We present a case study of one simulated dataset to illustrate the methods and describe simulation results to score which method best localizes the risk subregion. Our results lend support to the potential of genealogy-based methods for genetic fine-mapping of disease.

de paires de bases pour 50 cas et 50 témoins. Les variantes de risque se situaient dans une sous-région centrale. Nous présentons une étude de cas d'un ensemble de données simulées pour illustrer les méthodes et décrire les résultats de simulation afin de déterminer quelle méthode localise le mieux la sous-région du risque. Nos résultats appuient le potentiel des méthodes généalogiques pour la localisation détaillée génétique des maladies.

[Monday June 12/Lundi 12 juin, 14:00-14:15]

Lin Zhang (University of Toronto), **Lei Sun** (University of Toronto)

A Robust Allele-based Regression Framework for Testing Association and Hardy-Weinberg Equilibrium

Modèle de régression robuste basé sur les allèles pour tester l'association et l'équilibre Hardy-Weinberg

Existing allele-based association tests, examining association between genetic markers and complex human traits, are sensitive to the assumption of Hardy-Weinberg equilibrium (HWE) and limited to binary outcomes. We propose a new regression framework with individual allele as the response variable. We show that the score test statistic derived from this regression model contains a correction factor that explicitly adjusts for the departure from HWE, thus it maintains type 1 error control in the presence of HW disequilibrium (HWD). In the absence of HWD, the proposed method has comparable power as genotype-based association tests. Using this regression framework, we can then study quantitative traits, simultaneously test the HWE assumption, as well as handle more complex data, including correlated individuals from families, multiple populations and uncertain genotype observations. We support our analytical findings using evidence from both simulation and application studies.

Les tests d'association existants basé sur les allèles, qui étudient l'association entre les marqueurs génétiques et les caractéristiques humaines complexes, sont sensibles à l'hypothèse d'équilibre Hardy-Weinberg (HWE) et sont limités à des variables réponses binaires. Nous proposons un nouveau modèle de régression avec un allèle individuel comme variable de réponse. Nous démontrons que la statistique du score dérivée de ce modèle de régression contient un facteur de correction qui ajuste explicitement pour la déviation du HWE, maintenant ainsi un contrôle sur l'erreur de type 1 en présence du déséquilibre HW (HWD). En l'absence de HWD, la méthode proposée est aussi puissante que les tests d'association basés sur le génotype. En utilisant ce cadre de régression, nous pouvons étudier des traits quantitatifs, tester simultanément l'hypothèse HWE ainsi que traiter des données plus complexes telles que des individus corrélés d'une famille, des populations multiples et des observations de génotypes incertaines. Nous appuyons nos résultats analytiques avec des preuves provenant de simulations et d'études d'applications.

[Monday June 12/Lundi 12 juin, 14:15-14:30]

Elham Khodayari Moez (University of Alberta), **Irina Dinu** (School of Public Health, University of Alberta)

Gene Set Enrichment Analysis in Longitudinal Studies

Analyse de l'enrichissement des ensembles de gènes dans les études longitudinales

Gene-set enrichment methods aim to discover gene-sets associated with phenotypes. Although there has been progress in developing proper methods for analysis of high dimensional microarray datasets in cross-sectional studies, methods for dealing with longitudinal phenotypes are still limited. A two-step self-contained gene-set analysis method is developed to handle multiple longitudinal outcomes. Analysis of within-subject variation in the first step is followed by examining the between-subject vari-

Les méthodes d'enrichissement des gènes visent à découvrir des ensembles de gènes associés à des phénotypes. Bien qu'il y ait eu des progrès dans le développement de méthodes appropriées pour l'analyse des jeux de données de biopuces de grande dimension dans les études transversales, les méthodes pour traiter les phénotypes longitudinaux sont encore limitées. Une méthode d'analyse d'ensembles de gènes en deux étapes est développée pour analyser des variables réponse longitudinales multiples. L'analyse de la variation intra-sujet à la première étape est suivie par l'examen de la variation

ation utilizing Linear Combination Test (LCT) in the second step. This method is also applicable in analysis of time-course microarray data. The performance of the method is evaluated in a simulation study. The proposed method is very efficient in controlling Type I error and works well with small sample size, large number of genes and missing data.

entre les sujets en utilisant le test de combinaison linéaire (TCL) à la seconde étape. Cette méthode est également applicable dans l'analyse des données temporelles de biopuces. La performance de la méthode est évaluée dans une étude de simulation. La méthode proposée est très efficace pour contrôler l'erreur de type I et fonctionne bien avec une petite taille d'échantillon, un grand nombre de gènes et des données manquantes.

[Monday June 12/Lundi 12 juin, 14:30-14:45]

Agnes Nessie Amu (University of Manitoba), **Elif Fidan Acar** (University of Manitoba)

Factor Copula Analysis for Multivariate Ordinal Data

Analyse de copule avec facteurs pour données ordinales multivariées

In genetic studies of many neurological and psychological disorders, quantitative phenotypic traits are latent and inferred from the ordinal assessment scores in diagnostic questionnaires. In such cases, an accurate representation of the dependence structure in multivariate ordinal data is essential for correct identification of these latent traits and their use in genetic mapping. This work provides a detailed comparison of dependence measures for multivariate ordinal data and investigates the robustness of polychoric correlation estimation under settings with asymmetric dependence patterns and varying degrees of skewness in marginal distributions. An alternative strategy to quantify the latent traits is proposed using factor copula models. We compare the performances of these approaches in factor scores regression for testing genetic associations.

Dans les études génétiques sur plusieurs troubles psychologiques et neurologiques, les caractéristiques phénotypiques quantitatives sont latentes et inférées à partir des scores d'évaluation ordinaux des questionnaires de diagnostic. En pareil cas, une présentation précise de la structure de dépendance dans les données multivariées ordinales est essentielle pour une identification juste de ces caractéristiques latentes et de leur utilité dans la cartographie génétique. Cette étude présente une comparaison détaillée des mesures de dépendance pour des données ordinales multivariées et examine la robustesse de l'estimation de corrélation polychorique dans une structure de dépendance asymétrique et avec différents degrés d'asymétrie dans les lois marginales. Une stratégie alternative pour quantifier les caractéristiques latentes utilisant les modèles de copule avec facteurs est proposée. Nous comparons les performances de ces approches dans la régression des scores des facteurs pour tester les associations génétiques.

[Monday June 12/Lundi 12 juin, 14:45-15:00]

Maxime Turgeon (McGill University), **Stepan Grinek** (BC Cancer Agency), **Celia M.T. Greenwood** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University), **Auréli Labbe** (HEC Montréal)

Reduced-Rank Singular Value Decomposition for Dimension Reduction with High-Dimensional Data

Décomposition en valeurs singulières à rang réduit appliquée aux méthodes de réduction dimensionnelle pour l'étude de données de grande dimension

Recent technical advances in genomics have led to an abundance of high-dimensional and correlated data. Dimension reduction methods typically rely on matrix decompositions (e.g. SVD and EVD) to compute the quantities needed for further analysis. However, in a high-dimensional setting, these decompositions must be adapted to cope with the singularity of the matrices involved. We illustrate how this can be done in the context of a particular dimension reduction method, Principal Component of Explained Variance (PCEV). PCEV seeks a

Les avancées techniques récentes en génomique ont contribué à une abondance de données de grande dimension et corrélées. Les méthodes de réduction dimensionnelle utilisent habituellement des méthodes de décomposition matricielle (e.g. SVD et EVD) pour calculer les quantités nécessaires aux analyses de ces données. Or, dans un contexte de grande dimension, ces décompositions doivent être adaptées au cas où certaines matrices sont singulières. Nous illustrons comment faire dans le contexte d'une méthode de réduction dimensionnelle, Composante Principale de la Variance Expliquée (PCEV). Cette méthode cherche la combinaison linéaire des

linear combination of outcomes by maximising the proportion of variance explained by the covariates of interest. Using random matrix theory, we propose a heuristic that provides a fast way to compute valid p-values to test the significance of the decomposition. We compare the power of this approach with that of other common approaches to high-dimensional data. Finally, we illustrate our method using methylation data collected on a small number of individuals.

variables-réponses qui maximise la proportion de variance expliquée par les covariables. En se basant sur la théorie des matrices aléatoires, nous proposons une heuristique qui permet de calculer rapidement des valeurs-p valides pour tester la signification de la décomposition. Nous comparons la puissance de notre approche à celle d'autres méthodes populaires pour l'analyse de données de grande dimension. Enfin, nous illustrons notre approche à l'aide de données de méthylation mesurées chez un nombre restreint de patients.

Insurance Risk and Ruin Measures

Mesures de risque et de ruine en assurance

Chair/Président: Jose Garrido
Organizer/Responsable: Jose Garrido
Room/Salle: E2 155 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-16:00]

Ilie Radu Mitric (Université Laval), **Julien Trufin** (Université Libre de Bruxelles), **Amine Mohamed Lkabous** (Université du Québec à Montréal)

Properties of Risk Measures Inspired from the Ruin Probability, the Deficit at Ruin, and the Time of Ruin

L'étude des propriétés de quelques mesures de risque dérivées de la probabilité de la ruine, le déficit à la ruine et le temps de ruine

We study a risk measure derived from ruin theory defined as the amount of capital needed to cope in expectation with the first occurrence of a ruin event. Specifically, within the compound Poisson model, we investigate some properties of this risk measure with respect to the stochastic ordering of claim severities. Particular situations where combining risks yield diversification benefits are identified. Closed form expressions and upper bounds are also provided for certain claim severities. Further extensions are explored.

On étudie une mesure de risque dérivée de la probabilité de la ruine qui permet de déterminer le capital nécessaire pour faire face au déficit moyen à la ruine. Pour le modèle classique de Poisson, on étudie quelques propriétés de cette mesure de risque par rapport à des ordres stochastiques satisfaits par les montants de sinistres. Quelques situations particulières sont identifiées, où la fusion des risques ne dépasse pas la somme des risques individuels. Des expressions explicites et des bornes sont trouvées pour certains cas particuliers de sinistres. Des extensions pour un modèle avec diffusion et des fonctions particulières de type Gerber-Shiu sont explorées.

[Monday June 12/Lundi 12 juin, 16:00-16:30]

Georgios Pitselis (University of Piraeus)

Risk Measures Related to both the Information of an Individual Risk and Industry Risk

Les mesures du risque relatives à l'information en matière de risque individuel et de risque lié à l'industrie

In this paper we show how risk measures can be embedded within the framework of credibility theory. More specifically, we introduce a new type of risk measures, the credible risk measures, in order to capture the risk of an individual contract (or financial portfolio) as well as the industry risk consisting of several similar, but not identical, contracts (or financial portfolios). We consider the classical case, as well as the regression case. Examples are given based on the Fama/French data.

Cet article fait état du mode d'intégration des mesures du risque dans le cadre de la théorie de la crédibilité. Nous présentons plus précisément un nouveau type de mesures du risque, les mesures du risque plausible, afin de saisir le risque lié au contrat individuel (ou portefeuille financier) et le risque lié à l'industrie, constitué de plusieurs contrats (ou portefeuilles financiers) semblables sans être identiques. Nous étudierons le cas classique et le cas de la régression. Les exemples sont fondés sur les facteurs de Fama et French.

[Monday June 12/Lundi 12 juin, 16:30-17:00]

Jose Garrido (Concordia University), **Alejandro Balbás** (University Carlos III de Madrid, Spain), **Ramin Okhrati** (United of Southampton, United Kingdom)

Good Deal Indices in Asset Pricing: Actuarial and Financial Implications

Le concept d'aubaine en évaluation d'actifs : implications actuarielles et financières

We integrate into a single optimization problem a risk measure and, either arbitrage free real market quotations, or financial pricing rules generated by an arbitrage free stochastic pricing model. We call

Nous intégrons en un seul problème d'optimisation une mesure de risque et, soit des cotes de marché ou des règles de tarification générées par un modèle de tarification stochastique, sans arbitrage. Nous appelons aubaine une suite de stratégies

Insurance Risk and Ruin Measures

Mesures de risque et de ruine en assurance

a good deal (GD) a sequence of investment strategies such that the couple (expected-return, risk) diverges to $(+\infty; -\infty)$. The existence of such a sequence is equivalent to the existence of an alternative sequence such that the couple (risk, price) goes to $(-\infty; -\infty)$. Moreover, by appropriately adding the riskless asset, every GD may generate a new one composed only of strategies priced at 1. We show how GDs exist in practice, and study how to measure a good deal size. We also provide the minimum relative (per dollar) price modification that prevents the existence of GDs. This is a crucial tool to detect over/under-priced securities or marketed claims. Many classical actuarial and financial optimization problems can generate wrong solutions if the used market quotations or stochastic pricing models do not prevent the existence of GDs. We illustrate this and show how GD indices help overcome this caveat. Numerical illustrations are given.

d'investissement telle que le couple (retour-espéré, risque) diverge à $(+\infty; -\infty)$. L'existence d'une telle suite est équivalente à celle d'une suite alternative telle que le couple (risque, prix) diverge à $(-\infty; -\infty)$. De plus, en rajoutant l'actif sans risque approprié, chaque aubaine peut en générer une nouvelle, composée uniquement de stratégies au prix unitaire. Nous montrons comment ces aubaines existent en pratique et étudions la mesure de la taille d'une aubaine. Nous calculons aussi l'ajustement tarifaire relatif (par dollar) minimum qui permet de prévenir ces aubaines. Ceci s'avère être essentiel pour la détection d'actifs ou produits de marché aux prix sur/sous-évalués. Plusieurs problèmes classiques d'optimisation, actuariels et financiers, mènent à des solutions erronées si les cotes de marché ou les modèles de tarification stochastique utilisés permettent ces aubaines. Ceci sera présenté ainsi que comment des indices d'aubaines aident à surmonter ce problème. Des exemples numériques serviront d'illustrations.

Using Randomization and Simulation in the Teaching of Statistics

Enseigner la statistique par la randomisation et la simulation

Chair/Président: Jim Stallard
Organizer/Responsable: Jim Stallard
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-16:15]

Nathan Tintle (Dordt College)

Teaching the Statistical Investigation Process with Simulation-Based Inference to Improve Statistical Thinking

Enseignement du processus de recherche statistique avec inférence fondée sur la simulation en vue d'améliorer la pensée statistique

The statistics education community is increasingly focusing on the use of simulation-based methods, including bootstrapping and permutation tests, to illustrate core concepts of statistical inference within the context of the overall research process. This new focus presents an opportunity to address documented shortcomings in introductory and intermediate level statistics courses. In this talk I will (1) discuss the motivation and rationale behind the simulation-based approach, (2) share some concrete examples of how the approach works and can be integrated into existing courses, (3) present research evidence of its effectiveness at impacting students conceptual understanding and attitudes post-course and in the months following the courses completion and, (4) share a wealth of instructional resources available to support instructors trying out and using these approaches.

La communauté des enseignants en statistique accorde un intérêt croissant à l'utilisation de méthodes de simulation, y compris le bootstrap et les tests de permutation, pour illustrer les principaux concepts d'inférence statistique dans le contexte du processus global de recherche. Ce nouveau champ d'intérêt offre la possibilité d'étudier les faiblesses documentées des cours de statistique de base et intermédiaires. Cette allocution porte sur : (1) la motivation et le raisonnement qui sous-tendent les méthodes de simulation, (2) des exemples concrets du fonctionnement de ces méthodes et de la possibilité de les intégrer dans les cours déjà offerts, (3) des données de recherche quant à son efficacité à influencer sur la compréhension conceptuelle des étudiants et sur leur attitude à la fin du cours et dans les mois subséquents, et, (4) une mine de ressources instructives pour aider les enseignants à essayer et à utiliser ces méthodes.

[Monday June 12/Lundi 12 juin, 16:15-17:00]

John Sheriff (University of Lethbridge)

Incorporating Simulation-Based Inference in an Introductory Statistics Course

Intégration de l'inférence par simulation dans un cours de base en statistique

The use of simulation-based methods in the teaching of statistical inference has received considerable attention in recent years. The objective of this presentation is to share my experience incorporating simulation-based methods of inference in the introductory statistics course at the University of Lethbridge. This change was made for the first time in the 2016/17 academic year. I will discuss motivations for the change, and some of the consequences that arose from the change. I will address available resources, highlight student performance and feedback, as it pertains to the change, and discuss lessons learned for next time.

Ces dernières années, l'utilisation de méthodes de simulation dans l'enseignement de l'inférence statistique a fait l'objet d'un intérêt soutenu. Ma présentation porte sur mon expérience à intégrer des méthodes d'inférence fondée sur la simulation dans le cours de base en statistique qui se donne à la University of Lethbridge. Ce changement a été effectué pour la première fois pendant l'année universitaire 2016-2017. Je traiterai des motifs de ce changement et des conséquences qu'il a eu. En plus de présenter les ressources disponibles ainsi que les résultats des étudiants et leurs commentaires par rapport à ce changement, je toucherai un mot des leçons à retenir pour la prochaine fois.

Design and Analysis of Validation Studies for Health Outcomes in Electronic Health Databases

Conception et analyse d'études de validation pour les résultats cliniques dans les bases de données électroniques en santé

Chair/Président: Lisa M. Lix
Organizer/Responsable: Lisa M. Lix
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-16:00]

Elham Rahme (McGill University), **Jiayi Ni** (Research Institute of the McGill University Health Center), **Kaberi Dasgupta** (McGill University), **Denis Talbot** (Université Laval), **Geneviève Lefebvre** (Université du Québec à Montréal), **Lisa M. Lix** (University of Manitoba), **Lucie Blais** (Université de Montréal)

Comparing Bayesian External and Internal Validation Methods in Correcting Outcome Misclassification Bias in Logistic Regression

Comparaison de méthodes internes et externes de validation bayésienne pour corriger le biais de classification de l'issue dans la régression logistique

Misclassification is frequent in administrative data. Using simulations, we compared two Bayesian methods in reducing outcome misclassification bias in logistic regression. Sensitivity and specificity priors were based on external information in method 1 and on internal complementary data in method 2. Bias, 95% credible intervals (CI) coverage and mean squared error (MSE) were used to assess the performance of these methods. Both methods yielded estimates with less bias (-0.113 - 0.047) than the naïve analysis (-0.452 - 0.258) in all scenarios explored. Method1 performed well with reasonable sensitivity and specificity priors; performance decreased with decreased mean of prior sensitivity. CI coverage was high for both methods (95-100%). The Bayesian external validation method is practical and useful to reduce outcome misclassification bias in logistic regression. While the internal method may provide better adjustment, it has the disadvantage of requiring additional individual data.

Les erreurs de classification sont fréquentes dans les données administratives. À l'aide de simulations, nous comparons deux méthodes bayésiennes pour réduire les biais de classification de l'issue dans la régression logistique. La sensibilité et la spécificité a priori étaient fondées sur de l'information externe dans la méthode 1 et sur des données complémentaires internes dans la méthode 2. Le biais, la couverture des intervalles de crédibilité à 95% (IC) et l'erreur quadratique moyenne (EQM) ont été utilisés pour évaluer la performance de ces méthodes. Dans tous les scénarios étudiés, les deux méthodes ont produit des estimations avec moins de biais (-0.113 - 0.047) que les analyses naïves (-0.452 - 0.258). La méthode 1 a bien performé avec une sensibilité et une spécificité a priori raisonnables; la performance diminuait avec une moyenne de sensibilité réduite. La couverture des ICs était élevée pour les deux méthodes (95-100%). La méthode externe de validation bayésienne est pratique et utile pour réduire le biais de classification de l'issue dans la régression logistique. Bien que la méthode interne permette un meilleur ajustement, elle a le désavantage de nécessiter des données individuelles additionnelles.

[Monday June 12/Lundi 12 juin, 16:00-16:30]

Eric Benchimol (University of Ottawa)

Validation of Codes and Algorithms are an Essential Part of Research Using Routinely Collected Health Data

La validation des codes et des algorithmes est un élément essentiel de la recherche utilisant des données sur la santé recueillies systématiquement

Routinely collected health data (RCD) are defined as data collected without specific a priori research questions developed prior to utilization for research. In Canada, the most widely used population-based RCD are health administrative data, other examples include disease registries, public health reporting, and electronic health records. With the in-

Les données sur la santé recueillies systématiquement (RCD) sont définies comme étant des données recueillies sans questions de recherches préalables spécifiques développées avant l'utilisation pour la recherche. Au Canada, les RCD basées sur la population les plus largement utilisées sont les données administratives sur la santé; d'autres exemples comprennent les registres de maladies, les rapports sur la santé publique

Design and Analysis of Validation Studies for Health Outcomes in Electronic Health Databases

Conception et analyse d'études de validation pour les résultats cliniques dans les bases de données électroniques en santé

creased use of RCD, there is increased awareness of methodological concerns. The REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement was developed to address these limitations and to help meet the requirement of clear reporting of research using RCD. RECORD requires the description of validation studies of the codes/algorithms used to select the population, and to describe outcomes, confounders, or effect modifiers. I will emphasize the importance of algorithm and code validation prior to embarking on research using RCD, and present validation methods from the literature. We will encourage discussion on new and innovative methods to validate administrative data for use in health research.

et les dossiers médicaux électroniques. Avec l'utilisation accrue des RCD vient une sensibilisation accrue aux questions d'ordre méthodologique. La déclaration REporting of studies Conducted using Observational Routinely-collected Data (RECORD) a été développée pour adresser ces limites et pour faciliter l'atteinte des exigences de clarté des publications de la recherche utilisant les RCD. RECORD exige la description des études de validation des codes/algorithmes utilisés pour la sélection de la population ainsi que la description des résultats, des facteurs de confusion ou de modification d'effet. Je mettrai l'accent sur l'importance de la validation des algorithmes et des codes avant d'amorcer une recherche qui utilise des RCD et je présenterai des méthodes de validation provenant de la littérature. Nous encouragerons la discussion sur de nouvelles méthodes novatrices pour valider les données administratives dans la recherche en santé.

[Monday June 12/Lundi 12 juin, 16:30-17:00]

Tyler Williamson (University of Calgary), **Brendan Cord Lethebe** (University of Calgary), **Colin Weaver** (University of Calgary), **Tolulope T. Sajobi** (University of Calgary), **Hude Quan** (University of Calgary)

Feature Selection Methods for the Development of Case Definitions for Common Chronic Conditions in Primary Care Electronic Medical Record Data

Les méthodes de sélection de caractéristiques pour l'élaboration des définitions cliniques de cas des maladies chroniques courantes utilisant les dossiers médicaux électroniques de soins primaires

Chronic disease surveillance information is dependent on the quality of the EMR data, and the quality of the case identification algorithms. Data were obtained from the Canadian Primary Care Sentinel Surveillance Network. A chart review was conducted for the presence of 8 chronic conditions in 1920 primary care patients. The results of this review will be used as training data for classification models. Features will be selected from billing codes, medication prescriptions, laboratory values, encounter diagnoses and health-problem lists. CART, C5.0, and CHAID decision tree algorithms will be compared with LASSO and forward stepwise logistic regression in terms of case definition development. A bootstrap validation technique will be used to select optimal complexity parameter value. Validity measures will be determined using 10-fold cross validation. RESULTS: It has been shown that these methods are comparable with committee created case definitions in terms of predictive accuracy.

La surveillance des maladies chroniques dépend de la qualité des données des dossiers médicaux électroniques et de la qualité des algorithmes qui identifient les cas. Les données ont été obtenues du Réseau canadien de surveillance sentinelle en soins primaires. Un examen des dossiers médicaux a été effectué de 1920 patients de soins primaires pour déterminer la présence de 8 maladies chroniques. Les résultats de cet examen ont été utilisés pour l'entraînement des modèles de classification. Variables ont été : les codes de facturation, les ordonnances, les résultats des tests de laboratoire, et les diagnostics. Plusieurs algorithmes ont été comparés : CART, C5.0, et CHAID, ainsi que le LASSO et la régression logistique ascendante. La performance des modèles a été déterminée par une validation croisée de 10 échantillons. RÉSULTATS : La précision de ces algorithmes est comparable aux définitions cliniques de cas créées par les comités.

Design and Analysis Issues in Cluster Randomized Controlled Trials Problèmes de conception et d'analyse dans les essais randomisés contrôlés par grappes

Chair/Président: Lehana Thabane
Organizer/Responsable: Jinhui Ma
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-16:00]

Robert W. Platt (McGill University)

Observational Analyses of Cluster-Randomized Trials

Analyses d'observation d'essais randomisés par grappes

The cluster-randomized trial is an important tool for the study of interventions at a group level, and for interventions in which contamination may be a concern. Hospital-level program interventions are particularly appropriate for this design. Cluster-randomized trials can, however, be very expensive and time-consuming. Because these trials often involve collection of substantial data, observational analyses of follow-up data can provide useful information. I will discuss some analytic challenges related to observational analyses of cluster trials. Such challenges include methods to account for clustering, in particular how clustering affects missing data and measurement error, and appropriate statement of the research question and identification of correct comparisons to address the question. I will illustrate these problems through analyses of the PROBIT study of a breastfeeding promotion intervention.

L'essai randomisé par grappes est un outil important pour l'étude des interventions à un niveau de groupe, et pour les interventions dans laquelle la contamination peut être problématique. Les interventions dans le cadre d'un programme hospitalier sont particulièrement appropriées pour ce plan. Les essais en grappes aléatoires peuvent cependant être très chers et prendre du temps. Parce que ces essais impliquent souvent la collecte de données d'une grande richesse, des analyses de données de suivi peuvent fournir des renseignements utiles. Je discuterai de certains défis liés aux analyses observationnelles des essais en grappes. Ces défis comprennent les méthodes visant à prendre en compte la mise en grappe, notamment la façon dont la mise en grappe a une incidence sur les données manquantes et sur l'erreur de mesure, ainsi que sur la détermination de comparaisons adéquates pour aborder ces questions. J'illustrerai ces défis par des analyses d'intervention visant à promouvoir l'allaitement au sein (Promotion of Breastfeeding Intervention Trial PROBIT).

[Monday June 12/Lundi 12 juin, 16:00-16:30]

Eleanor M. Pullenayegum (The Hospital for Sick Children)

Analysis of Cluster-Randomized Trials with a Longitudinal Outcome Subject to Irregular Observation

Analyse d'essais randomisés par grappes avec variable dépendante longitudinale soumise à une observation irrégulière

Randomized trials may feature an outcome measured longitudinally as part of usual care. This reduces trial costs, as no special study visits are required, while allowing for an exploration of longer term outcomes. However, since the timing of the visits is not specified by protocol, this can lead to outcome observation times varying among subjects, and potentially related to the outcomes themselves. This may result in bias unless appropriate analytic methods are used. This talk will explore how two popular classes of methods (inverse-intensity weighting and semi-parametric joint models) can be

Les essais randomisés peuvent caractériser une variable mesurée de façon longitudinale dans le cadre de soins courants. Ceci réduit les coûts des essais cliniques, puisqu'aucune visite spéciale dans le cadre de l'étude n'est exigée, tout en permettant l'exploration de variables à plus long terme. Cependant, le fait que le moment des visites ne soit pas précisé par un protocole peut engendrer des périodes d'observation de variables variant parmi les sujets, et potentiellement liées aux variables elles-mêmes. Cela peut engendrer un biais, sauf si des méthodes d'analyse appropriées sont utilisées. Cet exposé explorera la façon dont deux catégories de méthodes d'analyse (modèles de pondération de l'intensité inverse et modèles

Design and Analysis Issues in Cluster Randomized Controlled Trials Problèmes de conception et d'analyse dans les essais randomisés contrôlés par grappes

extended to account for within-cluster correlation.

conjointes semi-paramétriques) peuvent être généralisées pour prendre en compte une corrélation au sein de la grappe.

Design and Analysis Issues in Cluster Randomized Controlled Trials

Problèmes de conception et d'analyse dans les essais randomisés contrôlés par grappes

[Monday June 12/Lundi 12 juin, 16:30-17:00]

Jinhui Ma (Children's Hospital of Eastern Ontario Research Institute), **Monica Taljaard** (Ottawa Hospital Research Institute), **Lisa Dolovich** (McMaster University), **Janusz Kaczorowski** (University of Montreal), **Larry Chambers** (University of Ottawa), **Lehana Thabane** (McMaster University)

Strategies to Handle Missing Binary Outcomes in Cluster Randomized Trials

Stratégies pour traiter les variables indépendantes binaires manquantes dans des essais randomisés par grappes

Missing data is a common issue in cluster randomized trials (CRTs). Choosing the most appropriate strategy to handle a missing binary outcome in CRTs may be very challenging due to great variability in the design and implementation of such trials. According to published literature, the optimal methods for handling missing outcome data in CRTs are infrequently used in practice. In my presentation, I will provide a brief overview of the methods available in the literature for handling missing binary outcomes in CRTs, and illustrate how to choose and implement the most appropriate missing data strategy based on the design characteristics of the CRTs using both a real CRT and simulated scenarios as examples. I will also address the issues using the population-averaged and cluster-specific methods in analyzing the binary data in CRTs when missing data present.

Les données manquantes sont un problème courant dans les essais randomisés par grappes. Le choix de la stratégie la plus adaptée pour traiter la variable dépendante binaire manquante dans les essais randomisés par grappes pose un défi de taille en raison de la grande variabilité dans la conception et la mise en œuvre de ces essais. D'après la littérature existante, les méthodes optimales de traitement des données manquantes ne sont pas utilisées fréquemment en pratique dans le cadre d'essais randomisés par grappes. Dans mon exposé, je donnerai un aperçu des méthodes existantes dans la littérature visant à traiter les variables dépendantes binaires manquantes dans les essais randomisés par grappes. J'illustrerai à l'aide de deux exemples, l'un réel et l'autre simulé, la façon de choisir et de mettre en œuvre la stratégie la plus adaptée en fonction des caractéristiques de conception des essais randomisés par grappes. J'examinerai aussi cette problématique à l'aide de méthodes relatives à la moyenne d'une population et spécifiques à une grappe donnée.

Double Robustness and Dynamic Treatment Regimes

La double robustesse et les régimes de traitement dynamiques

Chair/Président: Michael Wallace
Organizer/Responsable: Michael Wallace
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-16:00]

David A. Stephens (McGill University), **Michael Wallace** (University of Waterloo), **Erica E.M. Moodie** (McGill University)

G-estimation and Model Selection for Dynamic Treatment Regimes

G-estimation et sélection de modèles pour régimes de traitement dynamiques

Dynamic treatment regimes (DTRs) formalize personalized medicine by tailoring treatment decisions to individual patient characteristics. G-estimation for DTR identification targets the parameters of a structural nested mean model (the blip function), from which the optimal DTR is derived. Despite much work focusing on deriving such estimation methods, there has been little focus on extending G-estimation to the case of non-additive effects, non-continuous outcomes or on model selection. We demonstrate how G-estimation can be more widely applied through the use of iteratively-reweighted least squares procedures, and illustrate this for log-linear models. We then derive a quasi-likelihood function for G-estimation within the DTR framework, and show how it can be used to form an information criterion for blip model selection. These developments are demonstrated through simulation, as well as in application to data from the Sequenced Treatment Alternatives to Relieve Depression study.

Les régimes de traitement dynamiques (RTD) formalisent la médecine personnalisée en adaptant les décisions de traitement aux caractéristiques individuelles du patient. La g-estimation pour identification d'un RTD cible les paramètres d'un modèle moyen structurel emboîté (fonction Blip), pour en tirer ensuite un RTD optimal. Malgré de nombreux travaux sur la dérivation de ces méthodes d'estimation, on a peu mis l'accent sur une extension de la g-estimation au cas des effets non additifs, des issues non continus ou de la sélection de modèles. Nous montrons comment la g-estimation peut être étendue par l'utilisation de procédures des moindres carrés à repondération itérative, et en illustrons l'utilisation pour les modèles log-linéaires. Nous dérivons ensuite une fonction de quasi-vraisemblance pour la g-estimation dans le cadre des RTD et montrons comment elle peut être utilisée pour créer un critère d'information pour la sélection de modèle Blip. Nous illustrons ces développements à l'aide de simulations et par l'application à des données tirées de l'étude Sequenced Treatment Alternatives to Relieve Depression.

[Monday June 12/Lundi 12 juin, 16:00-16:30]

Olli Saarela (University of Toronto)

On the Bayesian Semi-Parametric Double Robustness Property

Double robustesse semi-paramétrique bayésienne

We have studied Bayesian doubly robust causal inferences under a framework where the causal contrast is specified in terms of a prediction problem under a hypothetical randomized experiment, and inverse probability of treatment (IPT) weights are introduced as importance sampling weights in Monte Carlo integration. A posterior distribution is produced by sampling from a non-parametric model, with parametric working models introduced for smoothing purposes. Marginal structural models can be estimated similarly through maximiza-

Nous étudions les inférences causales doubles robustes bayésiennes dans un cadre où le contraste causal est spécifié en termes d'un problème prévisionnel dans une expérience aléatoire hypothétique, et où les pondérations inverses sur la probabilité d'être traité (IPT) sont introduites comme poids d'échantillonnage d'importance dans l'intégration de Monte-Carlo. Une loi a posteriori est produite par échantillonnage d'un modèle non paramétrique, en introduisant des modèles fonctionnels paramétriques à des fins de lissage. De même, on peut estimer des modèles structurels marginaux par maximisation d'une fonction d'utilité paramé-

Double Robustness and Dynamic Treatment Regimes

La double robustesse et les régimes de traitement dynamiques

tion of a parametric utility function. However, it is still unclear how certain other causal estimands such as treatment effects among the treated and structural mean models, and estimation methods such as propensity score regression and g-estimation, could be incorporated into the Bayesian semi-parametric framework. We will review the ideas behind Bayesian doubly robust estimation using IPT weights, and consider extending these to the aforementioned directions.

[Monday June 12/Lundi 12 juin, 16:30-17:00]

Peisong Han (University of Waterloo)

Achieving Higher Efficiency and Robustness in Longitudinal Studies with Dropout

Amélioration de l'efficacité et de la robustesse dans les études longitudinales avec abandons

Intrinsic efficiency and multiple robustness are desirable properties in missing data analysis. We establish both for estimating the mean of a response at the end of a longitudinal study with drop-out. The idea is to calibrate the estimated missingness probability at each visit using data from past visits. We consider one working model for the missingness probability and multiple working models for the data distribution. Intrinsic efficiency guarantees that, when the missingness probability is correctly modelled, the multiple data distribution models, combined with data prior to the end of the study, are optimally accommodated to maximize efficiency. Multiple robustness ensures estimation consistency if the missingness probability model is misspecified but one data distribution model is correct. Our proposed estimators are all convex combinations of the observed responses, and thus always fall within the parameter space.

trique. Cependant, on ignore encore comment intégrer dans le cadre semi-paramétrique bayésien d'autres paramètres causals comme les effets du traitement sur les sujets traités et les modèles moyens structurels, ainsi que les méthodes d'estimation comme la régression par score de propension et la g-estimation. Nous examinons les idées qui sous-tendent l'estimation double robuste bayésienne à l'aide des pondérations IPT et en étudions l'extension éventuelle dans les directions susmentionnées.

L'efficacité intrinsèque et la multiple robustesse sont des propriétés avantageuses pour l'analyse de données manquantes. Nous établissons ces deux caractéristiques pour l'estimation de la moyenne d'une réponse à la fin d'une étude longitudinale avec abandons. L'idée est de calibrer la probabilité estimée de données manquantes à chaque visite à l'aide des données des visites précédentes. Nous considérons un modèle de travail pour la probabilité de données manquantes et plusieurs modèles de travail pour la distribution des données. L'efficacité intrinsèque garantit, si la probabilité de données manquantes est bien modélisée, que les multiples modèles de distribution des données, conjugués aux données collectées avant la fin de l'étude, sont intégrés au mieux pour maximiser l'efficacité. La multiple robustesse garantit une convergence de l'estimation si le modèle de probabilité de données manquantes contient des erreurs de spécification, mais que l'un des modèles de distribution des données est correct. Les estimateurs que nous proposons sont des combinaisons convexes des réponses observées, et sont donc toujours à l'intérieur de l'espace-paramètres.

Self-Normalized Processes: Dedicated to Miklos Csorgo on the Occasion of his 85th Birthday

Processus auto-normalisés : en l'honneur du 85e anniversaire de Miklos Csorgo

Chair/Président: Don McLeish
Organizer/Responsable: Barbara Szyszkowicz
Room/Salle: E2 150 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-16:00]

Miklos Csorgo (Carleton University)

Self-Normalized Weak and Strong Functional Limit Laws in the Domain of Attraction of Stable Laws

Lois limites fonctionnelles faibles et fortes autonormalisées dans les domaines d'attraction des lois stables

We will present a glimpse of Donsker-type weak invariance principles that have been established in these years for self-normalized partial sums processes in the domain of attraction of a stable law with index number in the interval $(0, 2]$. Special attention will be given to the case of index number 2, i.e., to DAN (domain of attraction of the normal law), and to applications to change-point analysis in this domain, with summands possibly having an infinite variance. Strassen-type strong functional laws and invariance principles in DAN will also be explored in the latter case.

Nous présenterons un aperçu des principes d'invariance faible de type Donsker qui ont été établis au cours de ces dernières années pour les processus de sommes partielles auto-normalisés dans le domaine d'attraction d'une loi stable avec un indice dans l'intervalle $(0, 2]$. Nous nous intéresserons plus particulièrement à l'indice 2, à savoir au domaine d'attraction de la loi normale et aux applications pour l'analyse des points de changement dans ce domaine, ainsi qu'aux opérandes ayant possiblement une variance infinie. Dans la dernière étude de cas, nous explorerons les lois fonctionnelles fortes du type Strassen et les principes d'invariance dans le domaine d'attraction de la loi normale.

[Monday June 12/Lundi 12 juin, 16:00-16:30]

Yuliya V. Martynsyuk (University of Manitoba)

Functional Central Limit Theorems for Self-Normalized Processes with Statistical Applications

Théorèmes centraux limites fonctionnels pour des processus auto-normalisés, et applications statistiques

We will revisit Donsker-type functional central limit theorems for self-normalized processes that are based on independent random variables. We will also present some statistical applications of these theorems, including the ones for linear errors-in-variables and regression models, and for construction of asymptotic confidence intervals for a population mean. Some modifications and generalizations of these theorems will be outlined.

Nous revisiterons les théorèmes centraux limites fonctionnels de type Donsker pour les processus auto-normalisés qui se fondent sur des variables aléatoires indépendantes. Nous présenterons aussi des applications statistiques de ces théorèmes, notamment pour les erreurs dans les variables et les modèles de régression linéaires, ainsi que pour les intervalles de confiance asymptotiques pour la moyenne d'une population. Nous décrirons les modifications qui ont été apportées à ces théorèmes et les généralisations de ceux-ci.

[Monday June 12/Lundi 12 juin, 16:30-17:00]

Masoud Nasari (Carleton University)

Self-Normalized Partial Sums of Randomly Weighted Data

Sommes partielles auto-normalisées de données pondérées au hasard

In this talk we address applications of self-normalized partial sums of randomly weighted data (i) in increasing the accuracy of the CLT based confidence intervals and (ii) in drawing inference based on subsamples drawn from large data sets.

Dans cet exposé, nous examinons les applications des sommes partielles auto-normalisées de données pondérées au hasard i) en augmentant la précision des intervalles de confiance basés sur les théorèmes centraux limites; ii) en faisant des inférences en fonction de sous-échantillons prélevés de grands jeux de données.

Data Science in Biomedical Studies
Science des données dans les études biomédicales

Chair/Président: X. Joan Hu and Lisa Le
Organizer/Responsable: X. Joan Hu and Lisa Le
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-16:00]

Juxin Liu (University of Saskatchewan), **Paul Gustafson** (University of British Columbia), **Dezheng Huo** (University of Chicago)

Discordance in Hormone Receptor Status Between Two Primary Breast Cancers: The Impact of Misclassification
Discordance dans l'état des récepteurs hormonaux entre deux cancers primitifs du sein : l'incidence des erreurs de classification

A Bayesian method is proposed to address misclassification errors in both independent and dependent variables. Our work is motivated by a study of women who have experienced new breast cancers on two separate occasions. Hormone receptors (HRs) are important in breast cancer biology, and it is well recognized that the measurement of HR status is subject to errors. This discordance in HR status for two primary breast cancers is of concern and might be an important reason for treatment failure. We consider a logistic regression model for the association between the HR status of the two cancers and introduce the misclassification parameters accounting for the misclassification in HR status. The prior distribution for sensitivity and specificity is based on how HR status is assessed in laboratory procedures. B-spline terms are used to account for the nonlinear effect of an error-free covariate. Our findings indicate that the true concordance rate of HR status between two primary cancers is greater than the observed value.

Nous proposons une méthode bayésienne permettant d'adresser le problème des erreurs de classification à la fois pour des variables dépendantes et indépendantes. Notre travail est motivé par une étude sur les femmes qui ont reçu un diagnostic de cancer du sein à deux occasions différentes. Les récepteurs hormonaux sont importants dans la biologie du cancer du sein, et il est bien reconnu que la mesure de l'état de ces récepteurs hormonaux peut être erronée. La discordance dans l'état des récepteurs hormonaux entre deux cancers primitifs du sein est préoccupante et pourrait être un motif important d'échec de traitement. Nous nous penchons sur un modèle de régression logistique modélisant l'association entre l'état des récepteurs hormonaux de deux cancers, qui introduit des paramètres liés à la classification erronée de l'état des récepteurs hormonaux. La loi de probabilité a priori pour la sensibilité et la spécificité du modèle est basée sur la façon dont l'état des récepteurs hormonaux est déterminé dans les laboratoires cliniques. Une B-spline est utilisée pour prendre en compte l'effet non linéaire d'une covariable sans erreur. Nos résultats indiquent que le taux de concordance réel de l'état des récepteurs hormonaux entre deux cancers primitifs du sein est supérieur à la valeur observée.

[Monday June 12/Lundi 12 juin, 16:00-16:30]

Laura L.E. Cowen (University of Victoria)

Estimating the Population Size of Greater Victoria's Injection Drug Users using Capture-Recapture Methods
Estimation de la taille de la population des utilisateurs de drogues injectables dans la région de Victoria au moyen des méthodes de « capture-recapture »

Population size estimation is critical for planning public health programs for injection drug users. Estimation is difficult, as these populations are considered “hidden” or “hard to reach”. The accepted population size estimate for Greater Victoria, Canada is between 1500 and 2000 individuals, which is dated prior to the year 2000, and is likely an underestimate. I will discuss the use of

L'estimation de la taille de la population est essentielle pour planifier les programmes de santé publique à l'intention des utilisateurs de drogues injectables. L'estimation est difficile, car ces populations sont considérées comme « cachées » ou « difficiles à atteindre ». L'estimation acceptée de la taille de la population dans la région de Victoria, au Canada, est de 1 500 à 2 000 personnes; elle date d'avant l'an 2000 et cette taille est probablement sous-estimée. Je discuterai des mo-

Data Science in Biomedical Studies

Science des données dans les études biomédicales

both closed and open mark-recapture models to estimate population size using cross-sectional survey data collected in 2003, 2005 and 2009. All methods provided population size estimates that were higher than the currently accepted estimate. The open-population estimate of the number of the injection drug users in Greater Victoria was relatively stable over time with fewer than 3000 individuals over the 6-year study. Improved estimates of population size and dynamics will assist in improving access to harm reduction services, which may reduce higher risk drug use practices.

dèles de capture-marquage-recapture à populations fermées et ouvertes pour estimer la taille de la population au moyen de données provenant d'une enquête transversale menée entre 2003, 2005 et 2009. Toutes les méthodes ont permis d'estimer une taille de la population plus élevée par rapport à l'estimation acceptée. L'estimation du nombre d'utilisateurs de drogues injectables dans la région de Victoria a été relativement stable au fil du temps, avec moins de 3 000 personnes au cours de l'étude pendant les six années. De meilleures estimations de la taille et de l'évolution de la population contribueront à améliorer l'accès à des services, ce qui peut engendrer une réduction de l'usage de drogues.

[Monday June 12/Lundi 12 juin, 16:30-17:00]

Wendy Lou (University of Toronto)

Statistical Opportunities and Challenges in Multiple-Platform Biomedical Studies

Opportunités et défis statistiques dans les études biomédicales multiplateformes

It has become increasingly common in biomedical research to involve multiple fields (e.g. neuroscience, nutrition, psychiatry, pediatrics etc.) with researchers sharing common overarching goals, under which specific project aims and study designs are carried out by different teams. Due to the nature of linked and often parallel research programs, information regarding the data structure and integration steps is essential when considering statistical analyses. Examples taken from ongoing multiple platform studies will be presented to illustrate some statistical applications, and selected statistical methods for quality assurance, dimension reduction, and visualization, will be discussed. The pros and cons of commonly used statistical approaches in multiple platform studies will be illustrated through real examples.

Il est de plus en plus courant en recherche biomédicale de regrouper ensemble plusieurs domaines d'expertise (ex : neuroscience, nutrition, psychiatrie, pédiatrie, etc.) réunissant des chercheurs partageant des objectifs communs pour lesquels des projets spécifiques sont exécutés par différentes équipes. En raison de la nature des programmes de recherche liés et souvent menés en parallèle, l'information concernant la structure des données ainsi que des étapes d'intégration est essentielle lorsque des analyses statistiques sont envisagées. Nous présentons des exemples tirés d'études multiplateformes en cours pour illustrer certaines applications statistiques ainsi que certaines méthodes statistiques en matière d'assurance qualité, de réduction de la dimension et de visualisation. Des exemples réels illustreront les avantages et inconvénients des approches statistiques utilisées dans ces études multiplateformes.

Industrial Statistics Statistique industrielle

Chair/Président: Jean-François Plante
Room/Salle: E2 165 (EITC)
Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-15:45]

Da Zhong (Dexen) Xi (Western University), **Charmaine B. Dean** (Western University), **Steve Taylor** (Pacific Forestry Centre)

Dependent Models for the Duration and Size of BC Fires

Modèles dépendants pour la durée et la taille des feux en Colombie-Britannique

The goal of our research is to jointly model time spent (duration) in days and area burned (size) in hectares from ground attack to final control of a fire as a bivariate survival outcome using either a copula model framework that connects the two outcomes functionally, or a joint modeling framework that connects the two outcomes with a shared random effect. The challenges include the specific framework to be employed, how the longitudinal environmental variables (e.g. precipitation, drought indices) are incorporated, and the complexity of computation associated with two outcomes considered jointly. A key aspect of the project is knowledge translation through collaboration with fire scientists at the federal government to implement the optimal framework developed as a component in the fire prediction system that is concurrently under development by Natural Resources Canada. The direct comparison between the statistical properties of the two broad frameworks is also of interest.

L'objectif de notre recherche est de modéliser conjointement le temps (durée) en jours et la superficie (taille) brûlée en hectares à partir de l'attaque au sol jusqu'au contrôle final du feu en un résultat de survie à deux variables en utilisant soit un cadre de modèle de copules qui connecte les deux résultats fonctionnellement, ou un cadre de modèle conjoint qui connecte les deux résultats avec un effet aléatoire partagé. Parmi les défis à relever figurent le cadre spécifique à employer, comment incorporer les variables environnementales longitudinales (par exemples les précipitations et les indices de sécheresse) ainsi que la complexité computationnelle associée à la considération conjointe de deux résultats. Un aspect essentiel du projet est le partage des connaissances, par la collaboration avec les experts en incendies au gouvernement fédéral, pour la mise en place du cadre optimal développé en tant que composante du système de prévision des feux de forêts qui est actuellement en développement par Ressources naturelles Canada. La comparaison directe entre les propriétés statistiques des deux cadres généraux est aussi d'intérêt.

[Monday June 12/Lundi 12 juin, 15:45-16:00]

Yongho Lim (Memorial University of Newfoundland), **Candemir Cigsar** (Memorial University of Newfoundland)

Testing for Parallel Carryover Effects in Redundant Systems

Tester les effets de report parallèle dans les systèmes redondants

Redundancy is an important method to improve the reliability of a repairable system. In this study, we consider clustering of failures in redundant systems due to parallel type of carryover effects, in which the event intensity of a recurrent event process is temporarily increased after event occurrences in other processes. Our main goal is to develop formal test procedures for the assessment of such effects in redundant systems with repairable components connected in parallel and subject to recurrent failures. We therefore develop partial score tests for testing the presence of parallel carryover effects in recurrent event settings. Asymptotic properties of the

La redondance est une méthode importante pour améliorer la fiabilité d'un système réparable. Dans cette étude, nous considérons le regroupement de défaillances dans des systèmes redondants en raison d'effets de report parallèles, dans lesquels l'intensité d'événement d'un processus d'événement récurrent est temporairement augmentée suite à des événements se produisant dans d'autres processus. Notre objectif principal est d'élaborer des procédures de test formelles pour évaluer ces effets dans des systèmes redondants avec des composantes réparables connectées en parallèle et sujettes à des défaillances récurrentes. Nous développons donc des tests de score partiel pour tester la présence d'effets de report parallèle dans des situations d'événements récurrents. Les propriétés asymptotiques

Industrial Statistics Statistique industrielle

test statistics are discussed analytically as well as through simulations. A data set including failure times of diesel power generators operating in remote and isolated communities is analyzed to illustrate the methods developed.

tiques des statistiques de test sont discutées analytiquement et par le biais de simulations. Les méthodes développées sont illustrées en analysant un jeu de données incluant les temps de panne des générateurs électriques au diesel fonctionnant dans des communautés éloignées et isolées.

[Monday June 12/Lundi 12 juin, 16:00-16:15]

Steven Wu (Simon Fraser University)

Building a Basketball Analytics Startup as a Statistics Student

Développer une entreprise à titre d'étudiant en statistiques à partir d'une analytique du basket-ball

We hear it time and time again: data surrounds us, data scientist is the career of the future, etc. With all of this opportunity for us, what does a statistician with entrepreneurial interests need to know before taking the leap? In this talk I'll share my experiences of building a startup that helps university coaches prepare for their next opponent by leveraging modern statistical techniques on play-by-play data. I'll discuss lessons learned that I didn't get in school about product development, business strategy, user acquisition, and customer success. I'll briefly acknowledge successes before focusing on the failures that were specific to the product being developed and marketed as a statistics product, hopefully deriving some useful general advice. I'll conclude with useful resources. My hope is to provide the information that I wish I saw someone speak about before I started this endeavour, and that it encourages anyone with interests in starting their own venture.

On l'entend sans cesse : les données sont partout, spécialiste des données est une carrière d'avenir, etc. Avec toutes ces possibilités qui se présentent, que doit savoir un statisticien intéressé à devenir entrepreneur avant de se lancer en affaires ? Dans le cadre de cette allocution, j'aborde mon expérience en démarrage d'entreprise, une entreprise visant à aider les entraîneurs universitaires dans leurs préparatifs pour affronter leur prochain adversaire, en tirant profit des techniques statistiques modernes de couverture intégrale des rencontres sportives. Je touche un mot des leçons apprises ailleurs que sur un banc d'école : développement de produits, stratégie commerciale, acquisition d'utilisateurs et succès auprès de la clientèle. Évoquant brièvement les succès, je m'attarde ensuite aux échecs relevant précisément du produit développé et commercialisé comme produit statistique, espérant en tirer des conseils pratiques généraux. En guise de conclusion, je présente des ressources utiles. Je vise à fournir l'information que j'aurais souhaité obtenir avant de me lancer dans cette aventure, en espérant que ce soit un encouragement pour d'autres gens intéressés à démarrer leur entreprise.

[Monday June 12/Lundi 12 juin, 16:15-16:30]

Abdollah Safari (Simon Fraser University), **Rachel MacKay Altman** (Simon Fraser University)

Efficient Conversion Probability Estimator for Display Advertising

Estimateur efficace de la probabilité de conversion pour affichage publicitaire

The goal of online display advertising is to entice users to “convert” (i.e., take a pre-defined action such as making a purchase) after clicking on the ad. An important measure of the value of an ad is the probability of conversion. The focus of our project is finding an efficient, accurate, and precise estimator of conversion probability. Challenges associated with the data are the delays in observing conversions and the size of the data set (both number of observations and number of predictors). Two models have been previously considered as a basis for estimation: A logistic regression model and a joint model for the conversion status and delay

L'objectif de la publicité en ligne est d'inciter les utilisateurs à « convertir » (c'est-à-dire faire une action prédéfinie telle que de faire un achat) après avoir cliqué sur la publicité. Une mesure importante de la valeur d'une publicité est la probabilité de conversion. L'objectif de notre projet est de trouver un estimateur de probabilité de conversion précis et efficace. Les défis associés aux données sont les délais dans l'observation des conversions et la taille de l'ensemble de données (le nombre d'observations ainsi que le nombre de prédicteurs). Deux modèles ont précédemment été considérés comme base d'estimation : un modèle de régression logistique et un modèle conjoint pour le statut de conversion et les délais. Ajuster le premier est simple mais ignorer les délais dans la conversion

times. Fitting the former is simple, but ignoring the delays in conversion leads to an under-estimate of conversion probability. The latter is more realistic but computationally expensive to fit. Our proposed estimator is a compromise between these two estimators. We apply our results to a data set from Criteo which is a subset of the clicks occurred over a two-month period along with their final conversion status.

[Monday June 12/Lundi 12 juin, 16:30-16:45]

Bryan Yates (University of Waterloo), **Martin Lysy** (University of Waterloo), **Aleks Labuda** (Asylum Research)

Parametric Spectral Density Estimation in Atomic Force Microscopy

Estimation paramétrique de la densité spectrale dans la microscopie à force atomique

Atomic force microscopy (AFM) experiments provide nanoscopic measurements of the dynamic properties of molecules and atoms. Parametric spectral density estimation features prominently in the analysis and calibration of these experiments, which typically generate large amounts of data. To this end, maximum likelihood estimation can be prohibitively expensive, while a much faster least-squares estimation method incurs a considerable loss in statistical precision. Moreover, both methods are highly sensitive to systematic sources of instrument vibration. We propose a variance-stabilizing procedure to combine the simplicity of least-squares with the efficiency of maximum likelihood, and a de-noising procedure based on significance testing for spectral periodicities. Simulation and experimental results indicate that a two- to ten-fold reduction in mean square error can be expected by applying our methodology.

entraîne une sous-estimation de la probabilité de conversion. Le deuxième est plus réaliste mais très coûteux au niveau des calculs pour l'ajustement. Notre estimateur proposé est un compromis entre ces deux estimateurs. Nous appliquons nos résultats à un ensemble de données provenant de Criteo, qui est un sous-ensemble des clics obtenus sur une période de deux mois ainsi que les états terminaux de conversion.

Les expériences de microscopie à force atomique (MFA) fournissent des mesures nanoscopiques des propriétés dynamiques des molécules et des atomes. L'estimation paramétrique de la densité spectrale joue un rôle important dans l'analyse et le calage de ces expériences, ce qui génère habituellement de grandes quantités de données. À cette fin, l'estimation du maximum de vraisemblance peut être beaucoup trop coûteuse, alors qu'une méthode d'estimation des moindres carrés qui se veut beaucoup plus rapide entraîne une perte considérable de précision statistique. De plus, les deux méthodes sont très sensibles aux sources systématiques de vibration des instruments. Nous proposons une procédure de stabilisation de la variance pour combiner la simplicité des moindres carrés avec l'efficacité du maximum de vraisemblance et une procédure d'élimination du bruit basée sur le test de signification des périodicités spectrales. Les résultats obtenus par simulation et par expérience indiquent que nous pouvons nous attendre à une réduction de deux à dix fois de l'erreur quadratique moyenne en appliquant notre méthodologie.

[Monday June 12/Lundi 12 juin, 16:45-17:00]

Peter Tait (McMaster University), **Paul D. McNicholas** (McMaster University)

Clustering of Functional Data using Mixture Model-Based Clustering

Regroupement de données fonctionnelles à l'aide d'un regroupement basé sur un modèle de mélanges

Clustering of functional data using mixture model-based clustering is considered. Dimension reduction is carried out via functional PCA and the principal components are clustered using a Gaussian mixture model. The proposed method uses an MM algorithm to calculate the MLEs of the mixture model, which allows our algorithm to scale efficiently as the data dimensionality increases. This work is motivated by an applied problem whose goal is to iden-

Nous considérons le regroupement des données fonctionnelles à l'aide d'un regroupement basé sur un modèle de mélanges. La réduction de la dimension se fait par une ACP fonctionnelle et les composantes principales sont groupées en utilisant un modèle de mélanges gaussiens. La méthode proposée utilise un algorithme MM pour calculer les EMV du modèle de mélanges, ce qui permet à notre algorithme de se mettre efficacement à l'échelle à mesure que la dimensionnalité des données augmente. Ce travail est motivé par un problème concret

Industrial Statistics Statistique industrielle

tify different physical activity patterns in children's movement over time. The movement is measured by wearable accelerometers and measurements are recorded every 3 seconds over the course of a week.

dont l'objectif est d'identifier différents modèles d'activité physique dans le mouvement des enfants au fil du temps. Le mouvement est mesuré par des accéléromètres portables et les mesures sont enregistrées toutes les 3 secondes au cours d'une semaine.

Spatial, Environmental, and Network Data

Données spatiales, environnementales et de réseau

Chair/Président: Patrick E. Brown

Room/Salle: E2 160 (EITC)

Abstract/Résumé

[Monday June 12/Lundi 12 juin, 15:30-15:45]

Dongmeng Liu (Simon Fraser University), **Jinko Graham** (Simon Fraser University)

Simple Measures of Individual Cluster-Membership Certainty for Hard Partitional Clustering

Mesures simples de la certitude d'appartenance à un regroupement individuel pour le regroupement de partition dur.

Hard partitional clustering methods such as the Partitioning Around Medoids (PAM) algorithm give no measure of individual cluster-membership certainties. We propose two posterior-probability-like measures of individual cluster-membership certainty which can be applied to a hard partition of the sample. One measure is an extension of the individual silhouette widths and the other is based on the pairwise dissimilarities in the sample. We apply both measures to a set of dissimilarity matrices for categorical data with the PAM algorithm and evaluate their performance on an individual with ambiguous cluster membership in simulated datasets. As a benchmark, we also present the results of a soft clustering algorithm and two model-based clustering methods. The proposed measures behave similarly to posterior probabilities from the soft clustering and model-based clustering methods and are worth considering as a way to augment hard partitional clustering methods.

Des méthodes de regroupement de partition durs telles que l'algorithme « Partitioning Around Medoids » (PAM) ne donnent aucune mesure de certitude d'appartenance à un regroupement individuel. Nous proposons deux mesures de certitude comme une probabilité a posteriori d'appartenance à un regroupement, qui peuvent être appliquées à une partition dure de l'échantillon. Une des mesures est un agrandissement des largeurs individuelles de la silhouette et l'autre est basée sur les dissimilarités entre paires dans l'échantillon. Nous appliquons les deux mesures à un ensemble de matrices de dissimilarité pour les données catégoriques avec l'algorithme PAM et évaluons leur performance sur un individu avec l'appartenance à des groupes ambigus dans des jeux de données simulés. Comme point de repère, nous présentons également les résultats d'un algorithme de regroupement souple et de deux méthodes de regroupement basées sur un modèle. Les mesures proposées se comportent de la même manière que les probabilités postérieures à partir des méthodes de regroupement souples et de regroupement par modélisation, et méritent d'être considérées comme un moyen d'ajouter aux méthodes de partitionnement durs.

[Monday June 12/Lundi 12 juin, 15:45-16:00]

Trevor Thomson (Simon Fraser University), **X. Joan Hu** (Simon Fraser University), **John Braun** (University of British Columbia - Okanagan)

Statistical Modelling of Spot Fires in Forest Fire Control

Modélisation statistique des incendies de surface dans la lutte contre les incendies de forêt

A spot fire is a term used in wildfire management to indicate that an airborne ember ignited a new fire. Under certain meteorologic and wildfire conditions, a spot fire can cross a fuel break, such as a river or road, and allow the wildfire to progress over an otherwise unreachable location. We model the progression of a wildfire under a variety of conditions, and the associated generation of airborne embers that can result in a spot fire. We then derive the probability distributions of the time to the first spot fire occurring across a fuel break. A wildfire

Un incendie de surface est un terme utilisé dans la gestion des incendies de forêt pour indiquer qu'une braise aérienne a allumé un nouvel incendie. Dans certaines conditions météorologiques et d'incendies de forêt, un incendie de surface peut traverser un coupe-feu, comme une rivière ou une route, et permettre la progression du feu de forêt sur un endroit autrement inaccessible. Nous modélisons la progression d'un incendie dans une multitude de conditions et la génération associée de braises aériennes pouvant provoquer un incendie. Nous dérivons ensuite les distributions de probabilité du temps que prendra le premier incendie à se développer au-

Spatial, Environmental, and Network Data

Données spatiales, environnementales et de réseau

simulator is developed, and we present procedures with various types of practical data to estimate the model parameters.

delà d'un coupe-feu. Un simulateur d'incendie de forêt est développé et nous présentons des procédures pour estimer les paramètres du modèle en utilisant différents types de données réelles.

[Monday June 12/Lundi 12 juin, 16:00-16:15]

Shabnam Balamchi (University of Manitoba), **Shabnam Balamchi** (University of Manitoba)

Spatial Modeling of Repeated Events: An Application to Disease Mapping

Modélisation spatiale des événements répétés : une application à la cartographie des maladies

Mixed models are commonly used to analyze spatial data, which frequently occur in practice such as health sciences and life studies. It is customary to incorporate spatial random effects into the model to account for the spatial variation of the data. In particular, Poisson mixed models are used to analyze the spatial count data. We assume that the observations in each area, conditional on spatial random effects, are independent of each other. However, this may not be a valid assumption in practice. E.g., multiple asthma visits by a patient to clinics within a year are not clearly independent observations. To address this issue, we develop spatial models with repeated events. In particular, compound Poisson mixed models are introduced to account for the repeated events as well as the spatial variation of the data. Performance of the proposed approach is evaluated through a simulation study and by application to a dataset of patients' asthma visits to clinics in the province of Manitoba, Canada.

Les modèles mixtes sont couramment utilisés pour analyser les données spatiales, que nous rencontrons souvent en pratique, comme en sciences de la santé et dans les études de la vie. Il est d'usage d'intégrer des effets aléatoires spatiaux dans le modèle pour prendre en compte la variation spatiale des données. En particulier, les modèles mixtes de Poisson sont utilisés pour analyser les données de dénombrement spatial. Nous supposons que les observations dans chaque zone, conditionnelles aux effets aléatoires spatiaux, sont indépendantes les unes des autres. Toutefois, il se peut que cette hypothèse ne soit pas valide en pratique. Par exemple, les visites multiples d'un patient asthmatique dans des cliniques dans un délai d'un an ne sont pas des observations clairement indépendantes. Pour résoudre ce problème, nous développons des modèles spatiaux avec des événements répétés. En particulier, des modèles mixtes de Poisson composés sont introduits pour prendre en compte des événements répétés ainsi que de la variation spatiale des données. La performance de l'approche proposée est évaluée au moyen d'une étude de simulation et d'une application à un jeu de données sur les visites reliées à l'asthme des patients dans les cliniques de la province du Manitoba au Canada.

[Monday June 12/Lundi 12 juin, 16:15-16:30]

Kevin D.J. McGregor (McGill University), **Auréli Labbe** (HEC Montréal), **Celia M.T. Greenwood** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University)

A Penalized Approach to Estimating Network Changes in the Human Microbiome

Une approche pénalisée pour estimer les changements de réseau dans le microbiome humain

The microbiome is the collection of microorganisms colonizing the human body, and plays an integral part in human health. A growing trend in microbiome analysis is to construct a network to estimate the level of co-occurrence between different taxa. Current methods do not facilitate investigation into how these networks change with respect to some phenotype. We propose a model to estimate network changes with respect to a binary phenotype. The counts of individual taxa in the samples are modeled through a Poisson distribution whose

Le microbiome est l'ensemble des micro-organismes qui colonisent le corps humain et il joue un rôle fondamental dans la santé humaine. Une tendance croissante dans l'analyse du microbiome est de construire un réseau pour estimer le niveau de concomitance entre différents taxons. Les méthodes actuelles ne facilitent pas l'étude sur la façon dont ces réseaux changent par rapport au phénotype. Nous proposons un modèle pour estimer les changements du réseau par rapport à un phénotype binaire. Les dénombrements de taxons individuels dans les échantillons sont modélisés par une loi de Poisson dont la moyenne dépend d'un terme gaussien à

Spatial, Environmental, and Network Data

Données spatiales, environnementales et de réseau

mean depends on a Gaussian random effect term. The penalized precision matrix of all the random effect terms determines the co-occurrence network among taxa. The model fit is obtained and evaluated using Monte Carlo methods. Finally, we introduce a means of coding established biological information into the model. The performance of the model is evaluated through an extensive simulation study, and tested in samples from recent gut microbiome studies.

effet aléatoire. La matrice de précision pénalisée des termes à effet aléatoire détermine la concomitance du réseau entre taxons. L'ajustement du modèle est obtenu et évalué en utilisant des méthodes Monte-Carlo. Finalement, nous présentons un moyen de coder l'information biologique bien établie dans le modèle. La performance du modèle est évaluée par une simulation exhaustive et est testée sur des échantillons provenant d'études récentes sur le microbiome de l'intestin.

[Monday June 12/Lundi 12 juin, 16:30-16:45]

Fang He (University of Western Ontario), **Duncan Murdoch** (University of Western Ontario), **Reg Kulperger** (University of Western Ontario)

Statistical Modelling of Carbon Dioxide Flux Data

Modélisation statistique de données de flux de dioxyde de carbone

Climate change is now having such a profound impact on life systems that scientists are preoccupied with phenology. Usually, two sources of data are used in phenology studies, the first of which is remote sensing data, while the other is ground based data. It is costly to get ground-based data at a new location. For this reason, we are interested in building a model and using remote sensing data to predict ground-based data at one location, then connecting multiple nearby locations to predict ground-based data across a surface at a wider range. The data are seasonal with variable phase. The data need to be registered (phase standardized) by year. We build a model based on FDA and time warping function. This gives a statistic model linking ground and remote data. Future work will include spatial aspects of this idea.

Le changement climatique a désormais un impact si profond sur les systèmes de vie que les scientifiques s'inquiètent pour la phénologie. Généralement, on utilise deux sources de données pour les études de phénologie, des données de télédétection et des données terrestres. Or, il est coûteux d'obtenir des données terrestres pour un nouveau site. Nous étudions donc la possibilité de construire un modèle qui utilise des données de télédétection pour prédire les données terrestres d'un site, puis de relier plusieurs sites voisins pour prédire les données terrestres sur une région plus large. Les données sont saisonnières et à phase variable. Elles doivent être enregistrées (normalisées par phase) par année. Nous construisons un modèle fondé sur l'analyse de données fonctionnelles et la fonction de distorsion temporelle. Cela nous donne un modèle statistique qui relie données terrestres et de télédétection. Nos prochains travaux tiendront compte des aspects spatiaux de cette idée.

[Monday June 12/Lundi 12 juin, 16:45-17:00]

Fahimeh Moradi (University of Alberta), **Elham Khodayari Moez** (University of Alberta), **Irina Dinu** (University of Alberta)

Identifying Expression Quantitative Trait Loci in Genome-Wide Association Studies using Matrix eQTL

Identifier les loci de traits d'expression quantitatifs dans des études d'association pangénomiques par l'utilisation de Matrix eQTL

The objective of this study is to identify an efficient, statistically sound and user friendly method for analysis of Expression quantitative trait loci (eQTL) studies. In this study, we performed expression quantitative trait loci (eQTL) analysis using the Matrix eQTL R package. This technique implements matrix covariance calculation and efficiently runs linear regression analysis. False discov-

L'objectif de cette étude est d'identifier une méthode efficace, statistiquement valide et facile d'utilisation pour l'analyse d'études sur les loci de traits d'expression (eQTL). Dans cette étude, nous réalisons une analyse eQTL en utilisant le package Matrix eQTL en R. Cette technique calcule la matrice de covariance et exécute efficacement une analyse de régression linéaire. Le taux de fausses découvertes (FDR) est utilisé pour identifier des eQTL cis et trans significatifs

ery rate (FDR) is used to identify significant cis and trans eQTL for multiple testing corrections. We applied matrix eQTL to a real data set consisting of 730,256 SNP and 33,298 RNA for 173 samples. After processing data, gene SNPs associations can be identified using the ANOVA model. In this study, 15,408 cis eQTL and 27,562 trans eQTL are identified, at a FDR less than 0.05. We found out that matrix eQTL is a computationally efficient and user friendly method for analysis of eQTL studies. The results provide insight into the genomic architecture of gene regulation in inflammatory bowel disease (IBD).

pour des corrections de tests multiples. Nous avons appliqué la matrice eQTL à un ensemble de données réelles composé de 730 256 SNP et de 33 298 RNA pour 173 échantillons. Après le traitement des données, les associations gènes-SNP peuvent être identifiées à l'aide d'un modèle ANOVA. Dans cette étude, 15 408 eQTL cis et 27 562 eQTL trans sont identifiés avec un FDR de moins de 0,05. Nous avons découvert que Matrix eQTL est une méthode efficace d'analyse d'études eQTL facile d'utilisation. Les résultats offrent un aperçu de l'architecture génomique de la régulation des gènes dans les maladies inflammatoires de l'intestin (IBD).

SSC Gold Medal Address
Allocution du récipiendaire de la Médaille d'or de la SSC

Chair/Président: John Petkau
Organizer/Responsable: Michael Evans
Room/Salle: UC 210 (UC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 08:40-09:45]

Harry Joe (University of British Columbia)

Classical and Modern Multivariate Statistical Analysis

Analyse statistique multivariée classique et moderne

Modern multivariate statistical analysis has flexible distributional assumptions, can be applied to continuous and discrete variables, and does not require the classical Gaussian assumption. The most popular methods are based on copulas and the most flexible copula constructions for high dimensions are based on graphical objects called vines. Vine pair-copula constructions, including those with latent variables, can be considered as Gaussian extensions after the correlation or loading matrix is parametrized in terms of correlations and partial correlations that are algebraically independent. It will be shown that some Gaussian models with parsimonious dependence and their copula extensions can be represented with the same graphical models. The truncated vine also provides a parsimonious dependence model for multivariate Gaussian applications. Concrete examples will be used to illustrate the main ideas.

L'analyse statistique multivariée moderne a des hypothèses de distribution flexibles, peut s'appliquer à des variables continues et discrètes et ne nécessite pas l'hypothèse classique de Gauss. Les méthodes les plus populaires se basent sur des copules et les constructions de copules les plus flexibles pour dimensions élevées reposent sur des objets graphiques appelés vignes. Les constructions de vignes de copules par paires, dont celles avec variables latentes, peuvent être considérées comme des extensions de Gauss une fois que la corrélation ou matrice de chargement a été paramétrisée en termes de corrélations ou de corrélations partielles algébriquement indépendantes. Il sera démontré que certains modèles de Gauss avec une dépendance parcimonieuse et leurs extensions de copules peuvent être représentés avec les mêmes modèles graphiques. La vigne tronquée donne aussi un modèle de dépendance parcimonieuse pour des applications gaussiennes multivariées. Des exemples concrets illustreront les idées centrales.

Statistical Modelling in Life Insurance Modélisation statistique en assurance-vie

Chair/Président: Johnny Li
Organizer/Responsable: Johnny Li
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-11:05]

Rui Zhou (University of Manitoba)

Modelling Multi-Population Mortality Dependence with a Regime-Switching Copula

Modéliser la dépendance de la mortalité multi-populations avec une copule à changement de régimes

Multi-population mortality modelling plays a crucial role in the securitization of longevity risk. For example, the payoff of Swiss Re's Kortis bond involves the mortality improvements of both the U.S. and U.K. male populations. Capturing the mortality dependence between the two populations is critical for accurate longevity risk pricing. Recently, some researchers investigated the use of copula in modelling mortality dependence, and observed that mortality dependence is stronger during mortality deteriorations than during mortality improvements. In order to capture observed asymmetric dependence, we propose to use a multivariate regime-switching copula. We study how the choice of copula affects the risk profile of a longevity security and examine the impact of asymmetric dependence and regime-switching in pricing the security.

La modélisation de la mortalité multi-populations joue un rôle essentiel dans la titrisation du risque de longévité. Par exemple, le gain sur l'obligation suisse Re's Kortis implique l'amélioration du niveau de mortalité des populations masculines des États-Unis et du Royaume-Uni. Capturer la dépendance de la mortalité entre les deux populations est critique pour la tarification exacte du risque de longévité. Récemment, quelques chercheurs ont étudié l'utilisation de copules dans la modélisation de la dépendance de la mortalité et ont observé que la dépendance est plus forte durant une détérioration de la mortalité que durant une amélioration de la mortalité. Pour saisir la dépendance asymétrique observée, nous proposons l'utilisation d'une copule multivariée à changement de régime. Nous étudions l'effet du choix de la copule sur le profil de risque d'un titre de longévité et nous examinons l'impact d'une dépendance asymétrique et d'un changement de régime sur la tarification d'un titre.

[Tuesday June 13/Mardi 13 juin, 11:05-11:50]

Cary Chi-Liang Tsai (Simon Fraser University), **Adelaide Di Wu** (Simon Fraser University)

A Hierarchical Credibility Approach to Modeling Mortality Rates for Multiple Populations

Une approche de crédibilité hiérarchique pour modéliser les taux de mortalité de populations multiples

In this talk, we propose a hierarchical credibility approach to modeling multi-population mortality rates using data from the Human Mortality Database. Hierarchical credibility is used in a premium calculation with grouped data from different levels in a tree structure. Traditional mortality models for a single population estimate the model parameters with mortality data grouped into two levels (age and year). We group the mortality data for both genders of several developed countries into four levels (country, gender, age and year) or three levels (population, age and year), apply the hierarchical credibility to each level to better reflect the correlation of mortality rates among populations or genders/countries, and compare the fore-

Dans cet exposé, nous proposons une approche de crédibilité hiérarchique pour modéliser des taux de mortalité multi-populations à l'aide de données provenant de la base de données sur la longévité (« Human Mortality Database »). La crédibilité hiérarchique est utilisée dans le calcul de primes avec les données regroupées de différents niveaux dans une structure en arbre. Les modèles de longévité traditionnels pour population unique estiment les paramètres du modèle avec les données de mortalité regroupés en deux niveaux (âge et année). Nous regroupons les données de mortalité pour les deux sexes de plusieurs pays développés en quatre niveaux (pays, sexe, âge et année) ou en trois niveaux (population, âge et année), nous appliquons la crédibilité hiérarchique à chaque niveau pour mieux représenter la corrélation des taux de mortalité entre les populations ou sexes/pays et nous com-

Statistical Modelling in Life Insurance Modélisation statistique en assurance-vie

casting performances between the model and some multi-population mortality ones.

parons les performances prévisionnelles entre le modèle et quelques modèles de mortalité multi-population.

Business Problems, Data Science Solutions
Problèmes commerciaux, solutions de sciences des données

Chair/Président: Nathaniel T. Stevens
Organizer/Responsable: Nathaniel T. Stevens
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-10:50]

James David Wilson (University of San Francisco), **Jon Mackay** (Waterloo Institute for Complexity Innovation)
The Silicon Valley Wage Cartel: Understanding its Effects on the Regional Labour Market with Dynamic Networks
Le cartel des salaires de la Silicon Valley : Quels en sont les effets sur le marché du travail régional et ses réseaux dynamiques ?

In 2013, Google, Apple, Adobe, Intel settled a class action lawsuit for \$415 million. The suit alleged that these companies colluded with other major corporations in Silicon Valley to suppress employee wages by agreeing not to poach workers from other cartel members. We build on niche theory in a dynamic context through the analysis of a network of manager job changes before and during the period the cartel was active. We investigate two related questions. First, how did the cartel affect managerial labour flows? And second, how did the cartel affect other firms competing for talent in the labour market? We use the temporal exponential graph model (TERGM) to examine the changes in manager job transitions before and during the cartel's activities. We find that firms that were not part of the cartel lost relatively more workers than expected when their niches overlapped with a higher proportion of cartel member firms. This work suggests paths for future labour market policy making.

En 2013, Google, Apple, Adobe et Intel ont réglé un recours collectif pour 415 millions de dollars. La poursuite alléguait que ces entreprises s'étaient entendues avec d'autres grandes entreprises de la Silicon Valley pour supprimer des salaires d'employés en convenant de ne pas débaucher d'employés des autres membres du cartel. Nous nous fondons sur la théorie des niches dans un contexte dynamique en analysant un réseau des changements d'emploi de gestionnaires avant et pendant la période d'activité du cartel. Nous étudions deux questions connexes. Premièrement, comment le cartel a-t-il affecté les flux de main-d'œuvre de gestionnaires ? Et deuxièmement, comment le cartel a-t-il affecté les autres entreprises qui cherchaient à attirer chez-eux des employés de talent présents sur le marché du travail ? Nous utilisons le modèle graphique exponentiel temporel (modèle TERGM) pour examiner l'évolution des transitions d'emploi de gestionnaires avant et pendant les activités du cartel. Nous constatons que les entreprises qui ne faisaient pas partie du cartel ont perdu relativement plus d'employés que prévu quand leurs niches coïncidaient avec une proportion plus élevée d'entreprises membres du cartel. Cette étude suggère des pistes pour l'élaboration future de politiques de l'emploi.

[Tuesday June 13/Mardi 13 juin, 10:50-11:20]

Ali Ghodsi (University of Waterloo)
Generative Mixture of Networks
Mélange de réseaux génératif

We introduce a generative model based on training deep architectures. The model consists of K networks that are trained together to learn the underlying distribution of a given data set. The process starts with dividing the input data into K clusters and feeding each into a separate network. We use an EM-like algorithm to train the networks together and update the clusters of the data. We call this model Mixture of Networks. The provided model is a platform that can be used for any deep struc-

Nous présentons un modèle génératif fondé sur l'apprentissage d'architectures profondes. Le modèle consiste en K réseaux qui apprennent ensemble la distribution sous-jacente d'un ensemble de données déterminé. Le processus commence par une division des données d'entrée en K regroupements qui alimentent des réseaux distincts. Nous utilisons un algorithme de type EM afin que les réseaux apprennent ensemble et mettent à jour les regroupements de données. Nous appelons ce modèle Mélange de réseaux. Le modèle proposé est une plateforme qui peut être utilisée sur n'importe quelle

Business Problems, Data Science Solutions

Problèmes commerciaux, solutions de sciences des données

ture and be trained by any conventional objective function for distribution modeling. As the components of the model are neural networks, it has a high capability in characterizing complicated data distributions as well as clustering data. We apply the algorithm on MNIST handwritten digits and Yale faces datasets. The model can learn the distribution of these data sets. One can sample new data points from these distributions that look like a real handwritten digit or a real face. We also demonstrate the clustering ability of the model using some real-world and toy examples.

structure profonde et formée par n'importe quelle fonction objective conventionnelle à la modélisation des distributions. Comme les composants du modèle sont des réseaux neuronaux, celui-ci est capable de caractériser des distributions de données compliquées et de regrouper des données. Nous appliquons l'algorithme à des chiffres manuscrits de la base de données MNIST et à des ensembles de données de visages de Yale. Le modèle est capable d'apprendre la distribution de ces ensembles de données. On peut échantillonner de ces distributions de nouveaux points de données qui ressemblent à un chiffre manuscrit ou à un visage. Nous prouvons également la capacité de regroupement du modèle à l'aide d'exemples réels et simulés.

[Tuesday June 13/Mardi 13 juin, 11:20-11:50]

Nick Ross (University of San Francisco)

Practical Testing

Les tests pratiques

Technology companies, especially consumer facing ones, have seen an explosion in the use of experiments. Many companies now consider statistical hypothesis testing (such as AB testing) to be an appropriate solution to many questions. Outside of a few well-defined examples, however, a lack of statistical knowledge hampers the efforts of real-world practitioners. The purpose of this talk is to document a few experiments and explain why they failed. We will also consider steps that we, as researchers, academics and practitioners, can take to avoid such failures in the future.

Les entreprises technologiques, surtout celles en rapport direct avec les consommateurs, ont vu le recours à des expériences augmenter de façon vertigineuse. Bon nombre d'entreprises voient aujourd'hui le test d'hypothèse en statistique (tel que le test A/B) comme la réponse appropriée à une foule de questions. À l'exception de quelques exemples bien fondés, un manque de connaissances sur la statistique ralentit cependant les efforts des praticiens dans le monde réel. Mon propos est de documenter certaines expériences et d'expliquer la raison de leur échec. Il sera aussi question des mesures que nous, chercheurs, universitaires et praticiens, pouvons prendre pour éviter ce genre d'échecs à l'avenir.

Beyond “Correlation is not Causation” Au-delà de “Corrélation n’est pas causalité”

Chair/Président: Georges A. Monette
Organizer/Responsable: Georges A. Monette
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-11:05]

Alison L. Gibbs (University of Toronto)

Is All That Coffee I’m Drinking Hurting or Harming Me? Understanding Causality

Une forte consommation de café : nuisible ou nocif? Comprendre la causalité

Students know the adage “Correlation is not Causation” yet research shows that, even after completing an introductory statistics course, they can be fooled into making causal conclusions by relationships that, on the surface, appear plausible, and they perform poorly on assessments of the purpose of randomization in study design. Moreover, most research questions are causal, and yet most data science problems rely on observational data. Unfortunately, research in the statistics education literature has focused largely on sample-to-population inference rather than experiment-to-causation inference. I will talk about some of the problems I have observed that may be inhibiting students’ ability to make appropriate causal inferences, including ambiguity in language and lack of facility in multivariate thinking and I will consider what we might do in the first (and often last) non-mathematical course in statistics, to give our students a deeper, practical understanding of causal inferences.

Les étudiants connaissent l’adage voulant que « corrélation n’égale pas causalité », mais la recherche indique pourtant que même après avoir suivi un cours de base en statistique, ils peuvent être induits en erreur en tirant des conclusions causales concernant des relations qui, en apparence, semblent plausibles. Par ailleurs, ils évaluent mal le but de la randomisation dans un devis d’étude. De plus, la plupart des questions en recherche relèvent de la causalité, quand la plupart des problèmes soulevés en science des données reposent pourtant sur des données d’observation. Malheureusement, la recherche en enseignement de la statistique est surtout centrée sur l’inférence échantillon-population plutôt que sur l’inférence expérience-causalité. Je traite ici de certains problèmes observés qui peuvent inhiber la capacité des étudiants à faire des inférences causales appropriées, y compris l’ambiguïté du langage et le manque de possibilités de réflexion concernant les problèmes multivariés. Je mets aussi de l’avant ce que nous pouvons faire dans le premier (et souvent dernier) cours de statistique non mathématique afin de donner aux étudiants une compréhension plus profonde et plus pratique des inférences causales.

[Tuesday June 13/Mardi 13 juin, 11:05-11:50]

Georges A. Monette (York University)

How to tell when Association might be Causation

Comment savoir quand une association peut être une causalité

The most important decisions we make, as individuals and collectively, are based on our perceptions of answers to causal questions. The cacophony of conflicting claims about the effects of diets, drugs, social and economic policies, are largely the result of the fact that, for most causal questions, we only have observational evidence. As statisticians, we have a unique appreciation of the issues involved with causal inference and we can help improve the public understanding of causality by helping students in our service courses to develop the judgment

Les décisions les plus importantes, à titre personnel ou collectivement, se prennent à partir de notre perception des réponses à des questions causales. La cacophonie des allégations conflictuelles au sujet des effets des diètes, des drogues ou des politiques sociales et économiques, découle en grande partie du fait qu’à la plupart des questions causales ne répondent que des données d’observation. Les statisticiens ont un point de vue unique des problèmes impliquant l’inférence causale et peuvent aider le grand public à mieux comprendre la causalité, en aidant les étudiants suivant leurs cours hors programme à acquérir le jugement nécessaire pour évaluer

Beyond “Correlation is not Causation” Au-delà de “Corrélation n’est pas causalité”

to assess causal claims. The talk will discuss some attempts to achieve this.

les allégations causales. Cette allocution présentera quelques tentatives dans ce but.

Contemporary Statistical Methodology and Application Méthodes statistiques contemporaines et leurs applications

Chair/Président: Liqun Wang
Organizer/Responsable: Liqun Wang
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-10:50]

Hsing-Ming Chang (National Cheng Kung University, Taiwan)

Ergodicity and Approximation of the Conditional Causation Probability of Engineering Systems

Ergodicité et approximation de la probabilité conditionnelle de causalité des systèmes d'ingénierie

We report on the ergodicity of and the approach to approximate the conditional causation probabilities and the conditional cause of failure probabilities via finite Markov chains. This is significant because they provide additional important information to the failure distribution of a system. They especially provide information about the reliability for an aged system. Assisted by a fast algorithm, we demonstrate to readers how to obtain the probabilities and verify our theoretical derivations. The paper should be of interest to a broad range of audiences, including academic researchers and reliability system engineers. This work is a continuation of that entitled "Distributions and causation probabilities of multiple-run-rules and their applications in system reliability, quality control and start-up tests" which has already been accepted for publication.

Cet exposé porte sur l'ergodicité et le mode d'approximation des probabilités conditionnelles de causalité et des probabilités conditionnelles des causes de défaillance, à l'aide de chaînes de Markov à espace d'états fini. Ce sujet important fournit de précieux renseignements sur la distribution des défaillances d'un système. Ces probabilités fournissent en particulier de l'information sur la fiabilité d'un système vieillissant. À l'aide d'un algorithme rapide, l'article décrit le moyen d'obtenir des probabilités et de vérifier les dérivations théoriques. L'article devrait intéresser divers publics, notamment les chercheurs universitaires et les ingénieurs en fiabilité des systèmes. Il fait suite à l'article « Distributions and causation probabilities of multiple-run-rules and their applications in system reliability, quality control and start-up tests », article déjà accepté pour publication.

[Tuesday June 13/Mardi 13 juin, 10:50-11:20]

Su-Fen Yang (National Chengchi University, Taiwan), **Chung-Ming Yang** (Ling Tung University, Taiwan), **Sin-Hong Wu** (National Chengchi University, Taiwan)

A Simple Variance Control Chart with Double Sampling Scheme

Graphique de contrôle simple pour la variance sous un plan d'échantillonnage double

Control charts are effective tools for signal detection in both manufacturing processes and service processes. Much of the data in service industries come from processes exhibiting non-normal or unknown distributions. The commonly used Shewhart variable control charts, which depend heavily on the normality assumption, are not appropriately used here. This paper thus proposes a DS EWMA-AV chart for monitoring process variability. We further explore the sampling properties of the new monitoring statistics, and calculate the average run lengths when using the proposed DS EWMA-AV chart. The performance of the DS EWMA-AV chart and those of non-parametric variance charts are

Les graphiques de contrôle sont des outils efficaces pour la détection de signal dans les processus de fabrication et de service. Dans les industries de service, la plupart des données proviennent de processus qui présentent des distributions non normales ou inconnues. Les graphiques de contrôle par variables de Shewhart, si couramment utilisées, dépendent fortement de l'hypothèse de normalité, si bien qu'elles ne sont pas bien adaptées ici. Cette présentation propose un graphique de contrôle DS EWMA-AV qui permet un suivi de la variabilité d'un processus. Nous explorons ensuite les propriétés d'échantillonnage des nouvelles statistiques de contrôle et calculons la durée de traitement moyenne pour l'utilisation de DS EWMA-AV. Nous comparons la performance du graphique DS EWMA-AV à celle de graphiques non paramé-

Contemporary Statistical Methodology and Application

Méthodes statistiques contemporaines et leurs applications

compared by considering cases in which the critical quality characteristic presents a normal and non-normal distribution. Comparison results show that the proposed chart always outperforms the latter.

triques en étudiant des cas où la caractéristique de qualité critique présente une distribution normale et non normale. Les résultats de la comparaison montrent que le graphique proposé surpasse toujours les autres.

[Tuesday June 13/Mardi 13 juin, 11:20-11:50]

Melody Ghahramani (University of Winnipeg), **Scott S. White** (University of Winnipeg), **Alexander R. de Leon** (University of Calgary)

A Combined Estimating Function Approach to Count Data Regression Modeling

Approche de fonctions d'estimation combinées pour modéliser la régression des données de dénombrement

A flexible semi-parametric regression model for autocorrelated count data is proposed. Unlike earlier models available in the literature, the model does not require construction of a likelihood function and only entails the specification of the first two conditional moments. A simple and flexible combined estimating function approach that makes efficient use of the information contained in the data is adopted for the model in the absence of a likelihood function. Simulation studies are conducted to study the performance of the semi-parametric method when the likelihood is either correctly specified or misspecified. The methodology is illustrated using a monthly polio counts dataset.

Nous proposons un modèle de régression semi-paramétrique flexible pour les données de dénombrement auto-corrélées. Contrairement aux modèles précédemment disponibles, celui-ci ne nécessite pas la construction d'une fonction de vraisemblance et n'exige la spécification que des deux premiers moments conditionnels. Nous adoptons pour le modèle une approche de fonctions d'estimation combinées simple et flexible qui utilise efficacement l'information contenue dans les données, en l'absence d'une fonction de vraisemblance. Nous menons des études par simulation pour déterminer la performance de la méthode semi-paramétrique lorsque la vraisemblance est spécifiée correctement ou incorrectement. Nous illustrons la méthodologie à l'aide d'un ensemble de données de dénombrements mensuels de la polio.

Functional Data Analysis: Models and Applications

Analyse de données fonctionnelles : modèles et applications

Chair/Président: Haocheng Li
Organizer/Responsable: Haocheng Li
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-10:50]

Zhenhua Lin (University of Toronto), **Fang Yao** (University of Toronto)

Adaptive Functional Regression with Manifold Structures

Régression fonctionnelle adaptative avec structures de variété

Statistical methods that adapt to unknown population structures are attractive due to both numerical and theoretical advantages over their non-adaptive counterparts. We contribute to adaptive modelling of functional regression, where challenges arise from the infinite dimensionality in their underlying spaces. We are interested in the scenario where the functional predictor process lies in a nonlinear manifold that is intrinsically finite-dimensional and embedded in an infinite-dimensional function space. We proposed a novel approach built upon local linear manifold smoothing that achieves a polynomial convergence rate that adapts to contamination level and intrinsic manifold dimension even when functional data are observed intermittently and contaminated by noise, in contrast to the logarithm rate in nonparametric functional regression literature. We demonstrate that the proposal enjoys favourable finite sample performance relative to commonly used methods via simulated and real examples.

Les méthodes statistiques qui s'adaptent à des structures de population inconnues sont intéressantes en raison de leurs avantages numériques et théoriques par rapport à leurs contreparties non adaptatives. Nous contribuons à la modélisation adaptative de la régression fonctionnelle, où les défis proviennent d'une dimensionnalité infinie dans leurs espaces sous-jacents. Nous nous intéressons au scénario selon lequel le processus fonctionnel prédicteur réside dans une variété non linéaire qui est intrinsèquement de dimension finie et plongée dans un espace fonctionnel dimensionnel infini. Nous proposons une nouvelle approche fondée sur le lissage de la variété linéaire locale qui engendre une vitesse de convergence polynomiale qui s'adapte au niveau de contamination et à la variété de dimension intrinsèque même si les données fonctionnelles sont observées par intermittences et contaminées par le bruit, en contraste avec le taux logarithmique dans la littérature de régression fonctionnelle non paramétrique. Nous démontrons que la proposition profite d'un rendement sous échantillons finis favorable relativement aux méthodes couramment utilisées au moyen d'exemples simulés et réels.

[Tuesday June 13/Mardi 13 juin, 10:50-11:20]

Jiguo Cao (Simon Fraser University), **Peijun Sang** (Simon Fraser University), **Liangliang Wang** (Simon Fraser University)

Parametric Functional Principal Component Analysis

Analyse paramétrique des composantes principales fonctionnelles

Functional principal component analysis (FPCA) is a popular approach in functional data analysis to explore major sources of variation in a sample of random curves. These major sources of variation are represented by functional principal components (FPCs). Most existing FPCA approaches use a set of flexible basis functions such as B-spline basis to represent the FPCs, and control the smoothness of the FPCs by adding roughness penalties. However, the flexible representations pose difficulties in interpretation. We consider a variety of applications of FPCA and find that, in many situa-

L'analyse fonctionnelle des composantes principales (FPCA) est une approche populaire dans l'analyse des données fonctionnelles pour explorer les principales sources de variation dans un échantillon de courbes aléatoires. Ces principales sources de variation sont représentées par des composantes principales fonctionnelles (FPC). La plupart des approches FPCA existantes utilisent un ensemble de fonctions de base flexibles telles que la base B-spline pour représenter les FPC, et contrôlent la régularité des FPCs en ajoutant des pénalités de rugosité. Cependant, les représentations flexibles posent des difficultés pour les utilisateurs, à savoir, comment interpréter les FPC. Dans cet article, nous considérons une va-

Functional Data Analysis: Models and Applications

Analyse de données fonctionnelles : modèles et applications

tions, the shapes of top FPCs are simple enough to be approximated using simple parametric functions. We propose a parametric approach to estimate the top FPCs to enhance their interpretability for users. Our parametric approach can circumvent the smoothing parameter selecting process in conventional nonparametric FPCAs. Our simulation study shows that the proposed parametric FPCA is more robust when outlier curves exist. The parametric FPCA is demonstrated in several datasets from a variety of settings.

riété d'applications de la FPCA et constatons que, dans de nombreuses situations, les formes des FPC supérieures sont assez simples pour être approchées en utilisant des fonctions paramétriques simples. Nous proposons une approche paramétrique pour estimer les FPC supérieures afin d'améliorer leur interprétation pour les utilisateurs. Notre approche paramétrique peut également contourner le processus de sélection des paramètres de lissage dans les méthodes classiques nonparamétriques FPCA.

[Tuesday June 13/Mardi 13 juin, 11:20-11:50]

Linglong Kong (University of Alberta), **Zhengwu Zhang** (SAMSI), **Xiao Wang** (Purdue University), **Hongtu Zhu** (MD Anderson Cancer Centre)

Spatial Quantile Regression Models for High-Dimensional Imaging Data

Modèles de régression quantile spatiale pour données d'imagerie de grande dimension

We aim to develop a spatial quantile regression framework for accurately quantifying high-dimensional image data conditional on some scalar predictors. This new framework allows us to delineate spatial quantile association between neuroimaging data and covariates, while explicitly modeling spatial dependence in neuroimaging data. Theoretically, we establish the minimax rates of convergence for the prediction risk under both fixed and random designs. We further develop efficient algorithms such as the ADMM and the primal-dual algorithm to estimate the varying coefficients. Our method is able to estimate the whole conditional distribution of the image response given the scalar covariates. Simulations and real data analysis are conducted to examine the finite-sample performance.

Nous visons à mettre au point un cadre de régression quantile spatiale afin de quantifier précisément des données d'images de grande dimension conditionnelles à des prédicteurs scalaires. Ce nouveau cadre nous permet de délimiter l'association entre les données de neuroimagerie et les covariables, tout en modélisant explicitement une dépendance spatiale dans des données de neuroimagerie. Nous établissons de façon théorique des taux minimax de convergence pour la prédiction du risque dans le cadre de plans fixes et aléatoires. Nous poursuivons le développement d'algorithmes efficaces existants comme la méthode de direction alternée des multiplicateurs (ADMM) et l'algorithme primal-dual afin d'estimer les coefficients variables. Notre méthode permet d'estimer la loi conditionnelle de la réponse d'image étant donné les covariables scalaires. Nous analysons des données de simulation et réelles pour examiner l'efficacité sous échantillons finis.

Methods for Censored and Recurrent Events

Méthodes pour les événements censurés et récurrents

Chair/Président: Karen Kopciuk

Room/Salle: E2 165 (EITC)

Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-10:35]

Chien-Lin Mark Su (McGill University), **Russell Steele** (McGill University), **Ian Shrier** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University)

Analysis of Recurrent Events Data Based on Trend-Renewal Process

Analyse de données d'événements récurrents fondée sur un processus de renouvellement de tendance

Recurrent events data arises in many biomedical and longitudinal studies when failure events can occur repeatedly for each subject during the follow-up time. We are interested in the gap times between recurrent events. We propose a semiparametric accelerated transform gap time model based on a trend-renewal process which contains a trend and a renewal component. We use the Buckley-James imputation approach to deal with censored transform gap times. The proposed estimators are shown to be consistent and asymptotically normal. Model diagnostic plots of residuals and a prediction method for predicting number of recurrent events given a specified covariate and follow-up time are also presented. Simulation studies are conducted to assess the finite sample performances of the proposed method. The proposed technique is demonstrated through an application to two real data sets.

Les données d'événements récurrents sont présentes dans de nombreuses études biomédicales et longitudinales, dans un contexte où chaque sujet peut présenter des événements de santé répétés pendant une période de suivi. Nous nous intéressons au laps de temps entre ces événements récurrents. Nous proposons un modèle semi-paramétrique pour les laps de temps transformés et accélérés, fondé sur un processus de renouvellement de tendance qui contient une composante de tendance et une composante de renouvellement. Nous utilisons l'approche d'imputation de Buckley-James pour traiter les laps de temps transformés tronqués. Les estimateurs proposés s'avèrent convergents et asymptotiquement normaux. Nous présentons des graphiques de diagnostic de validation de modèle basés sur les résidus ainsi qu'une méthode de prédiction du nombre d'événements récurrents pour une covariable et une durée de suivi données. Nous effectuons des études de simulation pour évaluer la performance de la méthode proposée sur des échantillons finis. Nous illustrons la technique proposée par une application à deux ensembles de données réelles.

[Tuesday June 13/Mardi 13 juin, 10:35-10:50]

Taehan Bae (University of Regina)

On the Mixtures of Length-biased Weibull Distributions

À propos des mélanges de distributions de Weibull avec biais de longueur

In this talk, a new class of length-biased Weibull mixtures will be introduced and a review of its basic distributional properties will be given. As a generalization of the Erlang mixture distribution, the length-biased Weibull mixture distribution has an increased flexibility to fit various shapes of data distributions including heavy-tailed ones. Methods of statistical estimation and model selection will be discussed with applications on real catastrophic loss data sets.

Dans cet article, nous présenterons une nouvelle catégorie de mélanges de distributions de Weibull avec biais de longueur et examinerons ses propriétés distributionnelles de base. Dans le cadre d'une généralisation du mélange de distributions d'Erlang, le mélange de distributions de Weibull avec biais de longueur a plus de souplesse pour s'adapter aux différentes formes de distributions de données, notamment les distributions à queue épaisse. Nous discuterons des méthodes d'estimation statique et de la sélection du modèle, ainsi que des applications sur des jeux de données réelles de sinistre catastrophique.

[Tuesday June 13/Mardi 13 juin, 10:50-11:05]

Defen Peng (University of British Columbia and ICVHealth), **Kevin Burke** (University of Limerick), **Gilbert**

Methods for Censored and Recurrent Events Méthodes pour les événements censurés et récurrents

MacKenzie (University of Limerick)

An Interval Censored MPR Survival Model

Modèle multiparamétrique de régression pour données de survie censurées par intervalle

We develop multi-parameter regression survival models for the interval censored survival data setting by means of the multi-parameter Weibull regression survival model which is wholly parametric and non-proportional hazards. We describe the model, develop the interval-censored likelihood with or without frailty involved, conduct a more detailed simulation study designed to investigate the effect of sample size and the proportion of right censored observations, and analyse data from a population-based study of survival from lung cancer conducted in the UK. We compare the findings obtained with an analysis of these data using right censored multi-parameter regression survival methods. To the best of our knowledge this is the first time the effect of interval censoring has been investigated in the MPR survival model setting.

Nous développons des modèles multiparamétriques de régression pour des données de survie censurées par intervalle au moyen du modèle multiparamétrique de régression de Weibull pour données de survie, qui est un modèle à risques paramétriques et non proportionnels. Nous décrivons le modèle, calculons la vraisemblance censurée par intervalle avec ou sans fragilité, menons une étude de simulation plus poussée afin d'examiner l'effet de la taille d'échantillon et la proportion d'observations censurées à droite, et nous analysons les données à partir d'une étude de survie sur la population ayant survécu à un cancer du poumon en Grande-Bretagne. Nous comparons les résultats obtenus par une analyse de ces données au moyen de méthodes multiparamétriques de régression sur des données de survie censurées à droite. À notre connaissance, c'est la première fois que l'effet de la censure par intervalle a été étudié dans le cadre du modèle multiparamétrique de régression.

[Tuesday June 13/Mardi 13 juin, 11:05-11:20]

Rachel MacKay Altman (Simon Fraser University), **Andrew Henrey** (Simon Fraser University)

Practical Considerations when Analyzing Small Samples of Discrete Survival Times Using the Grouped Relative Risk Model

Considérations pratiques pour l'analyse de petits échantillons de données de temps de survie discrets à l'aide du modèle de risque relatif de groupe (GRRM)

The grouped relative risk model (GRRM) is a popular semi-parametric model for analyzing discrete survival time data. The maximum likelihood estimators (MLEs) of the regression coefficients in this model are often asymptotically efficient relative to those based on a more restrictive, parametric model. However, in settings with a small number of sampling units, the usual properties of the MLEs are not assured. In this talk, we discuss computational issues that can arise when fitting a GRRM to small samples, and describe conditions under which the MLEs can be ill-behaved. We find that, overall, estimators based on a penalized score function behave substantially better than the MLEs in this setting and, in particular, can be far more efficient.

Le modèle de risque relatif de groupe est un modèle semi-paramétrique populaire pour l'analyse de données de temps de survie discrets. Les estimateurs du maximum de vraisemblance (EVM) des coefficients de régression de ce modèle sont souvent asymptotiquement efficaces par rapport à ceux basés sur un modèle paramétrique plus restrictif. Dans un contexte où il n'y a qu'un petit nombre d'unités d'échantillonnage, les propriétés ordinaires des EVM ne sont pas assurées. Cette présentation porte sur les problèmes computationnels qui peuvent se poser en adaptant un modèle de risque relatif de groupe à de petits échantillons ainsi que sur les conditions sous lesquelles le comportement des EVM est incorrect. Nous voyons qu'en règle générale, les estimateurs basés sur une fonction de score pénalisée se comportent notablement mieux que les EVM dans ce contexte et plus particulièrement, ils peuvent être vraiment plus efficaces.

[Tuesday June 13/Mardi 13 juin, 11:20-11:35]

Saima K. Khosa (University of Saskatchewan), **Shahedul A. Khan** (University of Saskatchewan)

A Flexible Parametric Proportional Hazards Model for Time-to-Event Data

Un modèle de risques proportionnels paramétriques flexible pour les données de durées de vie

Methods for Censored and Recurrent Events

Méthodes pour les événements censurés et récurrents

Proportional hazard (PH) models can be formulated with or without assuming a distribution for survival times. The former assumption leads to parametric models, whereas the latter leads to the semi-parametric Cox model which is by far the most popular in survival analysis. However, a parametric model may lead to more efficient estimates than the Cox model under certain conditions. Only a few parametric models are closed under the PH assumption, the most common of which is the Weibull that accommodates only monotone hazard functions. We propose a three-parameter distribution, which is closed under the PH assumption. The model is flexible and parsimonious in the sense that it accommodates all four basic shapes of the hazard function (increasing, decreasing, unimodal and bathtub shape) at the small cost of estimating one additional parameter compared to the Weibull PH model. The model can be used in analyzing time-to-event data, recurrent event data and joint modeling of time-to-event and longitudinal data. Comparative studies based on real and simulated data reveal that the model can be valuable in adequately describing different types of time-to-event data.

Les modèles de risque proportionnel (RP) peuvent être formulés avec ou sans hypothèse de distribution des temps de survie. La première situation mène à des modèles paramétriques, tandis que la seconde mène au modèle semi-paramétrique de Cox, de loin le plus populaire en analyse de survie. Cependant, un modèle paramétrique peut mener à des estimations plus efficaces que le modèle de Cox dans certaines conditions. Seuls quelques modèles paramétriques sont fermés sous l'hypothèse RP, le plus commun étant celui de Weibull qui n'accepte que les fonctions de risque monotones. Nous proposons une distribution à trois paramètres, qui est fermée sous l'hypothèse RP. Le modèle est souple et parcimonieux en ce sens qu'il permet de s'adapter aux quatre formes fondamentales de la fonction de risque (croissante, décroissante, unimodale et en forme de baignoire) au faible coût d'estimer un paramètre supplémentaire par rapport au modèle RP de Weibull. Le modèle peut être utilisé pour analyser les données de durées de survie, les données d'événements récurrents et la modélisation conjointe des données de durées de vie et longitudinales. Des études comparatives à l'aide de données réelles et simulées montrent que le modèle peut décrire adéquatement différents types de données de durées de vie.

New Research in Survey Methods Nouvelles recherches en méthodes d'enquête

Chair/Président: Wesley Yung
Room/Salle: E2 155 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-10:35]

Puying Zhao (University of Waterloo), **David Haziza** (Université de Montréal), **Changbao Wu** (University of Waterloo)

Empirical Likelihood and Semiparametric Inference with Survey Data

Vraisemblance empirique et inférence semi-paramétrique en présence de données d'enquête

Survey data are often collected under a complex sampling design and the finite population parameters are typically defined as the solution to the so-called population (census) estimating equations. In this paper, we address inferential problems where the population estimating equations also involve an unknown nuisance function which does not depend on the parameters. Survey weighted estimating equations with a semiparametric plug-in component are constructed for empirical likelihood based inferences. Maximum empirical likelihood estimators are introduced together with a set of sufficient conditions that ensure the root-n consistency and asymptotic normality of the estimators. Effects of using a plug-in estimator of the nuisance function on the empirical likelihood inference of the finite population parameters are examined. The proposed method is applicable to a wide range of problems, including the generalized Lorenz curve and the Gini measure of income inequalities. Numerical results based on simulated and real data are provided.

Les données d'enquête sont souvent recueillies sous échantillonnage complexe et les paramètres de population finie sont typiquement définis comme la solution aux équations d'estimation de ladite population (recensement). Dans cette présentation, nous abordons les problèmes d'inférence qui se posent lorsque les équations d'estimation de la population incluent aussi une fonction de nuisance inconnue qui ne dépend pas des paramètres. Nous construisons des équations d'estimation pondérées selon l'enquête avec un composant plug-in semi-paramétrique pour les inférences fondées sur la vraisemblance empirique. Nous introduisons des estimateurs du maximum de vraisemblance empirique, ainsi qu'un ensemble de conditions suffisantes qui garantissent la convergence de la racine-n et la normalité asymptotique des estimateurs. Nous examinons les effets de l'utilisation d'un estimateur plug-in de la fonction de nuisance sur l'inférence de la vraisemblance empirique des paramètres de population finie. La méthode proposée peut s'appliquer à de multiples problèmes, dont la courbe de Lorenz généralisée et le coefficient de Gini qui mesure l'inégalité des revenus. Nous présentons des résultats numériques fondés sur des données simulées et réelles.

[Tuesday June 13/Mardi 13 juin, 10:35-10:50]

Matei Mireuta (Statistics Canada), **Jessica Andrews** (Statistics Canada, Business Surveys Methods Division (BSMD)), **Pierre Daoust** (Statistics Canada, Business Surveys Methods Division (BSMD))

The Relative Deviation from Predicted Values as a Tool to Prioritize Units for Failed Edit Follow-Up in IBSP

La différence relative par rapport aux valeurs prédites comme outil de priorisation du suivi des unités dans le cadre du PISE

The production of official statistics can be considerably hindered by errors in collected data. The process of correcting these errors is referred to as statistical data editing and has traditionally been an expensive and time consuming manual effort. However, it is becoming increasingly accepted in practice that correction of a small subset of influential errors is often sufficient to guarantee a good quality estimate and therefore, that additional correction of data past a certain threshold yields only

La production de statistiques officielles est souvent compliquée par la présence d'erreurs dans les données collectées. La correction de ces erreurs se fait traditionnellement par un long processus de vérification manuelle des données. Cependant, il est de plus en plus reconnu que la correction d'un petit sous-ensemble d'erreurs influentes est souvent suffisante pour garantir une bonne qualité des estimations finales et donc, que la correction au-delà d'un certain seuil n'a que peu d'impact. La vérification sélective des données s'efforce d'identifier ces erreurs influentes, d'en évaluer l'impact sur

New Research in Survey Methods

Nouvelles recherches en méthodes d'enquête

minor improvements in quality. The concept of selective editing involves identifying these significant errors, estimating their impact on the final estimate and providing for a signal to halt editing operations when an error tolerance is reached. In this work, we compare several methods based on predicted values for estimating reporting errors and for prioritizing units for failed-edit follow-up in the context of the Integrated Business Statistics Program at Statistics Canada.

[Tuesday June 13/Mardi 13 juin, 10:50-11:05]

Min Jiang (Statistics Canada)

Job Vacancy and Wage Survey: Balancing Sampling and Operational Requirements using the Cube Method

Enquête sur les postes vacants et les salaires : équilibrer l'échantillonnage et les contraintes opérationnelles avec la méthode du Cube

Balanced sampling is a sampling method where the totals estimated with the Horvitz-Thompson estimator are the same or close to the true population totals for a given set of auxiliary variables on the survey frame. If the auxiliary variables are highly correlated to the variables of interest, the variances of the estimators for totals will be small. In Statistics Canada's Job Vacancy and Wage Survey (JVWS), the quarterly sample of business locations has to be split into three monthly subsamples for data collection. In parallel, all locations under the same enterprise must be collected the same month. In this presentation, we will show how one can use the balanced sampling method to allocate the sampling units at the enterprise level in a way that keeps the number of employees and the number of locations balanced for each province and industry between months. This allocation strategy was implemented for the JVWS survey using the Cube method which was proposed by Deville and Tillé (2004).

les estimations finales et de signaler l'arrêt des opérations de vérification lorsqu'un certain seuil d'erreur est atteint. Dans ce travail, nous comparerons des méthodes utilisant des valeurs prédites pour l'évaluation de l'ampleur des erreurs de réponse et pour la priorisation du suivi des unités dans le cadre du Programme Intégré de Statistiques des Entreprises de Statistique Canada.

L'échantillonnage équilibré est une méthode d'échantillonnage selon laquelle les totaux estimés par l'estimateur de Horvitz-Thompson sont les mêmes ou près des vrais totaux de population pour un certain ensemble de variables auxiliaires de la base de sondage. Quand ces variables auxiliaires sont bien corrélées aux variables d'intérêt, la variance des estimateurs pour des totaux est faible. Dans l'Enquête sur les postes vacants et les salaires (EPVS) de Statistique Canada, l'échantillon trimestriel d'emplacements commerciaux doit être réparti en trois sous-échantillons mensuels pour la collecte des données. En parallèle, la collecte des emplacements d'une même entreprise doit se faire le même mois. Dans cette présentation, nous montrerons comment l'échantillonnage équilibré peut être utilisé pour répartir l'échantillon au niveau de l'entreprise de sorte à balancer le nombre d'employés et le nombre d'emplacements selon les provinces et les secteurs industriels entre les mois. Cette stratégie de répartition a été mise en production dans l'EPVS en utilisant la méthode du Cube proposée par Deville et Tillé (2004).

[Tuesday June 13/Mardi 13 juin, 11:05-11:20]

Amanda Halladay (Statistics Canada), **Ming Jie Yang** (Statistics Canada), **François Brisebois** (Statistics Canada)

Segmenting the Canadian Population to Remedy Non-Response in Statistics Canada Household Surveys

Segmentation de la population canadienne afin de remédier à la non-réponse aux enquêtes auprès des ménages de Statistique Canada

Statistical agencies are facing the challenge of declining response rates with an increased demand for national statistics. We must find innovative ways to improve collection strategies while maintaining

Les instituts de statistiques font face au défi d'une baisse des taux de réponse combinée à une demande accrue pour des statistiques nationales. Nous nous devons de trouver des moyens novateurs afin d'améliorer les stratégies de collecte

New Research in Survey Methods

Nouvelles recherches en méthodes d'enquête

high-quality data. One innovation is to use characteristics associated with survey non-response in order to segment the population into homogenous clusters. Advertising campaigns and communication tools can then be tailored to specific clusters of the population, emphasizing areas with higher non-response. In this talk, we combine data from various household surveys with sociodemographic population characteristics from the 2011 Canadian Census in order to profile survey non-response. Logistic regressions identify characteristics associated with survey non-response and then a cluster analysis creates clusters of geographic areas having similar prevalence for these characteristics associated with non-response.

tout en continuant de fournir aux besoins en données de grande qualité. Une innovation consiste à utiliser les caractéristiques associées à la non-réponse afin de segmenter la population en groupes homogènes. Des campagnes publicitaires et des outils de communication peuvent alors être adaptés pour des groupes spécifiques, en mettant l'accent sur les endroits où la non-réponse est plus élevée. Dans cet exposé, nous combinons les données provenant de diverses enquêtes auprès des ménages aux caractéristiques sociodémographiques de la population provenant du recensement de 2011, afin de dresser un portrait de la non-réponse aux enquêtes. Des régressions logistiques sont effectuées afin de déterminer les caractéristiques associées à la non-réponse, puis une analyse de regroupements permet de créer des regroupements d'aires géographiques ayant une prévalence semblable pour ces caractéristiques associées à la non-réponse.

[Tuesday June 13/Mardi 13 juin, 11:20-11:35]

Jane Wang (Statistics Canada)

The Collection Challenges of the 2016 Reverse Record Check Survey

Défis à la collecte de données dans la contre-vérification des dossiers 2016

The 2016 Reverse Record Check (RRC) Survey is used to estimate the number of in-scope persons who were missed by the 2016 Census of Population. The RRC is one of the most important and challenging surveys done by Statistics Canada. Successfully tracing and interviewing the persons selected in the RRC sample is one of the biggest challenges, even though several sources of administrative data are used to provide more information to find these persons. The RRC collection operations faced quite a few challenges for the 2016 survey, including the redesign of the survey contents and of the multi-mode collection methodology, a new Computer Assisted Telephone Interview (CATI) application and a new tracing application. This article describes the 2016 RRC collection challenges and strategies.

La contre-vérification des dossiers (CVD) de 2016 est utilisée pour estimer le nombre de personnes dans le champ de l'enquête qui n'ont pas été dénombrées dans le Recensement de la population 2016. La CVD est l'une des enquêtes les plus importantes et complexes menées par Statistique Canada. Le dépistage et l'interview des personnes sélectionnées dans l'échantillon de la CVD est l'un des plus gros défis, même si plusieurs sources de données administratives sont utilisées pour retrouver ces personnes. Les opérations de collecte de CVD ont dû relever plusieurs défis lors du recensement de 2016, notamment la refonte du contenu de l'enquête et des méthodes de collecte multimode, une nouvelle application téléphonique assistée par ordinateur et une nouvelle application de dépistage. Cet article décrit les défis de collecte et les stratégies adoptées pour la CVD en 2016.

[Tuesday June 13/Mardi 13 juin, 11:35-11:50]

Gabrielle Poirier (Statistics Canada), **Kevin Bosa** (Statistics Canada), **François Gagnon** (Statistics Canada)

Methods of Selecting a Person within a Household: Mailed Invitations for an Electronic Questionnaire

Méthodes de sélection d'un membre à l'intérieur d'un ménage : invitations postales pour un questionnaire électronique

Currently, within-household selection for electronic questionnaire surveys consists of sending a paper invitation to a random sample of households asking for an online roster of the occupants. Then, a random selection of a person is made within the elec-

Pour les questionnaires électroniques, la sélection aléatoire d'une personne dans un ménage consiste à envoyer par la poste une lettre d'invitation demandant de fournir une liste électronique des membres du ménage. Ensuite, l'application sélectionne aléatoirement une personne. Cependant, dans le

New Research in Survey Methods

Nouvelles recherches en méthodes d'enquête

tronic questionnaire. A household with more than one occupant might require two people to connect to the application. There is also a risk that the second person does not want to answer, which creates non-response even if the first person completed their part. The challenges of the roster method raised some questions. What if the selected person could instead be determined from instructions on the mailed invitation? Would it be clear enough so people know who is selected? Will they follow these instructions? To answer these questions, the last-birthday method and a new version of the age-order method were tested and compared to the roster method. Response rates and selection inaccuracy rates were analyzed, and results will be presented.

cas d'un ménage ayant plus d'un membre, il est possible que la personne ayant fourni l'information sur les membres du ménage ne soit pas choisie et donc qu'une deuxième personne ait à se connecter à l'application. Il y a alors un risque que la deuxième personne ne veuille pas participer, ce qui augmenterait le taux de non-réponse. Les défis de la méthode d'énumération en ligne soulèvent quelques questions. Que se passerait-il si nous pouvions sélectionner un membre à partir d'instructions sur la lettre? Serait-ce assez compréhensible pour que les membres du ménage sachent qui est sélectionné? Suivront-ils ces instructions? Pour répondre à ces questions, la méthode basée sur la date d'anniversaire et une méthode basée sur l'âge ont été mises à l'essai et comparées à la méthode classique d'énumération en ligne. Les taux de réponses et d'inexactitude de la sélection ont été analysés et les résultats seront présentés.

Statistical Computing and Complex Models Calcul statistique et modèles complexes

Chair/Président: Russell Steele

Room/Salle: E2 160 (EITC)

Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-10:35]

Jiaxiu Li (University of New Brunswick), **Guohua Yan** (University of New Brunswick), **Renjun Ma** (University of New Brunswick)

Joint Modelling of Count, Continuous and Semi-Continuous Longitudinal Data

Modélisation conjointe de données longitudinales de dénombrement, continues et semi-continues

In medical studies, different types of outcomes are frequently collected over time on each of many subjects. For example, both CD4 cell counts and viral loads of patients are usually observed longitudinally in human immunodeficiency virus (HIV) research. Therefore, the effectiveness of treatments or interventions can be profiled based on these bivariate responses while accounting for their association. As the nature of relationships between these two outcomes are of particular interest, these two outcomes should be analyzed jointly instead of separately. In this study, we propose to model data of mixed types jointly by incorporating both subject-specific and temporally dependent random effects into Tweedie models of different index parameters. An optimal estimation of our model has been developed using orthodox best linear unbiased predictor of random effects. Our analysis results do not rely on any distributional assumption of random effects.

Dans les études médicales, on collecte souvent différents types d'issues au fil du temps pour chacun des sujets. Par exemple, on effectue des observations longitudinales du compte de cellules CD4 et de la charge virale des patients dans les études sur le virus de l'immunodéficience humaine (VIH). L'efficacité des traitements ou interventions peut donc être décrite sur la base de ces réponses bivariées, tout en tenant compte de leur association. Étant donné que la nature des relations entre ces deux issues présente un intérêt particulier, ces issues devraient être analysées ensemble plutôt que séparément. Dans cette étude, nous nous proposons de modéliser des données de divers types ensemble, en intégrant des effets aléatoires spécifiques au sujet et dépendant du temps dans des modèles de Tweedie pour divers paramètres d'indices. Nous mettons au point une estimation optimale de notre modèle à l'aide du meilleur prédicteur linéaire sans biais (BLUP) des effets aléatoires. Les résultats de notre analyse sont indépendants de toute hypothèse de distribution des effets aléatoires.

[Tuesday June 13/Mardi 13 juin, 10:35-10:50]

Kelly Ramsay (University of Manitoba), **Stephane Durocher** (University of Manitoba), **Alexandre Leblanc** (University of Manitoba)

Efficiently Computing the Projection Median in R^3

Calculer efficacement la médiane de projection dans R^3

This talk presents efficient algorithms for computing the projection median exactly in R^3 . The projection median, first introduced by Durocher and Kirkpatrick (2009), is a robust, non-parametric, descriptor of multivariate location. Computing the projection median in R^d involves averaging projections over infinite directions on the d-dimensional unit sphere. Previously, only approximation algorithms existed for computing the projection median for $d > 2$. Our algorithm begins by transforming the problem from computing this average over infinite directions to computing the median level in an arrangement of a finite set of hyperplanes. Next, the

Cet exposé présente des algorithmes efficaces pour le calcul de la médiane de projection en R^3 . La médiane de projection, présentée pour la première fois par Durocher et Kirkpatrick (2009), est un descripteur de position multivariée robuste et non paramétrique. Calculer la médiane de projection dans R^d implique de faire la moyenne des projections dans un nombre infini de directions sur la sphère unitaire de d-dimensions. Auparavant, il n'existait que des algorithmes d'approximation pour calculer la médiane de projection pour $d > 2$. Notre algorithme commence par transformer le problème du calcul de moyenne dans un nombre infini de directions à calculer le niveau médian d'un arrangement d'un ensemble fini d'hyperplans. Ensuite, l'algorithme aléatoire d'Agarwal et al. (1998)

Statistical Computing and Complex Models Calcul statistique et modèles complexes

randomized algorithm of Agarwal et al. (1998) for calculating the median level in an arrangement of planes is applied, before transforming the solution back to the original setting. Exact and efficient computation opens the door to practical use of this statistic in higher dimensions, as well as further study on the properties of the projection median.

pour le calcul du niveau médian d'un arrangement de plans est appliqué, avant de transformer à nouveau la solution dans la structure initiale. Le calcul exact et efficace ouvre la porte à l'utilisation pratique de cette statistique dans des dimensions supérieures, ainsi qu'à des études complémentaires sur les propriétés de la médiane de projection.

[Tuesday June 13/Mardi 13 juin, 10:50-11:05]

Yizhou Fang (University of Waterloo), **Martin Lysy** (University of Waterloo), **Don Mcleish** (University of Waterloo)

Common-Factor Multivariate Stochastic Volatility Modeling with Observable Proxy

Modèle de volatilité stochastique multivarié à facteur commun avec un intermédiaire observable

Multivariate modeling of financial assets often plays a critical role in portfolio analysis and risk management. While Stochastic Volatility (SV) models have enjoyed enormous success in the univariate setting, multivariate SV models are more challenging to design and implement, owing to the large amount of latent variables and the need for correlations to satisfy a positive-definiteness constraint. We propose a common-factor multivariate SV model, naturally extending the univariate case, with an interpretable hierarchical correlation structure for which positive-definiteness is guaranteed. We show how the common factor can be flexibly proxied by observable volatilities, resulting in enormous computational savings for parameter inference. The model is used to conduct a multivariate SV analysis for several major components of the S&P500.

La modélisation multivariée des actifs financiers joue un rôle critique dans l'analyse du portefeuille et dans la gestion du risque. Bien que les modèles de volatilité stochastique (SV) aient connu un grand succès dans le contexte univarié, les modèles multivariés SV sont plus difficiles à concevoir et à implanter dû à la grande quantité de variables latentes et au besoin de corrélations satisfaisant la contrainte d'une définitivité positive. Nous proposons un modèle SV multivarié à facteur commun, élargissant ainsi naturellement le cas univarié, avec une structure de corrélation hiérarchique pour laquelle la définitivité positive est garantie. Nous démontrons comment des volatilités observables peuvent agir en tant qu'intermédiaire flexible pour le facteur commun, entraînant de grandes économies computationnelles dans l'inférence de paramètre. Le modèle est utilisé pour procéder à une analyse SV multivariée de plusieurs des principales composantes du S&P 500.

[Tuesday June 13/Mardi 13 juin, 11:05-11:20]

Yun Ling (University of Waterloo), **Martin Lysy** (University of Waterloo)

Superfast Inference for Stationary Gaussian Processes

Inférence très rapide pour processus gaussiens stationnaires

Stationary Gaussian processes are popular models in many areas of statistical applications. When the observations are regularly spaced, the structure of the variance matrix admits "fast" likelihood evaluations via the famous Durbin-Levinson algorithm, scaling as $O(N^2)$ in the number of observations instead of the usual $O(N^3)$. Here we adapt the lesser-known Generalized Schur algorithm to further reduce this cost to $O(N \log^2 N)$. Our "superfast" method extends to both gradient and hessian calculations, and thus is applicable to many inference algorithms Bayesian and Frequentist alike. Our implementation is available via the R package "Super-

Les processus gaussiens stationnaires sont des modèles très populaires dans plusieurs domaines de statistiques appliquées. Lorsque les observations sont réparties également, la matrice de variance admet des évaluations de vraisemblance « rapides » au moyen de l'algorithme bien connu de Durbin-Levinson d'ordre $O(N^2)$ au lieu de $O(N^3)$ dans le nombre d'observations. Nous adaptons ici un algorithme moins connu, l'algorithme généralisé de Schur, pour réduire encore plus le coût à $O(N \log^2 N)$. Notre méthode « très rapide » s'étend aux calculs de gradient et de hessien et s'applique ainsi à plusieurs algorithmes d'inférence bayésiens et fréquentistes. Notre implémentation est disponible via le package R « SuperGauss », surpassant Durbin-Levinson au-

Statistical Computing and Complex Models

Calcul statistique et modèles complexes

Gauss”, beating Durbin-Levinson around $N = 300$. We present an application to Gaussian process factor analysis.

tour de $N=300$. Nous présentons une application au processus gaussien d’analyse factorielle.

Statistical Computing and Complex Models Calcul statistique et modèles complexes

[Tuesday June 13/Mardi 13 juin, 11:20-11:35]

Sahir R. Bhatnagar (McGill), **Yi Yang** (McGill University), **Alexia Jolicoeur-Martineau** (Jewish General Hospital), **Ashley Wazana** (McGill University), **Celia M.T. Greenwood** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University)

Strong Heredity Penalized Regression Models for Non-Linear Gene-Environment Interactions

Modèles de régression pénalisés ayant la propriété d'hérédité forte pour les interactions entre facteurs génétiques et environnementaux

Diseases are now thought to be the result of changes in entire biological networks whose states are affected by a complex interaction of genetic and environmental factors. In general, power to estimate interactions is low, the number of possible interactions could be enormous and their effects may be non-linear. Existing approaches such as the lasso might keep an interaction but remove a main effect, which is problematic for interpretation. We develop a model for linear and non-linear interactions in penalized regression models that automatically enforces the strong heredity property. A computationally efficient fitting algorithm combined with a non-parametric screening approach scales to high-dimensional datasets and has been implemented in an R package. We apply our method to identify gene-prenatal maternal depression interactions on negative emotionality in mother-infant dyads from the Maternal Adversity, Vulnerability, and Neurodevelopment (MAVAN) cohort.

On pense maintenant que les maladies sont le résultat de changements dans des réseaux biologiques entiers dont l'état est affecté par une interaction complexe entre facteurs génétiques et environnementaux. En général, la puissance pour estimer les interactions est faible, le nombre d'interactions possibles peut être élevé et leurs effets peuvent être non linéaires. Les approches existantes, telles que le lasso, pourraient inclure une interaction mais exclure un effet principal, ce qui rend l'interprétation problématique. Nous proposons un modèle pour les interactions linéaires et non linéaires dans le contexte des modèles de régression pénalisés qui force automatiquement la propriété d'hérédité forte. Un algorithme computationnellement efficace, combiné à une approche de dépistage non paramétrique, s'adapte à des ensembles de données de grande dimension et a été implémenté dans un paquet R. Nous appliquons notre méthode pour identifier les interactions entre gènes et dépression maternelle prénatale sur l'émotivité négative chez les dyades mère-enfant de la cohorte MAVAN (Maternal Adversity, Vulnerability, and Neurodevelopment).

[Tuesday June 13/Mardi 13 juin, 11:35-11:50]

Reza Ramezan (California State University, Fullerton)

Computational Neuroscience: A Romance of Many Disciplines

La neuroscience computationnelle : un lien étroit entre plusieurs disciplines

Computational neuroscience studies the brain from an information processing point of view. A strong body of literature suggests that the nervous system has a fundamental stochasticity, making statistical models very attractive in computational neuroscience. This being said, the adage "all models are wrong, but some are useful" applies. Among the useful is a collection of data-driven models tailored to understand neurophysiological phenomena and the process of information coding in the brain. This talk reviews, briefly, some of the common statistical methods and recent advancements in the analysis of neural spike trains. Data visualization, computational challenges, biological justification, and

La neuroscience computationnelle étudie le cerveau sous l'angle du traitement de l'information. Une documentation étoffée laisse entendre que le système nerveux est doté d'une stochasticité fondamentale, rendant les modèles statistiques très attrayants dans ce domaine d'études. Cela dit, l'adage selon lequel « tous les modèles sont erronés, mais certains sont utiles » s'y applique. Parmi ceux qui sont utiles on compte les modèles guidés par les données, taillés sur mesure pour comprendre les phénomènes neurophysiologiques et le processus de codage de l'information dans le cerveau. Cette allocution se veut un bref tour d'horizon de certaines méthodes statistiques courantes ainsi que des progrès récents dans l'analyse des trains d'impulsions neuronales. Nous abordons également la visualisation des données, les problèmes computationnels,

Statistical Computing and Complex Models

Calcul statistique et modèles complexes

performance of these models will also be discussed in addition to exciting opportunities for future research.

la justification biologique et la performance de ces modèles, en plus de voir d'intéressantes possibilités pour de futurs travaux de recherche.

Statistical Methods for Infectious Diseases and Other Health Outcomes

Méthodes statistiques pour les maladies infectieuses et autres résultats cliniques

Chair/Président: Tyler Williamson

Room/Salle: E2 150 (EITC)

Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 10:20-10:35]

Zhaoyang Tian (University of Waterloo), **Pengfei Li** (University of Waterloo), **Kun Liang** (University of Waterloo)

Nonparametric Estimation in a Compound Mixture Model with Application to a Malaria Study

Estimation non paramétrique dans un modèle de mélanges composé avec application à une étude sur la malaria

Malaria can be diagnosed by the presence of parasites and symptoms (usually fever). However, an individual may have fever attributable either to malaria or to other causes. As such, the parasite level of an individual with fever follows a two-component mixture distribution. Further, the parasite levels of some nonmalaria individuals are exactly zero so that the distribution of the parasite levels in nonmalaria individuals also follows a mixture distribution. We propose a maximum multinomial likelihood approach for estimating the proportion of clinical malaria using parasite-level data from two groups of individuals. The first is collected from the endemic area, whereas the second is from the community, where all parasite levels are from nonmalaria individuals. We develop an EM-algorithm and show its convergence locally. Simulations show that the proposed estimator is more efficient than existing nonparametric estimators using only the frequencies of zero and nonzero data. The proposed method is used to analyze data from a malaria survey carried out in Tanzania.

La malaria peut être diagnostiquée par la présence de parasites et de symptômes (fièvre, généralement). Cependant, un individu peut souffrir de fièvre pour d'autres raisons. Ainsi, le niveau de parasites d'un individu souffrant de fièvre suit une distribution de mélange à deux composants. De plus, les niveaux de parasites de certains individus ne souffrant pas de la malaria sont d'exactement zéro, si bien que la distribution des niveaux de parasites des individus ne souffrant pas de la malaria suit également une distribution de mélange. Nous proposons une approche de maximum de vraisemblance multinomiale pour estimer la proportion de malaria clinique, avec des données sur les niveaux de parasites de deux groupes d'individus. Le premier provient de la région endémique, tandis que le second vient de la communauté, où tous les niveaux de parasites proviennent d'individus ne souffrant pas de la malaria. Nous mettons au point un algorithme EM et en montrons la convergence locale. Nos simulations montrent que l'estimateur proposé est plus efficace que les estimateurs non paramétriques existants qui n'utilisent que la fréquence de zéros et de non zéros. Nous utilisons la méthode proposée pour analyser les données d'une étude sur la malaria effectuée en Tanzanie.

[Tuesday June 13/Mardi 13 juin, 10:35-10:50]

Thuva Vanniyasingam (McMaster University), **Caitlin Daly** (McMaster University), **Xuejing Jin** (McMaster University), **Yuan Zhang** (McMaster University), **Charles Cunningham** (McMaster University), **Gary Foster** (McMaster University), **Lehana Thabane** (McMaster University)

Investigating the Impact of Design Characteristics on Statistical Efficiency within Discrete Choice Experiments: A Systematic Survey

Étude d'impact des caractéristiques d'un plan d'étude sur l'efficacité statistique dans le cadre d'expériences de choix discrets : une étude systématique

This systematic survey aimed to review simulation studies of discrete choice experiments (DCEs) to determine what design features affect statistical efficiency. Electronic searches were conducted in JSTOR, Science Direct, PubMed and OVID. Screening and data extraction were performed independently and in duplicate. The reporting quality

Cette étude systématique vise à passer en revue les études en simulation d'expériences de choix discrets (ECD) afin de déterminer quelles caractéristiques d'un plan d'étude affectent l'efficacité statistique. Des recherches par voie électronique ont été menées dans les bibliothèques JSTOR, Science Direct, PubMed et OVID. Le tri et l'extraction des données ont été exécutés de façon indépendante et en duplicata. La

Statistical Methods for Infectious Diseases and Other Health Outcomes

Méthodes statistiques pour les maladies infectieuses et autres résultats cliniques

of simulation studies were also evaluated in each study. From 371 potentially relevant studies, 9 proved eligible. Studies showed statistical efficiency improved when: increasing the number of choice tasks or alternatives; decreasing the number of attributes, attribute levels, or overlaps; incorporating response behaviour or heterogeneity; correctly specifying Bayesian priors; minimizing prior variances; or matching the method to the research question. Studies need to improve reporting of: study objectives, failures, random number generators, starting seeds, and software. These results may help to inform investigators during DCE design creation.

qualité des rapports de ces études en simulation a été évaluée pour chacune d'elles. Parmi 371 études potentiellement pertinentes, neuf se sont révélées admissibles. Les études indiquaient que l'efficacité statistique s'améliorait dans les cas suivants : nombre accru de choix de tâche ou de solution de rechange ; baisse du nombre et des niveaux des attributs, ou des chevauchements ; incorporation du comportement ou de l'hétérogénéité des réponses ; spécification correcte des lois à priori bayésiennes ; réduction des variances des lois à priori, ou ; concordance de la méthode avec la question. Il faut que les rapports d'études soient améliorés sur les aspects suivants : objectifs de l'étude, défaillances, générateurs de nombre aléatoire, valeurs de départ et équipement logiciel. Ces résultats peuvent servir à renseigner les chercheurs pour la conception d'un plan ECD.

[Tuesday June 13/Mardi 13 juin, 10:50-11:05]

Md. Mahsin (University of Calgary), **Rob Deardon** (University of Calgary)

Geo-Dependent Individual-Level Models for Infectious Diseases Transmission

Modèles géo-dépendants au niveau des individus pour la transmission des maladies infectieuses

In recent years, a class of complex statistical models, known as individual-level models (ILMs), have been effectively used to model infectious disease transmission in discrete time. These models are well developed but assume the probability of disease transmission between two individuals depends only on their spatial (or network-based) separation and not on their spatial locations. However, spatially varying demographic and environmental factors could influence the disease transmission. Thus, it might be useful to incorporate the effect of the spatial location itself into the ILMs. In this study, we extend ILMs to Geo-dependent ILMs (GD-ILMs) that allow the evaluation of the effect of spatially varying social risk factors, environmental factors as well as unobserved spatial structure, upon the transmission of infectious disease. We consider a conditional autoregressive (CAR) model to capture the effects of unobserved spatially structured latent covariates or measurement error. We show how GD-ILMs can be fitted to data within a Bayesian statistical framework using Markov chain Monte Carlo (MCMC) methods.

Une classe de modèles statistiques complexes, connus sous le nom de modèles au niveau des individus (MNI), a été récemment utilisée efficacement pour modéliser la transmission des maladies infectieuses en temps discret. Ces modèles bien développés supposent que la probabilité de transmission de la maladie entre deux individus ne dépend que de leur séparation spatiale (ou basée sur le réseau) et non de leur localisation spatiale. Des facteurs démographiques et environnementaux variant spatialement pourraient pourtant influencer la transmission d'une maladie. Incorporer l'effet de l'emplacement spatial dans les MNI pourrait s'avérer utile. Dans cette étude, nous étendons les MNI à des MNI géodépendantes (MNI-GD) permettant d'évaluer l'effet de divers facteurs de risque social, des facteurs environnementaux et de la structure spatiale non observée sur la transmission des maladies infectieuses. Nous considérons un modèle autorégressif conditionnel (MAC) pour capturer les effets des covariables latentes non observées structurées spatialement ou des erreurs de mesure. Nous montrons comment les MNI-GD peuvent être ajustés à des données dans un cadre statistique bayésien en utilisant des méthodes de chaînes de Markov.

[Tuesday June 13/Mardi 13 juin, 11:05-11:20]

Claire Boteler (Queen's University), **David Thomson** (Queen's University), **Troy Day** (Queen's University)

Seasonality in Influenza Mortality: Moving Beyond the Annual Cycle

Caractère saisonnier de la mortalité causée par la grippe : au-delà du cycle annuel

Statistical Methods for Infectious Diseases and Other Health Outcomes

Méthodes statistiques pour les maladies infectieuses et autres résultats cliniques

Influenza is an infectious disease, and its periodic patterns are commonly modeled focusing on a yearly cycle. However, considering that four pandemics have occurred in the past century, this poses the question of whether there could be any hidden patterns. Using times series analysis and multitaper spectral analysis techniques, an investigation of the mortality due to Influenza that occurred in the United States from 1910 to 2016 has shown some interesting features. Periodic signals of a one-year period were found to be significant, as well as signals in the low frequency range, near 30 years. The analysis also revealed significant frequencies of 2, 4 and 5 cycles per year. It is then possible that these frequencies have a biological connection to influenza, or that they are an artifact created by the fact that the yearly cycle is not absolutely identical each year. It then follows that there might be a more complex periodic structure to this influenza mortality data.

La grippe est une maladie infectieuse et ses tendances périodiques sont couramment modelées selon un cycle annuel. Compte tenu des quatre pandémies à survenir au siècle dernier, on doit cependant se demander s'il ne pourrait pas exister de tendances cachées. À l'aide de techniques d'analyse de séries chronologiques et d'analyse spectrale multitaper, une recherche sur la mortalité causée par la grippe aux États-Unis entre 1910 et 2016 a mis en lumière certaines caractéristiques intéressantes. Les signaux périodiques dans le cadre d'une année étaient significatifs de même que les signaux peu fréquents autour de 30 ans. L'analyse a aussi révélé des fréquences significatives de deux, quatre et cinq cycles par année. Il est possible que ces fréquences aient un lien biologique avec la grippe ou qu'elles soient façonnées par un cycle annuel pas tout à fait identique d'une année à l'autre. Il s'ensuit qu'il pourrait exister une structure périodique plus complexe de ces données sur la mortalité causée par la grippe.

[Tuesday June 13/Mardi 13 juin, 11:20-11:35]

Nicholas Mitsakakis (University of Toronto)

Benefits of Transformations in the Analysis of Health Utility Data

Bénéfices des transformations dans l'analyse des données d'utilité

Health utilities represent a single global measure of Health Related Quality of Life and are necessary for economic evaluation of health interventions. Their data distribution is semi-continuous, skewed and leptokurtic with upper bound at 1 and probability mass at the bound. Linear regression is not appropriate for the analysis of these data, since many assumptions (linearity, homoscedasticity and normality) are violated. Here we investigate the benefits of applying a mathematical transformation to the response variable, before fitting a linear regression model. We use simulated and real health utility data to compare a number of transformations with the untransformed model, using several measures of model accuracy and goodness of fit. Our investigation identifies the benefits of transformations in this context, offering suggestions for future analysis and modeling of health utilities.

En économie de la santé, l'utilité représente une mesure globale unique de la qualité de vie liée à la santé et sont nécessaires à l'évaluation économique des interventions de santé. La distribution des données d'utilité est semi-continue, asymétrique et leptokurtique avec une borne supérieure à 1 et une masse de probabilité à la borne. La régression linéaire n'est pas adéquate pour l'analyse de ces données car les hypothèses de linéarité, homoscedaticité et normalité ne sont pas vérifiées. Ici, nous étudions les avantages de transformer la variable réponse avant d'ajuster le modèle de regression linéaire. Nous utilisons des données simulées et des données d'utilité pour comparer un certain nombre de transformations avec le modèle sans transformation à l'aide de mesures d'exactitude des modèles et de l'exactitude de l'ajustement. Nos recherches ont identifié les avantages des transformations dans ce contexte et nous offrons des suggestions pour les analyses à venir et la modélisation des données d'utilité.

[Tuesday June 13/Mardi 13 juin, 11:35-11:50]

Katherine Davies (University of Manitoba), **William Volterman** (Syracuse University)

Progressively Type-II Censored Competing Risks Data from the Linear Exponential Distribution

Risques concurrents soumis à une censure progressive de type II à partir d'une distribution exponentielle linéaire

We consider competing risks under a progressively type-II censored sample where the competing risks each follow a linear exponential law. We develop likelihood inference when there is a known number of competing risks. To demonstrate the performance of the maximum likelihood estimates, we consider the case of two competing risks and carry out an extensive simulation study. We also apply our inferential method to a real data set.

Nous examinons des risques concurrents dans un échantillon soumis à une censure progressive de type II où chacun de ces risques suit une loi exponentielle linéaire. Nous élaborons l'inférence de vraisemblance lorsque le nombre de risques concurrents est connu. Pour démontrer la performance des estimateurs du maximum de vraisemblance, nous étudions le cas de deux risques concurrents et procédons à une vaste étude en simulation, appliquant par ailleurs notre méthode d'inférence à un ensemble de données réelles.

Poster Session
Séance d'affichage

Room/Salle: Atrium (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Angelina Pesevski (McMaster University), **Brian Franczak** (MacEwan University), **Paul D. McNicholas** (McMaster University)

High Dimensional Clustering: tHDDC

Regroupement de grande dimension : tRDGD

High-dimensional data are collected daily in many fields of work, such as DNA analysis and social media data analysis. The presence of clusters in a data set can be crucial to discovering new trends and patterns, e.g., disease subtypes. Therefore, it is necessary to have reliable and quick techniques which can detect and define clusters. Model-based clustering is a very popular method and uses a mixture of distributions where each component density corresponds to a cluster. The most common model-based clustering techniques are based on using a mixture of multivariate normal distributions. A method called high-dimensional data clustering (HDDC) has given rise to a very computationally efficient family of Gaussian mixture models for high-dimensional data. HDDC is based on the idea that high-dimensional data can be represented in much lower-dimensional subspaces. The HDDC family of models has gained vast attention due to its superior performance compared to other families of mixture models. We propose a t-analogue of this family, which we call the tHDDC family. The tHDDC family extends the high-dimensional data clustering models to include the multivariate-t distribution.

Les données de grande dimension sont recueillies quotidiennement dans de nombreux champs d'activité, comme l'analyse de l'ADN et l'analyse des données des médias sociaux. La présence de grappes dans un jeu de données peut être cruciale pour découvrir de nouvelles tendances et modèles, par exemple des sous-types de maladies. Par conséquent, il est nécessaire d'avoir des techniques fiables et rapides pour détecter et définir les grappes. Le regroupement selon le modèle est une méthode très populaire qui utilise un mélange de distributions où chaque composante de densité correspond à une grappe. Les techniques de regroupement selon le modèle les plus courantes reposent sur l'utilisation d'un mélange de distributions normales multivariées. Une méthode dite de regroupement de données de grande dimension (RDGD) a donné lieu à une famille de modèles de mélange gaussien très efficace sur le plan informatique pour traiter des données de grande dimension. Le RDGD est basé sur l'idée que des données de grande dimension peuvent être représentées dans des sous-espaces de dimensions beaucoup plus petites. L'intérêt porté à la famille de modèles RDGD s'est accru en raison de sa performance supérieure par rapport à d'autres familles de modèles de mélange. Nous proposons un t analogue de cette famille, que nous appelons la famille tRDGD. La famille tRDGD étend les modèles de regroupement de données de grande dimension pour inclure la distribution de t multivariée.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Mohsen Soltanifar, **Annie Dupuis** (Clinical Research Services, Sickkids Hospital, Toronto, Canada), **Russell Schachar** (Psychiatry Research, Neuroscience and Mental Health, Sickkids Hospital, Toronto, Canada), **Michael Escobar** (Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Canada)

Stop Signal Reaction Time: A New Weighted Frequentist Estimation Method

Temps de réaction du signal d'arrêt : une nouvelle méthode d'estimation fréquentiste pondérée

The stop signal reaction time (SSRT), a measure of the latency of the stop signal process, has been theoretically formulated using a horse race model of go and stop signal processes by the American scientist Gordon Logan in 1994. Current analysis from a general population sample of 13696 children age 6-17

Le temps de réaction du signal d'arrêt (TRSA), une mesure de la latence du processus de signal d'arrêt, a été formulé de manière théorique à l'aide d'un modèle de course de chevaux de processus de signaux de départ et d'arrêt par le scientifique américain Gordon Logan en 1994. L'analyse actuelle d'un échantillon de 13 696 enfants âgés de 6 à 17 ans à To-

Poster Session Séance d'affichage

in Toronto demonstrated significantly higher probability of successful inhibit on stop trials following a stop trial compared to those following a go trial, suggesting the need for a modified weighted mixture approach to estimate SSRT. Our Results showed: (i) Significantly reduced go response times after stop trials versus after correct go trials where mean go response times were assumed to follow an Ex-Gaussian distribution (Mean Difference=53.4 ms; p -value<0.0001); (ii) Significantly larger estimates of SSRT using a pooled estimate weighted by trial type versus the original Logan 1994 method with an Ex-Gaussian distribution assumption (Mean Difference=7.6 ms; p -value<0.0001).

ronto provenant d'une population générale a démontré une probabilité significativement plus élevée d'inhibition réussie des essais d'arrêt après un essai d'arrêt comparativement à ceux suivant un essai de départ, d'où la nécessité d'une approche modifiée de mélanges pondérés pour estimer le TRSA. Nos résultats ont montré : (i) une réduction significative des temps de réponse après les essais d'arrêt, comparativement aux essais corrects de départ, où l'on a supposé que les temps de réponse moyens au départ suivent une distribution ex-gaussienne (différence moyenne = 53,4 ms; valeur de $p < 0,0001$); (ii) estimations significativement plus grandes du TRSA en utilisant une estimation groupée pondérée par le type d'essai par rapport à la méthode originale de Logan de 1994 avec une hypothèse de distribution ex-gaussienne (différence moyenne = 7,6 ms, valeur de $p < 0,0001$).

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Junchi Bin (University of British Columbia - Okanagan), **Bryan Gardiner** (Data Nerds), **Zheng Liu** (University of British Columbia (Okanagan))

Advanced Regression Models for Automated Valuation Model

Modèles de régression avancés pour le modèle de valorisation automatisée

The automated valuation model (AVM) is a mathematical program to estimate the market value of real estates based on the real estate data. Currently, boosting is a popular predictive approach in AVM. In 2006, Geoffrey Hinton ignited the popularity of neural networks by showing substantially better performance with a “deep” neural network. However, the application of deep learning algorithm has not been fully explored in the AVM market. In this paper, we investigated advanced statistical learning approaches for AVM. With the linear regression model as the reference, this study investigated the accuracy of advanced statistical learning approaches such as boosting, random forest, support vector regression and other models for real estate data. Moreover, a constructing deep belief networks (DBN) for regression is considered in this application. An ensemble approach to integrate deep learning method and other statistical learning methods for AVM is also investigated in this research.

Le modèle de valorisation automatisé (MVA) est un programme mathématique permettant d'estimer la valeur marchande des biens immobiliers en fonction des données immobilières. Actuellement, la stimulation est une approche prédictive populaire dans les MVA. En 2006, Geoffrey Hinton a accru la popularité des réseaux de neurones en démontrant une performance nettement meilleure avec un réseau de neurones « profond ». Cependant, l'application de l'algorithme d'apprentissage en profondeur n'a pas été entièrement explorée sur le marché des MVA. Dans cet article, nous avons étudié des approches avancées d'apprentissage statistique pour les MVA. En prenant comme référence le modèle de régression linéaire, cette étude a évalué l'exactitude des approches avancées d'apprentissage statistique telles que la stimulation, la forêt d'arbres décisionnels, la régression du vecteur de soutien et d'autres modèles pour les données immobilières. De plus, nous considérons la construction de réseaux de conviction profonde (DCP) pour la régression dans cette application. Par l'entremise de cette recherche, nous étudions une approche d'ensemble pour intégrer la méthode d'apprentissage en profondeur et d'autres méthodes pour les MVA.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Tayler Dawn Scory (University of Calgary), **Karen Kopciuk** (Alberta Health Services)

Genetic Risk Prediction for Type X Colorectal Cancer Families

Prédiction du risque génétique du cancer colorectal familial de type X

Type X colorectal cancer (CRC) occurs in individu-

Le cancer colorectal de type X survient chez l'individu vers

Poster Session Séance d'affichage

als at approximately age 50, and they typically have at least two affected relatives. No major gene has been found so far to explain the clustering of CRC in these families. We will investigate several genetic models for the familial Type X CRC phenotype by carrying out a complex segregation analysis on Type X pedigree data, which will provide evidence in favour of a major gene, polygenic or environmental component. Furthermore, we will incorporate a polygenic threshold model into a genetic risk model to estimate the probability that an individual is a carrier of the disease-causing genetic component. We predict that individuals with a family history of Type X CRC will have an increased probability of carrying the genetic component, and thus have a higher chance of being diagnosed with colorectal cancer. Ultimately, this research will have applications in genetic counselling and cancer prevention strategies such as screening guidelines.

50 ans, et en général, au moins deux membres de sa famille sont affectés. On n'a pas encore trouvé de gène majeur qui explique cette concentration de cas de cancer colorectal dans ces familles. Nous examinerons plusieurs modèles génétiques pour le phénotype de cancer colorectal familial de type X en effectuant une analyse de ségrégation complexe sur les données de pédigrées de type X, ce qui donne des indices quant à l'existence d'un gène majeur et d'une composante polygénique ou environnementale. De plus, nous incorporerons un modèle polygénique à seuil dans un modèle de risque génétique afin d'estimer la probabilité qu'un individu est porteur de la composante génétique causant la maladie. Nous prévoyons que les individus avec des antécédents familiaux de cancer colorectal de type X ont plus de risque d'être porteurs de la composante génétique, et ainsi d'être atteints d'un cancer colorectal. Enfin, ces recherches trouveront des applications dans les stratégies en matière de conseil génétique et de prévention et de dépistage du cancer.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

René Ferland (Université du Québec à Montréal), **Sorana Froda** (Université du Québec à Montréal), **Anthony Coache** (Université du Québec à Montréal)

Modeling and Treatment of Surveillance Flu Data

Modélisation et traitement des données de la surveillance de la grippe

In this paper, we propose a novel look at the CDC (Centers for Disease Control and Prevention, USA) surveillance data on yearly influenza outbreaks. First we apply a flexible model, based on a non-homogeneous birth and death process, which allows to estimate a lower bound to the basic reproduction ratio, an indicator of the transmission potential of an epidemic. Further, we explore various ways of comparing such data across years and regions by resorting to a conditional model which allows to circumvent some of the drawbacks of the original data. In particular, we apply some standard procedures in mixture modeling in order to assess whether two outbreaks are similar, for example if they attain their peak at the same time.

Dans cet article, nous posons un regard nouveau sur les données annuelles de surveillance des épidémies de la grippe du CDC (Centers for Disease Control and Prevention, USA). Premièrement, nous appliquons un modèle flexible basé sur un processus de naissances et de décès non homogène, ce qui nous permet d'estimer une borne inférieure au rapport de reproduction de base, un indicateur du potentiel de transmission d'une épidémie. Ensuite, nous explorons différentes façons de comparer de telles données à travers les années et les régions en ayant recours à un modèle conditionnel qui permet de contourner quelques-uns des inconvénients des données originales. Nous appliquons notamment quelques procédures habituelles dans la modélisation de mélange pour évaluer si deux épidémies sont similaires, par exemples si elles atteignent leur sommet au même moment.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Dola Pathak (University of Nebraska at Lincoln), **Christopher Bilder** (University of Nebraska at Lincoln)

Estimating a Disease Prevalence with Group Testing Data and Imperfect Assays

Estimation de la prévalence d'une maladie à l'aide de données de tests de groupe et d'essais imparfaits

Infectious disease assays can be imperfect. When estimating disease prevalence, this imperfection is accounted for by incorporating assay sensitivity and

Les essais sur les maladies infectieuses peuvent être plus ou moins réussis. Pour estimer la prévalence d'une maladie, cette imperfection est prise en compte par l'incorporation de la

Poster Session Séance d'affichage

specificity into point and variance estimates. Unfortunately, these accuracy measures are often treated as fixed constants, rather than taking into account that they are actually estimates from an assay validation process. The purpose of this presentation is to show the detrimental effect of not taking into account this sampling variability when samples are obtained through group testing (a.k.a., pooled testing). We show that confidence interval coverage can dramatically decline as the sample size increases for the main sample of interest. As a remedy for this problem, we propose a new confidence interval which takes into account the extra sampling variability. This new interval is shown to obtain coverage near the nominal level.

sensibilité et de la spécificité de l'essai dans des estimations ponctuelles et de variance. Malheureusement, l'exactitude de ces mesures est souvent traitée comme une constante fixe, plutôt que pour ce qu'elle est, c.-à-d. une estimation dérivée d'un processus de validation de l'essai. Cette présentation a pour but de montrer l'effet préjudiciable de ne pas tenir compte de la variabilité de l'échantillonnage quand on obtient les échantillons à l'aide de tests de groupe. Nous montrons que la couverture de l'intervalle de confiance peut subir une baisse marquée lorsque la taille de l'échantillonnage s'accroît pour le principal échantillon d'intérêt. Pour remédier à ce problème, nous proposons un nouvel intervalle de confiance qui prend en compte la variabilité additionnelle de l'échantillonnage. Nous montrons que la couverture de ce nouvel intervalle est proche du niveau nominal.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Sarah A. MacQueen (University of British Columbia Okanagan), **Rebecca Tyson** (University of British Columbia Okanagan)

Modelling the Effects of Memory on Bumblebee Pollination Services

Modélisation des effets de la mémoire sur les services de pollinisation des bourdons

Experimental and observational studies have shown that bumblebees return faithfully to particular forage locations, which can have a significant effect on the pollination services they provide. This research project develops stochastic models for bumblebees with and without spatial memory to investigate the resulting pollination services. The topics of memory in movement models and pollination services have previously been researched separately, but have not yet been combined. In this work, movement is modelled with a combination of correlated and uncorrelated random walks, in an agent based model inspired by partial differential equations and informed by recent bumblebee radar tracking studies. Pollination services are quantified by flower visitation rates, tracked as the bumblebee moves throughout the landscape. The results of this work will inform future modelling work on the landscape requirements for optimal pollination of blueberry crops by bumblebees.

Les études expérimentales et d'observation montrent que les bourdons retournent fidèlement vers des sites de butinage particuliers, ce qui peut avoir un effet notable sur les services de pollinisation qu'ils fournissent. Avec ce projet de recherche, nous élaborons des modèles stochastiques pour les bourdons avec ou sans mémoire spatiale, afin d'étudier les services de pollinisation qui en découlent. Les aspects de la mémoire dans les modèles de mouvement et les services de pollinisation ont fait l'objet de recherches séparément, mais jamais en combinaison. Dans ce projet, le mouvement est modelé à l'aide d'une combinaison de trajets aléatoires corrélés et non corrélés, dans un modèle en mode agent inspiré d'équations différentielles partielles et enrichi par de récentes études de surveillance par radar des bourdons. Les services de pollinisation sont quantifiés par le taux de butinage des fleurs, calculé selon les déplacements des bourdons dans la nature. Les résultats de ce projet de recherche serviront à de futurs travaux de modélisation sur les aménagements paysagers nécessaires à une pollinisation optimale des récoltes de bleuets par les bourdons.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Hensley Hubert Mariathas (Memorial University of Newfoundland), **Shabnam Asghari** (Memorial University of Newfoundland), **Alvin Simms** (Memorial University of Newfoundland), **Masoud Mahdavian** (Memorial University of Newfoundland), **Oliver Hurley** (Memorial University of Newfoundland)

Spatial Mixed Model for Binary Response with an Application to Single Low-Density Lipoprotein Dyslipidemia

Modèle spatial mixte pour réponse binaire avec une application à la dyslipidémie de lipoprotéines de basse densité

Poster Session Séance d'affichage

Low-density lipoprotein (LDL) dyslipidemia, $LDL \geq 3.4$ mmol/L, is one of the leading factors contributing to cardiovascular diseases. Previous studies suggest regional variations in LDL-dyslipidemia; however, random effect due to geography could cause overdispersion in binary responses. To investigate the variability of spatial random effect and how factors impact the binary response, we fitted spatial mixed model to LDL dyslipidemia, a binary response from the individuals in local regions ($n=80$) in Newfoundland and Labrador (NL). Using the data from Laboratory Information System in 2014, among adults who had a LDL test ($N=170,506$), 32 % had LDL-dyslipidemia. Using method of moments, we are assessing the variability of the random component. The responses of the individuals of a given geographical region could be correlated through a common random effect shared by the individuals. Spatial mixed model is useful by obtaining consistent and efficient estimates of the parameters of the model.

La dyslipidémie de lipoprotéines de basse densité (LBD), $LBD \geq 3,4$ mmol / L, est l'un des principaux facteurs contribuant aux maladies cardiovasculaires. Des études antérieures suggèrent des variations régionales de la dyslipidémie des LBD. Cependant, l'effet aléatoire dû à la géographie pourrait provoquer de la surdispersion dans les réponses binaires. Pour étudier la variabilité de l'effet aléatoire spatial et la façon dont les facteurs influent la réponse binaire, nous avons ajusté un modèle spatial mixte à la dyslipidémie LBD, une réponse binaire des individus dans les régions locales ($n = 80$) à Terre-Neuve-et-Labrador. En utilisant les données du système d'information de laboratoire en 2014, chez les adultes qui avaient un test de LBD ($N = 170\,506$), 32% avaient la dyslipidémie LBD. En utilisant la méthode des moments, nous évaluons la variabilité de la composante aléatoire. Les réponses des individus d'une région géographique donnée pourraient être corrélées par un effet aléatoire commun partagé par les individus. Le modèle spatial mixte est utile pour obtenir des estimations convergentes et efficaces des paramètres du modèle.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Jingjia Chu (University of Western Ontario), **Reg Kulperger** (University of Western Ontario), **Hao Yu** (University of Western Ontario)

A Time Series Approach for the Underlying Driven Process in Stocks and its Large Sample Theory

Une approche par séries chronologiques pour le processus sous-jacent des actions et sa théorie des grands échantillons

An additive structure of multivariate GARCH type model is proposed to describe a dynamic common driven process in stocks and indices. The observable sequence is divided into two parts, a common risk term and an individual risk term, both following a GARCH type structure. The conditional volatilities of all stocks can increase dramatically together because of a sudden peak in the common volatility. We provide sufficient conditions for the strict stationarity and ergodicity of the model. All the parameters in the model are identifiable in terms of the conditional second moments under mild assumptions. Under the general assumptions we proposed, without specifying the distribution of the innovation, different initial values would lead to the same estimates asymptotically. Sufficient conditions for the strong consistency and asymptotic normality of the quasi maximum likelihood estimator (QMLE) are proposed.

Nous proposons une structure additive du modèle de type GARCH multivarié pour décrire un processus dynamique commun dans les actions et les indices. La séquence observable est divisée en deux parties, un terme de risque commun et un terme de risque individuel, tous deux suivant une structure de type GARCH. Les volatilités conditionnelles de toutes les actions peuvent augmenter considérablement ensemble en raison d'un pic soudain de la volatilité commune. Nous fournissons des conditions suffisantes pour la stationnarité stricte et l'ergodicité du modèle. Tous les paramètres du modèle sont identifiables en termes de deuxièmes moments conditionnels sous des hypothèses modérées. Sous les hypothèses générales que nous avons proposées, sans préciser la distribution de l'innovation, des valeurs initiales différentes conduiraient asymptotiquement aux mêmes estimations. Nous proposons des conditions suffisantes pour la cohérence forte et la normalité asymptotique de l'estimateur du quasi-maximum de vraisemblance (EQMV).

[Tuesday June 13/Mardi 13 juin, 12:00-17:00]

Poster Session Séance d'affichage

Bingrui (Cindy) Sun (University of Calgary), **MohanaGowri Arumugam** (University of Calgary)

A Second Chance? - A SoTL Project on Implementing Midterm Corrections in a Large Introductory-Level Statistics Course

Une seconde chance? - Un projet SoTL sur la mise en œuvre des corrections à mi-parcours dans un grand cours d'introduction à la statistique

A common practice students do towards assessments such as quizzes and/or exams is to look at the grade and file them away. In this case, the learning potential of assessments is not fully reached. Assessment corrections, which allow students a second chance to improve their performance and protect their GPA and thus reducing the D/Fail/Withdraw (DFW) rates, actually encourage students to identify their knowledge gaps and try to fill these gaps up. It is expected that by doing corrections students not only gain deeper understanding of the course materials but also accumulate confidence in their learning skills. The main goal of this poster presentation is to disseminate the research findings of a scholarship of teaching and learning (SoTL) project which aimed to examine the impact of midterm corrections on student learning in a large service course setup.

Une pratique courante des étudiants face à des évaluations telles que des quiz et/ou des examens est d'examiner la note et de les classer. Dans ce cas, le potentiel d'apprentissage des évaluations n'est pas entièrement atteint. Les corrections des évaluations, qui fournissent aux étudiants une seconde chance d'améliorer leur rendement et de protéger leur moyenne et de réduire ainsi les taux d'abandon/échec/désistement, encouragent les élèves à identifier les lacunes dans leurs connaissances et à tenter de combler ces lacunes. On s'attend à ce qu'en faisant des corrections, les étudiants non seulement acquièrent une compréhension plus profonde des contenus du cours, mais aient aussi davantage confiance en leurs compétences d'apprentissage. L'objectif principal de cette présentation sur affiche est de diffuser les résultats de la recherche d'un projet de bourse en enseignement et apprentissage (SoTL) qui visait à examiner l'impact des corrections à mi-parcours sur l'apprentissage des élèves dans le cadre d'un grand cours.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Kevin Matira (McMaster University), **Paul D. McNicholas** (McMaster University)

Discriminant Analysis for Longitudinal Data

Analyse discriminante pour des données longitudinales

Various approaches for discriminant analysis of longitudinal data are investigated, with some focus on model-based approaches. The latter are typically based on the modified Cholesky decomposition of the covariance matrix in a Gaussian mixture; however, non-Gaussian mixtures are also considered. Where applicable, the Bayesian information criterion is used to select the number of components per class. The various approaches are demonstrated on real and simulated data.

Nous étudions plusieurs approches de l'analyse discriminante pour des données longitudinales, en mettant l'accent sur les approches fondées sur des modèles. Ces dernières sont généralement basées sur la décomposition modifiée de Cholesky de la matrice de covariance dans un mélange de gaussiennes; cependant, nous étudions aussi les mélange non-gaussiens. Dans la mesure du possible, nous utilisons le critère d'information bayésien pour sélectionner le nombre de composantes par classe. Nous illustrons ces diverses approches sur des données réelles et simulées.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Rajib Dey (Memorial University of Newfoundland), **Noel Cadigan** (Fisheries and Marine Institute of Memorial University of Newfoundland), **Taraneh Abarin** (Memorial University of Newfoundland)

Sensitivity to Model Misspecification of a Von Bertalanffy Growth Model with Measurement Error in Age

Sensibilité à l'erreur de spécification d'un modèle de croissance de Von Bertalanffy avec erreur de mesure sur l'âge

The Von Bertalanffy growth function (VoB) specifies the length of fish as a function of age. However, in practice age is measured with error. We

La fonction de croissance de Von Bertalanffy (VoB) définit la longueur du poisson en fonction de l'âge. Cependant, l'âge est mesuré avec erreur dans la pratique. Nous étudions l'ap-

Poster Session Séance d'affichage

study the structural errors-in-variables (SEV) approach to account for ageing error. Recent research has proposed this approach for fish growth data. They assumed the distribution of true unobserved age was a Gamma distribution. They showed that this particular SEV approach provided improved regression parameter inferences compared to the standard nonlinear estimation approach that does not account for covariate measurement error (ME). In this presentation we investigate whether SEV VoB parameter estimators are robust to misspecification of the gamma true age distribution. By robust we mean no large sample bias. We outline a numerical method for evaluating the large sample bias in SEV VoB parameter estimators. Our simulation results demonstrate that the SEV VoB using a gamma distribution for true age is not robust. When ME in age is small, the misspecification bias is low. However, for large ME the bias of estimators may be large.

proche structurelle d'erreurs dans les variables (SEV) pour tenir compte de l'erreur sur l'âge. Des recherches récentes ont proposé cette approche pour les données sur la croissance des poissons. Elles ont supposé que la distribution de l'âge réel non observé était une distribution Gamma. Elles ont montré que cette approche SEV produisait des inférences de paramètres de régression améliorées par rapport à l'approche d'estimation standard non linéaire qui ne tient pas compte de l'erreur de mesure (EM) de covariable. Dans cette présentation, nous étudions si les estimateurs des paramètres SEV VoB sont robustes à une erreur de spécification de la distribution gamma de l'âge réel. Par robuste, nous entendons sans biais pour des échantillons de grande taille. Nous décrivons une méthode numérique pour évaluer le biais pour des échantillons de grande taille dans les estimateurs des paramètres SEV VoB. Les résultats de nos simulations démontrent que le SEV VoB utilisant une distribution gamma pour l'âge réel n'est pas robuste. Lorsque l'EM sur l'âge est faible, le biais attribué à la mauvaise spécification est faible. Cependant, lorsque l'EM sur l'âge est substantielle, le biais des estimateurs peut être important.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Sahar Arshadi (Memorial University of Newfoundland), **Taraneh Abarin** (Memorial University of Newfoundland)

On Identifiability of Regression Models with Interactions

Identifiabilité des modèles de régression avec interactions

Before performing any statistical inference on regression model parameters, we need to know whether the parameters of interest are “estimable” or “identifiable”. A model is said to be identifiable if all the unknown parameters of the model can be estimated uniquely provided data. We consider different interaction models, both with and without measurement error. We will look at different measurement error models such as the Berkson and classic models, and apply some remedies for non-identifiability such as instrumental variables as well as replicated and validated data.

Avant d'effectuer une inférence statistique sur les paramètres de modèles de régression, nous devons savoir si les paramètres d'intérêt sont « estimables » ou « identifiables ». Un modèle est dit identifiable si tous les paramètres inconnus du modèle peuvent être estimés de façon unique à partir des données fournies. Nous considérons différents modèles d'interaction, avec et sans erreur de mesure. Nous étudierons différents modèles d'erreur de mesure tels que celui de Berkson et les modèles classiques, et nous appliquerons des remèdes à la non-identifiabilité tels que les variables instrumentales ainsi que les données répliquées et validées.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Tyler Roick (McMaster University), **Dimitris Karlis** (Athens University), **Paul D. McNicholas** (McMaster University)

Clustering Discrete Valued Time Series

Regrouper des séries chronologiques à valeurs discrètes

There exists a need for the development of models that are able to account for discreteness in data, along with its time series properties and correla-

Le développement de modèles capables de rendre compte du caractère discret des données, ainsi que des propriétés et de la corrélation de la série chronologique, est nécessaire. Nous

Poster Session Séance d'affichage

tion. A review of the application of thinning operators to adapt the ARMA recursion to the integer-valued case is first discussed. A class of integer-valued ARMA (INARMA) models arises from this application. Our focus falls on INteger-valued AutoRegressive (INAR) type models. The INAR type models can be used in conjunction with existing model-based clustering techniques to cluster discrete valued time series data. This approach is then illustrated with the addition of autocorrelations. With the use of a finite mixture model, several existing techniques such as the selection of the number of clusters, estimation using expectation-maximization and model selection are applicable. The proposed model is then demonstrated on real data to illustrate its clustering applications.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Prageeth Manujaya Senadeera (Memorial University of Newfoundland), **Noel Cadigan** (Marine Institute of Memorial University of Newfoundland)

Local Influence Analysis of a State-Space Fish Stock Assessment Model Compared to Conventional Modelling Approaches

Analyse d'influence locale d'un modèle d'évaluation des stocks de poissons d'espace-état par rapport aux approches de modélisation conventionnelles

Surplus production models provide simple analytical methods of assessing fish populations by taking the annual biomass, the growth rate and the carrying capacity into account. However, these simple models may not adequately reflect fish stock dynamics that can be substantially more complex with age and length specific birth, growth, and death processes at play. To account for this, process errors can be included in the production model in a state-space modelling framework. In this paper, we compare the sensitivity of estimators of state-space and conventional nonlinear production models (without process errors) using the local influence analysis method introduced by R.D. Cook (1986). We apply these diagnostics to different fish stocks to assess how estimated parameters respond to small perturbations of the data. Our diagnostics reveal that the conventional model parameter estimators are more sensitive to small changes in the input data compared to the state space model estimators.

abordons d'abord l'examen de l'application des opérateurs d'amincissement pour adapter la récursion ARMA au cas de valeurs entières. Une classe de modèles à valeurs entières ARMA (INARMA) résulte de cette application. Nous nous concentrons sur des modèles de type autorégressif à valeurs entières (INAR). Les modèles de type INAR peuvent être utilisés conjointement avec les techniques existantes de regroupement basées sur le modèle pour regrouper des données de séries chronologiques à valeurs discrètes. Nous illustrons cette approche en ajoutant des autocorrélations. Avec l'utilisation d'un modèle de mélange fini, plusieurs techniques existantes telles que la sélection du nombre de grappes, l'estimation à l'aide de l'espérance-maximisation et la sélection du modèle s'appliquent. Nous illustrons les applications de regroupement du modèle proposé à l'aide de données réelles.

Les modèles de production excédentaire fournissent des méthodes analytiques simples pour évaluer les populations de poissons en tenant compte de la biomasse annuelle, du taux de croissance et de la capacité de charge. Cependant, ces modèles simples peuvent ne pas refléter adéquatement la dynamique du stock de poisson qui peut être beaucoup plus complexe avec des processus de naissance, de croissance et de mort spécifiques à l'âge et à la longueur. Pour en tenir compte, les erreurs de processus peuvent être incluses dans le modèle de production dans un cadre de modélisation d'espace-état. Dans cet article, nous comparons la sensibilité des estimateurs des modèles d'espace-état et de production non linéaires conventionnels (sans erreur de processus) en utilisant la méthode d'analyse d'influence locale introduite par R.D. Cook (1986). Nous appliquons ces diagnostics à différents stocks de poissons pour évaluer comment les paramètres estimés réagissent à de petites perturbations des données. Nos diagnostics révèlent que les estimateurs de paramètres des modèles classiques sont plus sensibles à de petits changements dans les données d'entrée par rapport aux estimateurs du modèle d'espace-état.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Jiajia Yue (Memorial University of Newfoundland), **Candemir Cigsar** (Memorial University of Newfoundland)

A Power Comparison of Robust Tests for Monotone Trend in Recurrent Event Processes

Poster Session Séance d'affichage

Comparaison de la puissance de tests robustes de la tendance monotone dans des processus d'événements récurrents

The analysis of the existence and form of time trends in repairable systems is an important issue in reliability studies. Hence, many trend tests have been proposed and studied in the literature. There has been a recent interest in developing robust trend tests based on the robust estimating function approach. These tests are appealing because they are powerful in a range of settings. In this study, we consider monotone time trends in recurrent event data from repairable systems, and develop a robust trend test based on rate functions of the power law process. Our main goal is to discuss the power of robust trend tests as well as to compare their power with other well-known trend tests under various settings. We conduct extensive Monte Carlo simulations to compute and compare the power of these tests under various cases. Finally, we analyze a data set from industry to illustrate the methodology.

L'analyse de l'existence et de la forme des tendances chronologiques dans des systèmes réparables est un aspect important dans les études de fiabilité. Ainsi, de nombreux tests de tendance ont été proposés et utilisés dans la littérature. Récemment un intérêt s'est exprimé pour l'élaboration de tests de tendance robustes basés sur l'approche des fonctions d'estimation robustes. Ces tests sont intéressants, car ils sont puissants dans différents contextes. Dans le cadre de cette étude, nous considérons les tendances de temps monotones dans les données d'événements récurrents provenant de systèmes réparables, et nous élaborons un test de tendance robuste selon les fonctions de taux d'un processus de loi de puissance. Nous examinons dans différents contextes la puissance des tests de tendance robustes et comparons leur puissance avec celle d'autres tests de tendance bien connus. Nous effectuons plusieurs simulations Monte-Carlo pour calculer et comparer la puissance de ces tests dans les différents contextes. Enfin, nous analysons un jeu de données provenant de l'industrie pour illustrer la méthodologie.

[Tuesday June 13/Mardi 13 juin, 12:00-17:30]

Sarah Ricciuti (McMaster University), **Paul D. McNicholas** (McMaster University)

Mixture Model Averaging for Clustering

Moyenne de modèles de mélange pour du regroupement

Cluster analysis is commonly described as the classification of unlabeled observations into groups such that they are more similar to one another than to observations in other groups. Model-based clustering assumes that data arise from a statistical (mixture) model. In most clustering applications, it is common to fit several models in a family and only report the results from the 'best' model chosen using a selection criterion, often the Bayesian information criterion. Recent interest has been placed on selecting a subset of solutions that are close to the 'best' model and averaging these to create a weighted averaging of clustering results. Two averaging approaches are explored, and both use Occam's window to select models that are in some sense close to the best one. In one of the methods the a posteriori probabilities are averaged and in the other method the models' parameters are averaged to generate a single interpretable model. The efficacy of these model-based averaging approaches are demonstrated for a family of skew-t mixture models using real and simulated data.

L'analyse de regroupement est communément décrite comme étant une classification en groupes d'observations non marquées de telle sorte que les observations sont plus similaires au sein du même groupe qu'avec celles dans d'autres groupes. Le regroupement basé sur un modèle suppose que les données proviennent d'un modèle (de mélange) statistique. Dans la plupart des applications de regroupement, il est courant d'ajuster plusieurs modèles d'une famille et de présenter seulement les résultats du « meilleur » modèle choisi à l'aide d'un critère de sélection, habituellement le critère d'information bayésien. Récemment, l'accent a été mis sur la sélection d'un sous ensemble de solutions qui sont proches du « meilleur » modèle et sur la moyenne de ces solutions pour créer une moyenne pondérée de ces résultats. Deux approches aux calculs de moyennes sont explorées et les deux utilisent la fenêtre d'Occam pour sélectionner les modèles qui sont en quelque sorte près du meilleur. L'un des modèles calcule la moyenne des probabilités a posteriori et l'autre modèle calcule la moyenne des paramètres du modèle pour produire un unique modèle. L'efficacité de ces approches de calcul de moyenne basée sur un modèle est démontrée pour une famille de modèles de mélange t-asymétriques à l'aide de données

Poster Session
Séance d'affichage

réelles et simulées.

Statistical Modelling in General Insurance

Modélisation statistique en assurance générale

Chair/Président: Taehan Bae
Organizer/Responsable: Taehan Bae
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-14:00]

Lysa Porth (University of Manitoba), **Ken Seng Tan** (University of Waterloo), **Wenjun Zhu** (Nankai University)

A Relational Model for Predicting Farm-Level Crop Yield Distributions in the Absence of Farm-Level Data

Un modèle relationnel pour la prédiction des répartitions du rendement des cultures au niveau des fermes en l'absence de données agricoles

Central to designing and delivering an individual crop insurance program are historical individual farm-level yields, which serve as the foundation for setting coverage levels and rates. However, data scarcity and credibility, particularly lack of farm-level yield data, make it difficult to calculate individual expected losses. As a result, aggregate county-level data are often used to establish a base premium rate, and this may contribute to adverse selection, and thus, program losses. I develop a new relational model to predict farm-level crop yield distributions in the absence of farm-level yield data, to improve the accuracy of computing crop insurance premium rates. The relational model developed defines a similarity measure based on a Euclidean distance metric to select an optimal county from a reference country, from which farm-level yield data are "borrowed". An empirical analysis shows that the relational model achieves lower mean and standard deviation prediction errors compared to the benchmark model, and is able to recover the actual premium rates more closely.

Les rendements historiques des exploitations agricoles sont au centre de la conception et de la prestation d'un programme d'assurance récolte individuel et ils constituent la base pour établir les niveaux et les taux de couverture. Par contre, la rareté et la crédibilité des données, principalement le manque de données sur le rendement des exploitations agricoles, rendent difficile le calcul des pertes individuelles attendues. Par conséquent, des données cumulées au niveau des pays sont souvent utilisées pour établir une prime de base, ce qui peut contribuer à des sélections adverses et ainsi à des pertes pour le programme. J'ai développé un nouveau modèle relationnel pour prédire les distributions des rendements des exploitations agricoles en l'absence de données sur ces rendements pour améliorer la précision du calcul des primes d'assurance récolte. Ce modèle relationnel définit une mesure de similitude fondée sur une distance euclidienne pour sélectionner un comté optimal comme comté de référence d'où des données de rendements seront « empruntées ». Une analyse empirique démontre que le modèle relationnel obtient des erreurs de prédiction inférieures au modèle de référence au niveau de la moyenne et de l'écart-type et il se rapproche plus étroitement de la prime réelle.

[Tuesday June 13/Mardi 13 juin, 14:00-14:30]

Lei Hua, Jianxi Su (Purdue University)

Full-Range Tail Dependence Copulas

Étendue complète de la dépendance des ailes des copules

In this talk, I will introduce some flexible new bivariate copulas, the R package CopulaOne, and a few applications for data analytics in insurance and finance. Popular multivariate copulas such as vine copulas and factor copulas are constructed based on bivariate copulas. An ideal bivariate copula should have the following features. First, both upper and lower tails are able to explain full-range tail dependence. That is, the dependence in each tail can

Lors de cet exposé, je présenterai des nouvelles copules bivariées flexibles, le paquet R CopulaOne et quelques applications à l'analyse de données dans le domaine des assurances et de la finance. Les copules multivariées les plus populaires, telles que les copules en vigne et les copules de facteurs, sont construites à partir de copules bivariées. La copule bivariée idéale devrait posséder les caractéristiques suivantes. Premièrement, les ailes supérieures et inférieures expliquent l'étendue complète de la dépendance des ailes.

Statistical Modelling in General Insurance Modélisation statistique en assurance générale

range among quadrant tail independence, intermediate tail dependence, and usual tail dependence. Second, it can capture upper and lower tail dependence patterns that are either the same or different. In this talk, I will discuss a general approach for constructing copulas that have the above features. Some promising parametric copula families are to be presented, and both the ideal features and the computational speeds were considered when constructing the copulas. Finally, a few applications using the copulas are to be demonstrated.

C'est-à-dire, la dépendance de chaque aile peut varier entre la dépendance d'aile quadrant, la dépendance d'aile intermédiaire et la dépendance d'aile habituelle. Deuxièmement, la copule peut capturer des patrons de la dépendance des ailes supérieures et inférieures qui sont soit identiques ou différents. Dans cet exposé, je discuterai d'une approche générale pour construire des copules qui possèdent les caractéristiques susmentionnées. Quelques familles de copules paramétriques prometteuses seront présentées. Les caractéristiques idéales ainsi que les vitesses computationnelles ont été considérées lors de la construction des copules. Finalement, quelques applications de l'utilisation des copules seront démontrées.

[Tuesday June 13/Mardi 13 juin, 14:30-15:00]

Di (Cindy) Xu (University of Nebraska-Lincoln), **David Landriault** (University of Waterloo), **Bin Li** (University of Waterloo)

Aggregate Claim Analysis in a Two-sided Exit Setting with Dependence

Analyse de réclamations agrégées dans un contexte de sortie bilatéral avec dépendance

The two-sided exit problem has been the subject of risk management analysis, used to better understand the dynamic of various insurance risk processes. In the two-sided exit setting, the discounted aggregate claims are investigated under a dependent renewal process (also known as dependent Sparre Andersen risk process). Utilizing Lundberg's generalized equation and Laplace transform, we identify the fundamental solutions to a given integral equation, which will be shown to play a role similar to the scale matrix for spectrally-negative Markov-additive processes. Explicit expressions and recursions are then identified for the two-sided probabilities and the moments of the aggregate claims respectively. A numerical example for the two-sided exit probabilities involving the Farlie-Gumbel-Morgenstern (FGM) copula is provided.

La question de la sortie bilatérale a fait l'objet d'analyses de gestion du risque qui sont utilisées pour mieux comprendre la dynamique de différents processus d'assurance risque. Dans le contexte de la sortie bilatérale, les réclamations agrégées réduites sont étudiées par un processus de renouvellement dépendant (aussi connu sous le nom de processus de risque dépendant Sparre Andersen). Nous identifions les solutions fondamentales d'une équation intégrale donnée en utilisant l'équation généralisée de Lundberg et la transformation de Laplace et il sera démontré qu'elles jouent un rôle similaire à celui de la matrice d'échelle pour les processus additifs de Markov aux spectres négatifs. Des expressions et des récursions explicites sont alors identifiées pour, respectivement, les probabilités bilatérales et les moments des réclamations agrégées. Un exemple numérique des probabilités de sortie bilatérales comprenant la copule Farlie-Gumbel-Morgenstern (FGM) est présenté.

New Directions in Quantitative Finance Nouvelles directions en finance quantitative

Chair/Président: Adam Metzler
Organizer/Responsable: Adam Metzler
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-14:15]

Ralf Kellner (University of Regensburg)

Improving Risk Assessment using Market Expectations and Information from Model Risks

Améliorer l'évaluation du risque à l'aide des renseignements et des attentes du marché provenant des modèles de risque

Market Risk is usually assessed by means of point estimates for risk measures which are exposed to randomness in data and/or misspecification of underlying model assumptions. In this paper, we extend the view of risk measurement on the distribution of estimators for risk measure forecasts to account for possible model risk. In addition, we analyze to which degree unexpected economic losses may be reduced when using information from the risk measure estimator distribution and combine it with future market expectations. The idea behind this approach is that more conservative risk measure forecasts than the point estimate of the estimator's distribution should be used when market expectations point to times of financial turmoil.

Le risque du marché est généralement évalué au moyen d'estimations ponctuelles des mesures du risque exposées au caractère aléatoire des données et/ou à l'imprécision des hypothèses sous-jacentes au modèle. Cet article étend la notion de mesure du risque à la distribution des estimateurs en vue d'obtenir des prévisions de mesure du risque qui rendent compte de divers modèles de risque. Nous y analysons aussi comment les pertes économiques peuvent être réduites en utilisant l'information tirée de la distribution des estimateurs de la mesure du risque et en la combinant aux futures attentes du marché. Cette approche repose sur l'idée que des prévisions de mesure du risque plus conservatrices que les estimations ponctuelles de la distribution des estimateurs devraient être utilisées lorsque les attentes du marché pointent vers une période de tourmente financière.

[Tuesday June 13/Mardi 13 juin, 14:15-15:00]

Adam Metzler (Wilfrid Laurier University)

Stochastic Correlation and Regulatory Capital

Corrélation stochastique et capital réglementaire

Current banking regulations require banks to use a so-called risk-weight function to determine the amount of capital that is held as a buffer against extraordinary losses. The function is based on a simple and intuitive model developed by Vasicek (2002), which makes several strong simplifying assumptions but retains its popularity due to its intuitive appeal (workarounds to some of the shortcomings caused by these assumptions have been proposed in the literature and adopted in practice). One of these assumptions is that correlations do not vary with the business cycle, which is in stark contrast to empirical evidence. In this talk we generalize the Vasicek model to allow for correlations that vary systematically with the state of the economy. The model includes a parameter that allows the user to control the degree to which correlations

La réglementation bancaire en vigueur exige des banques qu'elles utilisent une fonction de soi-disant pondération des risques afin de déterminer le montant du capital qui servira d'amortissement en cas de lourdes pertes financières. Cette fonction repose sur un modèle intuitif simple développé par Vasicek (2002). Ce modèle fondé sur plusieurs hypothèses simplificatrices audacieuses est resté populaire en raison de son caractère intuitif attrayant (des solutions de rechange à certaines lacunes causées par ces hypothèses ont été proposées dans la documentation et adoptées en pratique). L'une de ces hypothèses veut que les corrélations ne varient pas selon les cycles d'affaires, ce qui est en contradiction flagrante avec les données empiriques. Cet exposé propose une généralisation du modèle de Vasicek permettant des corrélations qui varient systématiquement en fonction de l'état de l'économie. Le modèle comprend un paramètre qui permet à l'utilisateur de contrôler le degré de hausse des corrélations selon divers

New Directions in Quantitative Finance

Nouvelles directions en finance quantitative

to rise during adverse economic scenarios, and we demonstrate that this parameter has a profound impact on regulatory capital calculations.

scénarios de conditions économiques défavorables. Nous montrons que ce paramètre a un impact marqué sur le calcul du capital réglementaire.

Recent Developments in Small Area Estimation Développements récents en estimation pour petits domaines

Chair/Président: Mahmoud Torabi
Organizer/Responsable: Mahmoud Torabi
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-14:00]

Sanjoy Kumar Sinha (Carleton University)

Robust Small Area Estimation in Generalized Linear Models

Estimation robuste sur petits domaines dans des modèles linéaires généralisés

In this talk, I will present novel methods for small area estimation when the outcome variable is discrete. The methods developed in the framework of the generalized linear mixed models for clustered correlated data will be robust against slight deviations from the underlying distributions for both outcome and auxiliary variables. The empirical properties of the proposed estimators will be studied using Monte Carlo simulations. An application will also be provided using actual survey data.

Dans cette allocution, je présenterai de nouvelles méthodes pour l'estimation sur petits domaines lorsque la variable réponse est discrète. Les méthodes élaborées dans le cadre des modèles mixtes linéaires généralisés pour des données corrélées en grappes seront robustes par rapport aux légères déviations provenant des distributions sous-jacentes des variables réponses et auxiliaires. Nous étudierons les propriétés empiriques des estimateurs proposés au moyen de simulations Monte-Carlo et présenterons son application au moyen de données d'enquête réelles.

[Tuesday June 13/Mardi 13 juin, 14:00-14:30]

Snigdhasu Chatterjee, **Kaibo Gong** (University of Minnesota), **Megan Heyman** (Rose-Hulman Inst of Technology), **Taps Maiti** (Michigan State University)

Small Area Model Selection

Sélection de modèles sur petits domaines

Complex statistical models that have multiple sources of dependencies and variability in the observations are of primary importance in studying data from multiple disciplines. These include spatial, temporal, spatio-temporal, various mixed effects and other statistical models. Of special importance among such models are those that are useful for studying problems where there is limited directly observed data, for example, as in small area models. In this talk we present a new resampling-based method that can be used for simultaneous variable selection and inference in several complex models, including small area and other mixed effects models. Theoretical results justifying the proposed resampling schemes will be presented, followed by simulations and real data examples. This talk is based on research involving several students and collaborators from multiple institutions.

Les modèles statistiques complexes qui ont plusieurs sources de dépendances et de variabilité dans les observations revêtent une importance primordiale dans l'étude de données provenant de plusieurs disciplines. Parmi eux on retrouve les modèles spatiaux, temporels, spatio-temporels, à effets mixtes variés, et d'autres modèles statistiques. Parmi ces modèles, certains sont particulièrement importants pour étudier les problèmes où le nombre de données qui sont observées directement est limité, comme dans des modèles sur petits domaines. Dans cet article, nous présentons une nouvelle méthode axée sur le rééchantillonnage qui peut être utilisée pour la sélection de variables simultanées et pour l'inférence dans divers modèles complexes, notamment les modèles sur petits domaines et à effets mixtes. Nous présenterons les résultats théoriques justifiant les plans de rééchantillonnage proposés, ainsi que des exemples de données de simulation et réelles. Cet exposé se fonde sur la recherche de plusieurs étudiants et collaborateurs provenant de plusieurs institutions.

[Tuesday June 13/Mardi 13 juin, 14:30-15:00]

Gauri Datta (U.S. Bureau of the Census), **Abhyuday Mandal** (University of Georgia), **Adrijo Chakraborty** (NORC)

Recent Developments in Small Area Estimation

Développements récents en estimation pour petits domaines

Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model

Estimation bayésienne hiérarchique robuste pour petites régions pour modèle de régression à erreur emboîtée

National statistical offices are mandated to produce reliable statistics for important characteristics for many sub-populations or small areas, defined by geography and/or demography. Model-based methods that “borrow strength” from other areas and variables are extensively used to generate reliable statistics. Standard model-based small area estimates perform poorly in presence of outliers. Sinha and Rao (2009) proposed a robust frequentist approach to handle outliers. We propose a robust hierarchical Bayes (HB) method to handle outliers in unit-level data. We consider a two-component scale mixture of normal distributions for the unit error to model outliers and produce noninformative HB predictors of small area means. An example and extensive simulations convincingly show robustness of HB predictors. Simulation evaluation of these two procedures shows their superiority over the M-quantile small area estimators. Our HB procedure enjoys dual (Bayes and frequentist) dominance.

Les services nationaux de statistique sont chargés de produire des statistiques fiables concernant d’importantes caractéristiques pour de nombreuses sous-populations ou petites régions, définies par la géographie et/ou la démographie. On utilise souvent des méthodes basées sur des modèles qui « exploitent la force » d’autres régions et variables afin de produire des statistiques fiables. Mais les estimations pour petites régions usuelles basées sur des modèles donnent de mauvais résultats en la présence de données aberrantes. Sinha et Rao (2009) proposent une approche fréquentiste robuste pour traiter les données aberrantes. Nous proposons une méthode bayésienne hiérarchique (BH) robuste pour traiter les données aberrantes au niveau des unités. Nous étudions un mélange d’échelles à deux composants de distributions normales pour l’erreur au niveau des unités qui modélise les données aberrantes et produit des prédicteurs BH non informatifs des moyennes pour petites régions. Un exemple et de nombreuses simulations prouvent de manière convaincante la robustesse des prédicteurs BH. Une évaluation par simulation de ces deux procédures montre leur supériorité par rapport aux estimateurs de quantile M pour petites régions. Notre procédure BH bénéficie d’une double dominance (bayésienne et fréquentiste).

Improving Statistics at CIHR Comment améliorer la statistique aux IRSC

Chair/Président: Lawrence McCandless
Organizer/Responsable: Lawrence McCandless
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-15:00]

Lawrence C. McCandless (Simon Fraser University), **Erica E.M. Moodie** (McGill University), **Russell Steele** (McGill University), **David A. Stephens** (McGill University), **Nancy Reid** (University of Toronto)

Panel discussion

Améliorer les statistiques dans les Instituts de recherche en santé du Canada (discussion en groupe)

In the most recent 2015 Transitional Operating Grants Competition, only two grants worth \$808,719 out of 391 grants worth \$250.1 million were awarded to projects to investigate biostatistics or statistical methodology. The results of the CIHR-NSERC Collaborative Health Research Program (CHRP) competitions have been similar: among 106 CHRP grants over three years (2013-15) worth \$32.2 million, there were two grants awarded to support biostatistics research, which were collectively worth \$582,179. This panel discussion session will assemble Canadian biostatisticians who hold operating grants from CIHR in order to discuss national strategies wherein Canadian biostatisticians can secure more CIHR funding for biostatistics research. The themes for the session will include 1) Strategies for successfully applying to CIHR; 2) the CIHR-NSERC collaborative health research projects program; and 3) potential leadership roles for CANSSI and the SSC.

Dans le plus récent concours transitoire de subvention de fonctionnement, seulement deux subventions valant au total 808 719 \$ des 391 subventions d'une enveloppe de 250,1 millions \$ ont été remises dans le cadre de projets de recherche sur la méthodologie biostatistique ou statistique. Les résultats des concours dans le cadre du programme de recherche concertée sur la santé des Instituts de recherche en santé du Canada (IRSC) et du Conseil de recherches en sciences naturelles et en génie (CRSNG) ont été semblables : parmi 106 subventions du CRSNG sur trois ans (2013-2015) s'élevant à 32,2 millions, deux subventions ont été remises pour soutenir la recherche biostatistique, et se sont ensemble élevées à 582 179 \$. Cette séance de discussion en groupe réunira les biostatisticiens canadiens qui détiennent des subventions de fonctionnement de l'IRSC et permettra de discuter des stratégies nationales grâce auxquelles les biostatisticiens canadiens peuvent obtenir plus de financement de la part de l'IRSC pour la recherche biostatistique. Les sujets de cette séance seront les suivants : 1) stratégies pour réussir à obtenir du financement l'IRSC ; 2) programme projets de recherche concertée sur la santé de l'IRSC et du CRSNG ; 3) rôles de leadership potentiels de l'Institut canadien de sciences statistiques et de la Société statistique du Canada.

Design and Methodological Issues in Registry-based Randomized Controlled Trials Problèmes de conception et de méthodologie dans les essais randomisés contrôlés sur registres

Chair/Président: Robert W. Platt
Organizer/Responsable: Tolulope T. Sajobi
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-14:00]

Lehana Thabane (McMaster University)

The Randomized Registry Trial — Opportunities, Challenges and Solutions

Essais sur registres randomisés : possibilités, défis et solutions

Registry-based randomized controlled trials (RRCTs) are defined as pragmatic trials that use registries as a platform for case records, data collection, randomization and follow-up. Recently the application of RRCTs has attracted increasing attention in health research to address comparative effectiveness research questions in real-world settings, mainly due to their low cost, enhanced generalizability of findings, rapid consecutive enrollment, and the potential completeness of follow-up for the reference population, when compared with conventional randomized effectiveness trials. However, several challenges of RRCTs have to be taken into consideration, including registry data quality, ethical issues and methodological challenges. In this talk, I will summarize the advantages and challenges and areas for future research related to RRCTs.

Les essais randomisés contrôlés sur les registres (ERCR) sont des essais pragmatiques qui utilisent des registres comme plateforme pour les dossiers, la collecte de données, la randomisation et le suivi. Récemment, l'application des ERCR a suscité un intérêt accru dans la recherche en santé car ils permettent de répondre à des questions de recherche relatives à l'efficacité comparative dans des contextes réels et qu'ils présentent les avantages suivants par rapport aux essais d'efficacité randomisés traditionnels : faible coût, généralisabilité accrue de leurs résultats, recrutement rapide et le potentiel de suivi de la population de référence. Cependant, il faut prendre en compte les divers défis que posent les ERCR, notamment en ce qui a trait à la qualité des données de registre, de certains problèmes éthiques ainsi que les défis méthodologiques. Dans cette présentation, je résume les avantages, les défis et les pistes de recherche futures pour les ERCR.

[Tuesday June 13/Mardi 13 juin, 14:00-14:30]

Kevin E. Thorpe (University of Toronto)

Registry-Based Randomized Controlled Trials: The Case for Explanatory-Pragmatic Continuum

Essais randomisés sur registres : le continuum explicatif-pragmatique

Randomized controlled trials (RCTs) have been broadly categorized as either having a pragmatic or explanatory attitude. Pragmatic trials are designed to evaluate the effectiveness of interventions in real-life routine practice conditions, whereas explanatory trials aim to test whether an intervention works under optimal situations. In this talk, I will discuss the development of a Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) for classifying trials and its application to registry-based trials.

Les essais randomisés contrôlés (ERC) sont généralement catégorisés comme étant pragmatiques ou explicatifs. Les essais pragmatiques sont conçus pour évaluer l'efficacité d'interventions dans les conditions de la pratique routinière réelle, tandis que les essais explicatifs visent à déterminer si une intervention fonctionne dans une situation optimale. Dans cette présentation, je discute de la mise au point d'un indicateur de résumé du continuum explicatif-pragmatique (Pragmatic-Explanatory Continuum Indicator Summary, ou PRECIS) pour classer les essais, ainsi que de son application aux essais sur registres.

[Tuesday June 13/Mardi 13 juin, 14:30-15:00]

Tolulope T. Sajobi (University of Calgary), **Oluwagbohunmi Awosoga** (University of Lethbridge), **Bijoy Menon** (University of Calgary), **Michael Hill** (University of Calgary), **Lehana Thabane** (McMaster University)

Statistical Methods for Synthesizing Evidence from Explanatory and Pragmatic trials

Design and Methodological Issues in Registry-based Randomized Controlled Trials

Problèmes de conception et de méthodologie dans les essais randomisés contrôlés sur registres

Méthodes statistiques pour la synthèse de résultats d'essais explicatifs et pragmatiques

Randomized controlled trials (RCTs) of treatments and interventions are typically described as either explanatory or pragmatic. Meta-analysis of RCT studies typically pools evidence of treatment effects from included studies, regardless of their classification as 'pragmatic' or 'explanatory' trials. Given that treatment effects in explanatory trials may be greater than those obtained in pragmatic trials, conventional meta-analytic approaches may not accurately account for the heterogeneity among the studies. Stratified meta-analysis of systematically review studies in which treatment effects from explanatory trials are meta-analyzed and reported separately from pragmatic trials is generally used in systematic reviews. In this study we investigate a variety of meta-analytic approaches for synthesizing evidence from pragmatic and explanatory trials. Discussions about the key statistical and design considerations when pooling evidence from both types of trial designs are provided.

Les essais contrôlés randomisés (ECR) de traitements et d'interventions sont généralement décrits comme explicatifs ou pragmatiques. La méta-analyse des essais ECR regroupe habituellement tous les résultats sur les effets du traitement, sans se préoccuper du fait que les essais en question soient « pragmatiques » ou « explicatifs ». Bien que dans les essais explicatifs, les effets de traitement puissent sembler supérieurs à ceux obtenus lors d'essais pragmatiques, les approches méta-analytiques traditionnelles ne tiennent pas forcément compte de l'hétérogénéité entre essais. Dans les analyses systématiques, on procède généralement à une méta-analyse stratifiée et à une présentation distincte des résultats des essais explicatifs et des essais pragmatiques. Nous étudions ici une variété d'approches méta-analytiques qui permettent une synthèse des résultats d'essais explicatifs et pragmatiques. Nous discutons des aspects statistiques et de planification à considérer lors de l'analyse de résultats de ces deux types d'essais.

Bayesian Methods and Applications

Méthodes bayésiennes et leurs applications

Chair/Président: Rob Deardon

Room/Salle: E2 165 (EITC)

Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-13:45]

Haoxuan Zhou (Simon Fraser University), **Dave Campbell** (Simon Fraser University)

A Bayesian Approach to Diagnosis of the Gaussian Random Field Assumption

Une approche bayésienne pour le diagnostic de l'hypothèse de champ aléatoire gaussien

The Latent Gaussian Model (LDM) is used in various fields such as ecology, the study of cancer, and the stock market. The model is usually used on high dimensional data, with the Integrated Laplace Approximation (INLA) an advanced approach to inference in the LDM. Compare to the traditional approach of Monte Carlo Markov Chain, INLA has much better performance in terms of running speed and error rate. However, one main assumption of using INLA is that the data come from a Gaussian Markov random fields (GMRF). In this paper, we present a method to diagnose the normality assumption of LDM in high dimensions. This method uses Ordinary Differential Equations (ODE) with a Bayesian approach.

Le modèle latent gaussien (LDM) est utilisé dans divers domaines tels que l'écologie, l'étude du cancer et le marché boursier. Le modèle est généralement utilisé en présence de données de grande dimension, avec l'approximation intégrée de Laplace (AILA) une approche avancée d'inférence dans le LDM. L'AILA performe nettement mieux que l'approche traditionnelle des chaînes de Markov en termes de vitesse de fonctionnement et de taux d'erreur. Cependant, une hypothèse principale de l'utilisation de l'AILA est que les données proviennent d'un champ aléatoire gaussien de Markov (CAGM). Dans cet article, nous présentons une méthode pour diagnostiquer l'hypothèse de normalité de LDM en grande dimension. Cette méthode utilise des équations différentielles ordinaires (EDO) avec une approche bayésienne.

[Tuesday June 13/Mardi 13 juin, 13:45-14:00]

Thierry Chekouo Tekougang (University of Minnesota Duluth), **Francesco Stingo** (University of Florence), **Kim-Anh Do** (The University of Texas MD Anderson Cancer Center)

A Bayesian Approach for Pattern Detection in Extremely Small Sample Sizes: Application to Proteomic Cell Line Data of Human Leukemia

Une approche bayésienne pour la détection de motifs dans des tailles d'échantillon extrêmement petites : application aux données de la lignée cellulaire protéomique de la leucémie humaine

Human cancer cell line experiments are a valuable tool for investigating drug sensitivity biomarkers. The number of biomarkers measured in these experiments is typically on the order of several thousand; whereas the number of samples is often limited to one or at most three replicates for each experimental condition. We have developed an innovative Bayesian approach that efficiently identifies clusters of markers with similar patterns. We define prior distributions on cluster parameters that allow us to obtain biologically meaningful clusters and to better discriminate them. Motivated by the availability of ion mobility mass spectrometry data on cell line experiments in myelodysplastic syndromes and acute myeloid leukemia from the "Moon Shots" Program at MD Anderson Cancer Center, our method-

Les expériences sur les lignées cellulaires humaines du cancer sont un outil précieux pour l'étude de la sensibilité des biomarqueurs aux médicaments. Le nombre de biomarqueurs mesurés dans ces expériences est typiquement de l'ordre de plusieurs milliers; tandis que le nombre d'échantillons est souvent limité à un ou au maximum trois répétitions pour chaque condition expérimentale. Nous avons développé une approche bayésienne novatrice qui identifie efficacement des grappes de marqueurs avec des motifs similaires. Nous définissons des distributions a priori sur des paramètres de grappe qui permettent d'obtenir des grappes biologiquement significatives et de mieux les discriminer. Motivée par la disponibilité de données sur la spectrométrie de masse par mobilité ionique sur les expériences de lignée cellulaire dans les syndromes myélodysplasiques et la leucémie myéloïde aiguë du programme « Moon Shots » à MD Anderson Cancer Cen-

Bayesian Methods and Applications

Méthodes bayésiennes et leurs applications

ology can identify protein markers that follow biologically meaningful trends.

ter, notre méthodologie permet d'identifier des marqueurs protéiques qui suivent des tendances biologiquement significatives.

[Tuesday June 13/Mardi 13 juin, 14:00-14:15]

Tian Li (Simon Fraser University), **Lawrence C. McCandless** (Simon Fraser University), **X. Joan Hu** (Simon Fraser University), **Julian M. Somers** (Simon Fraser University)

Bayesian Sensitivity Analysis for Non-ignorable Missing Data in Longitudinal Studies

Analyse de sensibilité bayésienne pour les données manquantes non ignorables dans les études longitudinales

The use of Bayesian statistical methods to handle missing data in biomedical studies has become popular in recent years. In this paper, we propose a novel Bayesian sensitivity analysis (BSA) model that accounts for the influences of missing outcome data on the estimation of treatment effects in randomized trials with non-ignorable missing data. We implement the method using the probabilistic programming language STAN, and apply it to data from the Vancouver At Home (VAH) Study, which is a randomized control trial that provided housing to homeless people with mental illness. We compare the results of BSA to those from an existing Bayesian longitudinal model that ignores missingness in the outcome. Furthermore, we demonstrate in a simulation study that BSA credible intervals have greater length and higher coverage rate of the target parameters than existing methods.

L'utilisation de méthodes bayésiennes pour traiter les données manquantes dans les études biomédicales est devenue très populaire ces dernières années. Dans cette présentation, nous proposons un nouveau modèle d'analyse de sensibilité bayésienne (ASB) qui tient compte de l'influence des données manquantes sur l'estimation des effets du traitement dans les essais randomisés avec données manquantes non ignorables. Nous mettons en œuvre la méthode à l'aide de la langue de programmation probabiliste STAN et l'appliquons à des données de l'étude Vancouver At Home (VAH), un essai contrôlé randomisé qui fournit des logements à des sans-abri souffrant de maladie mentale. Nous comparons les résultats de l'ASB à ceux d'un modèle longitudinal bayésien existant qui néglige les données manquantes. Par ailleurs, nous montrons par une étude de simulation que les intervalles plausibles de l'ASB sont plus longs et donnent un taux de couverture des paramètres cibles supérieur aux méthodes existantes.

[Tuesday June 13/Mardi 13 juin, 14:15-14:30]

Vineetha K.V. Warriyar (University of Calgary), **Rob Deardon** (University of Calgary)

Individual Level Modelling of Infectious Disease Data: EpiILM

Modélisation au niveau individuel des données sur les maladies infectieuses : EpiILM

The statistical modelling of infectious disease spread through a population generally requires the use of non-standard statistical models. This is primarily due to the fact that infection events depend upon the infection status of other members of the population and hence we cannot assume independence of infection events. Further complication is added by the fact that there are often complex heterogeneities in the population which we wish to account for, since, for example, populations do not tend to mix homogeneously. Sometimes, we may wish to account for such heterogeneities using spatial mechanisms that assume that transmission events are more likely to occur between individuals close together in space than individuals further apart. Sometimes, it is more natural to model such

La modélisation statistique de propagation des maladies infectieuses dans une population nécessite généralement l'utilisation de modèles statistiques non standards. Cela s'explique surtout par le fait que les cas d'infection dépendent du stade d'infection des autres membres de la population. Par conséquent nous ne pouvons pas supposer l'indépendance des cas d'infection. Le fait qu'on veuille souvent prendre en compte l'hétérogénéité complexe de la population, puisque, par exemple, les populations ne semblent pas se mélanger de façon homogène, constitue une complication supplémentaire. Parfois, nous souhaitons prendre en compte ces hétérogénéités en utilisant des mécanismes spatiaux qui supposent que les cas de transmission sont plus susceptibles de se produire entre les individus proches les uns des autres dans un espace par rapport à ceux qui sont plus éloignés. De plus, il est plus naturel de modéliser ces hétérogénéités au moyen de réseaux

Bayesian Methods and Applications

Méthodes bayésiennes et leurs applications

heterogeneities using contact networks that represent, for example, the sharing of supplier companies between farms. Typically, statistical inference for these models (e.g., parameter estimation) is done in a Bayesian context using computational techniques such as Markov chain Monte Carlo (MCMC). Here we examine the main characteristics of individual level infectious disease models, and how to fit them to data within a Bayesian statistical framework using the R package "EpiILM".

[Tuesday June 13/Mardi 13 juin, 14:30-14:45]

Shaun Zheng Sun (University of the Fraser Valley), **Richard Lockhart** (Simon Fraser University)

Bayesian Optimality for Beran's Class of Tests of Uniformity around the Circle

Optimalité bayésienne pour la classe de tests d'uniformité sur le cercle de Beran

We show that the locally most powerful tests of uniformity on the circle given by Beran have optimality properties not just against the parametric alternatives described by Beran but also against many non-parametric alternatives. We describe both local and asymptotic versions of this optimality.

de contacts qui représentent par exemple le partage de fournisseurs entre fermes. En général, l'inférence statistique de ces modèles (par exemple l'estimation de paramètres) est effectuée dans un contexte bayésien au moyen de techniques de calcul, comme les méthodes de Monte-Carlo par chaîne de Markov. Dans cet exposé, nous examinons les principales caractéristiques des modèles Bayésiens de maladie infectieuse au niveau individuel à l'aide de la librairie R EpiILM.

Nous montrons que les tests d'uniformité sur le cercle les plus puissants au niveau local tels que donnés par Beran présentent des propriétés d'optimalité non seulement par rapport aux alternatives paramétriques décrites par Beran mais aussi par rapport à de nombreuses alternatives non paramétriques. Nous décrivons des versions locales et asymptotiques de cette optimalité.

[Tuesday June 13/Mardi 13 juin, 14:45-15:00]

Chel Hee Lee (University of Calgary), **Martha E. Lyon** (University of Saskatchewan), **Krista Baerg** (University of Saskatchewan)

Development of Quality Assurance Program for Jaundice Meter Management

Élaboration d'un programme d'assurance de la qualité pour la gestion des ictomètres

Jaundice meters are intended for use as a screening device in newborns. Meters need to be validated before putting them into clinical use. Transcutaneous bilirubin readings are compared to the clinical laboratory standard, serum bilirubin obtained from blood samples. Meter reliability is evaluated by a regression model where errors are in both variables since an error is not only associated with the jaundice meter method but also the clinical standard. A Bayesian approach is discussed with independent and noninformative priors. Since reliability varies from meter to meter, a number of clustering techniques are examined to develop a statistical quality assurance program for the management of multiple meters using reliability index. A simulation study is carried out to see robustness of the developed program under various conditions, including unbalanced samples size over meters, over-ranged and outlying observations, error ratio, and other factors that cannot be controlled in practice.

Les ictomètres sont destinés à être utilisés comme dispositif de dépistage chez les nouveau-nés. Ils doivent être validés avant leur utilisation clinique. Les lectures de la bilirubine transcutanée sont comparées à la norme clinique de laboratoire, la bilirubine sérique obtenue à partir des échantillons de sang. La fiabilité des ictomètres est évaluée par un modèle de régression dans lequel il existe des erreurs dans les deux variables, car une erreur est non seulement associée à la méthode de l'ictomètre, mais aussi à la norme clinique. Nous discutons une approche bayésienne avec des préalables indépendants et non informatifs. Puisque la fiabilité varie d'un ictomètre à l'autre, un certain nombre de techniques de regroupement sont examinées pour élaborer un programme statistique d'assurance de la qualité pour la gestion de nombreux ictomètres à l'aide d'un indice de fiabilité. Une étude de simulation est réalisée pour vérifier la robustesse du programme développé dans diverses conditions, y compris la taille déséquilibrée des échantillons sur les ictomètres, les observations surdimensionnées et aberrantes, le taux d'erreur et d'autres facteurs qui ne peuvent pas être contrôlés en pratique.

Complex Regression and Inference Régression complexe et inférence

Chair/Président: Mohamed Belalia
Room/Salle: E2 160 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-13:45]

Jenny Tieu (Brock University), Ramona Rat (Brock University), Mei Ling Huang (Brock University)

A New Weighted Quantile Regression

Une nouvelle régression pondérée sur les quantiles

Quantile regression has wide applications in many fields. Studies of heavy tailed distributions have rapidly developed. For multivariate heavy tailed distributions, estimation of conditional quantiles at very high or low tails is of interest in numerous applications. This presentation proposes a new weighted quantile regression method on high quantile regression. The Monte Carlo simulations show good efficiency of the proposed weighted estimator relative to regular quantile regression estimator. The presentation also investigates a real-world example by using the proposed weighted method. Comparisons of the proposed method and existing methods are given.

La régression sur les quantiles a une portée très étendue dans de nombreux domaines. Des études sur les distributions à queue épaisse se sont rapidement développées. Pour les distributions multivariées à queue épaisse, l'estimation des quantiles conditionnels à des niveaux très élevés ou bas est intéressante dans de nombreuses applications. Cette présentation propose une nouvelle méthode de régression pondérée sur les quantiles sur la régression à quantiles élevés. Les simulations de Monte-Carlo montrent une bonne efficacité de l'estimateur pondéré proposé par rapport à l'estimateur régulier de régression sur les quantiles. La présentation examine également un exemple concret en utilisant la méthode pondérée proposée. Les méthodes proposée et existantes sont comparées.

[Tuesday June 13/Mardi 13 juin, 13:45-14:00]

Shahedul A. Khan (University of Saskatchewan)

Generalized Bent-Cable Regression for Changepoint Data

Régression généralisée «bent-cable» pour les données de points de rupture

The choice of the model framework in a regression setting depends on the nature of the data. The focus of this study is on changepoint data, exhibiting three phases: incoming and outgoing, both of which are linear, joined by a curved transition. Bent-cable regression is an appealing statistical tool to characterize such trajectories, quantifying the nature of the transition between the two linear phases by modeling the transition as a quadratic phase with unknown width. We demonstrate that a quadratic function may not be appropriate to adequately describe many changepoint data. We then propose a generalization of the bent-cable model by relaxing the assumption of the quadratic bend. The properties of the generalized model are discussed and Bayesian approach for inference is proposed. This study suggests that the proposed generalization can produce a fit, which is either comparable or superior to that of the quadratic bent-cable model.

Le choix du modèle cadre dans un contexte de régression dépend de la nature des données. Cette étude porte sur des données de points de rupture, présentant trois phases : entrante et sortante, toutes deux linéaires, jointes par une transition courbe. La régression bent-cable est un outil statistique attrayant pour caractériser de telles trajectoires, quantifiant la nature de la transition entre les deux phases linéaires en modélisant la transition comme une phase quadratique de largeur inconnue. Nous démontrons qu'une fonction quadratique peut ne pas être appropriée pour décrire correctement plusieurs données de point de rupture. Nous proposons ensuite une généralisation du modèle bent-cable en relaxant l'hypothèse de la courbure quadratique. Nous discutons des propriétés du modèle généralisé et proposons une approche bayésienne pour l'inférence. Cette étude suggère que la généralisation proposée peut produire un ajustement comparable ou supérieur à celui du modèle quadratique bent-cable.

Complex Regression and Inference Régression complexe et inférence

[Tuesday June 13/Mardi 13 juin, 14:00-14:15]

Jeffrey L. Andrews (University of British Columbia)

Combatting Overfitting in Finite Mixture Models

Combattre le surajustement dans les modèles de mélanges finis

The standard expectation-maximization approach to parameter estimation for Gaussian mixture models is modified by a within-cycle incorporation of the nonparametric bootstrap in order to prevent overfitting in model-based clustering applications. Benefits include more reasonable parameter estimates (including estimates of standard error), more ‘realistic’ clustering results, and increased avoidance of local maxima. In addition, we find better efficiency versus a traditional nonparametric bootstrap approach to the EM for mixture models. Proposals for issues that arise, such as lack of monotonicity and convergence, are discussed and the method is applied to real and simulated data.

L’approche standard d’espérance-maximisation pour l’estimation des paramètres des modèles de mélanges gaussiens est modifiée en incorporant le bootstrap non paramétrique à l’intérieur du cycle afin d’éviter le surajustement dans les applications de regroupement fondées sur le modèle. Les avantages comprennent des estimations plus raisonnables des paramètres (y compris des estimations de l’erreur-type), des résultats de regroupements plus «réalistes» et une plus grande diminution d’occurrence de maxima locaux. De plus, nous trouvons une meilleure efficacité par rapport à une approche bootstrap non paramétrique traditionnelle utilisant l’EM pour les modèles de mélanges. Nous discutons de propositions pour résoudre les problèmes qui se posent, comme le manque de monotonie et de convergence, et nous appliquons la méthode à des données réelles et simulées.

[Tuesday June 13/Mardi 13 juin, 14:15-14:30]

Olawale Fatai Ayilara (University of Manitoba), **Mohammad Jafari Jozani** (University of Manitoba)

Quantile Regression with Nominated Samples

Régression par les quantiles avec des échantillons nommés

In this paper, we study quantile regression with nominated samples. Nominated samples have a wide range of applications in medical, ecological and environmental research, and have been shown to perform better than SRS in estimating several population parameters. We propose a new objective function which takes into account the ranking information to estimate the unknown model parameters based on the maxima or minima nomination sampling designs. We compare the mean squared error of the proposed quantile regression estimates using maxima (or minima) nomination sampling design and observe that it provides higher relative efficiency when compared with its counterparts under SRS design for analyzing the upper (or lower) tails of the distribution of the response variable. We also evaluate the performance of our proposed methods when ranking is done with error.

Dans cet article, nous étudions la régression par les quantiles avec des échantillons nommés. Les échantillons nommés possèdent plusieurs applications en recherche médicale, écologique et environnementale, et il a été démontré qu’ils ont un meilleur rendement que l’ÉAS dans l’estimation de plusieurs paramètres de population. Nous proposons une nouvelle fonction objective qui prend en compte l’information de classement pour estimer les paramètres inconnus du modèle en se basant sur les plans d’échantillonnage des nominations maxima ou minima. Nous comparons l’erreur quadratique moyenne des estimations de régression par les quantiles proposées à l’aide d’un plan d’échantillonnage des nominations maxima (ou minima) et observons que leur efficacité relative est plus élevée que celle de leurs équivalents ÉAS pour analyser les queues supérieures (ou inférieures) de la variable réponse. Nous évaluons également la performance de nos méthodes proposées en présence d’erreurs de classement.

[Tuesday June 13/Mardi 13 juin, 14:30-14:45]

Md. Shakhawat Hossain (University of Winnipeg), **Mina Norouzirad** (Shahrood University of Technology), **Mohammad Arashi** (Shahrood University of Technology)

Estimation Strategies for Weighted Least Absolute Deviations Regression Models

Complex Regression and Inference Régression complexe et inférence

Stratégies d'estimation pour les modèles de régression des moindres écarts absolus pondérés

We consider the estimation problem of the weighted least absolute deviation (WLAD) regression parameter vector when there are some outliers in the response and the leverage points in the predictors. We propose the pretest and shrinkage WLAD estimators when some of the parameters may be subject to certain restrictions. We derive the asymptotic risk of the pretest and shrinkage WLAD estimators and show that the asymptotic risk of the shrinkage estimator is strictly less than the unrestricted WLAD estimator. On the other hand, the risk of the pretest WLAD estimator depends on the validity of restrictions. Furthermore, we study the WLAD absolute shrinkage and selection operator and compare its relative performance with the pretest and shrinkage WLAD estimators. A simulation study is conducted to evaluate the performance of the proposed estimators relative to the unrestricted WLAD estimator. A real life data example is used to illustrate the performance of the proposed estimators.

Nous considérons le problème d'estimation du vecteur des paramètres de régression des moindres écarts absolus pondérés (MEAP) en présence de quelques valeurs aberrantes dans la réponse et les points de levier des prédictors. Nous proposons les estimateurs MEAP prétest et rétrécissement lorsque certains des paramètres peuvent être soumis à certaines restrictions. Nous déduisons le risque asymptotique des estimateurs MEAP prétest et rétrécissement et montrons que le risque asymptotique de l'estimateur de rétrécissement est strictement inférieur à l'estimateur MEAP non restreint. D'autre part, le risque de l'estimateur MEAP prétest dépend de la validité des restrictions. De plus, nous étudions l'opérateur de rétrécissement et de sélection MEAP et comparons sa performance relative à celle des estimateurs MEAP prétest et rétrécissement. Une étude de simulation est menée pour évaluer la performance des estimateurs proposés par rapport à l'estimateur MEAP sans restriction. Un exemple utilisant des données réelles sert à illustrer la performance des estimateurs proposés.

[Tuesday June 13/Mardi 13 juin, 14:45-15:00]

Jinhong You

Estimation and Identification of Varying-Coefficient Additive Models for Locally Stationary Process

Estimation et identification de modèles additifs à coefficients variables pour des processus localement stationnaires

Nonparametric regression models allowing for locally stationary covariates have received increasing interest. To catch the dynamic nature of regression function, we adopt more flexible structural nonparametric model than classical varying-coefficient model (VC) and additive model (AM), named varying-coefficient additive model (VCAM). For this model, we propose a three-step spline estimation method for varying-coefficient and additive component functions and show their consistency and rate of convergence. Furthermore, we develop a two-stage penalty procedure to identify varying-coefficient and additive terms in the VCAM. We demonstrate that the proposed identification procedure is consistent, i.e., with probability approaching to one, we correctly select varying-coefficient and additive terms, respectively. Simulation studies explore the finite sample performance and validate the asymptotic theories. Two real data applications illustrating the methodologies are presented as well.

Les modèles de régression non paramétriques permettant des covariables localement stationnaires suscitent de plus en plus d'intérêt. Pour tenir compte de la nature dynamique de la fonction de régression, nous adoptons un modèle non paramétrique structurel plus flexible que les modèles à coefficients variables (CV) classiques et que les modèles additifs (MA), que nous appelons modèle additif à coefficients variables (MACV). À cet effet, nous proposons une méthode d'estimation spline en trois étapes pour les fonctions des composants à coefficients variables et additifs et en montrons la cohérence et le taux de convergence. Par ailleurs, nous mettons au point une procédure de pénalité en deux étapes pour identifier les termes à coefficients variables et additifs dans le MACV. Nous prouvons que la procédure d'identification proposée est cohérente, c'est-à-dire que lorsque la probabilité se rapproche de un, nous sélectionnons les termes à coefficients variables et additifs respectifs corrects. Nous effectuons des études de simulation pour explorer la performance de l'échantillon fini et valider les théories asymptotiques. Nous présentons aussi deux applications à des données réelles qui illus-

Complex Regression and Inference Régression complexe et inférence

trent ces méthodologies.

Chair/Président: Connie Stewart
Room/Salle: E2 150 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-13:45]

Khurram Nadeem (Natural Resources Canada), **Charmaine B. Dean** (Western University), **Douglas Woolford** (Western University), **Steve Taylor** (Pacific Forestry Centre, Natural Resources Canada)

Predicting the Unpredictable: Severe Wildfire Risk Prediction in Canada

Prévoir l'imprévisible : prévision du risque de feux de friches dévastateurs au Canada

About 8000 wildfires occur in the protected area of Canada each year. Approximately 2% of these fires exceed 150+ ha in size, but account for most of the suppression costs and are the greatest threat to our communities. Although statistical approaches to fire occurrence prediction (FOP) have evolved over the past 40 years, FOP has not been implemented at a national scale in Canada. We develop a big data (supervised machine learning) based statistical modeling approach to predict severe large wildfire occurrences in Canada using a set of spatially gridded meteorological, topographic and demographic covariates. Our modeling framework is focused on allowing for better preparedness and more effective decisions to share resources among provincial/territorial and federal agencies, and to enhance Canada's resilience to large wildfires. The talk will focus on large wildfire prediction in British Columbia, Canada, to showcase our methodology.

Chaque année, plus de 8 000 feux de friches surviennent dans une zone protégée du Canada. Environ 2 % de ces incendies ont une taille supérieure à 150 hectares, mais ils représentent la plus grande partie des coûts associés à la lutte contre les feux de friches et la plus grande menace pour nos collectivités. Bien que les approches statistiques de la prévision de l'occurrence des feux ont évolué au cours de ces 40 dernières années, la prévision de l'occurrence des feux n'a pas été mise en œuvre à l'échelle nationale au Canada. Nous avons créé une approche de modélisation statistique basée sur des mégadonnées (apprentissage automatique supervisé) afin de prévoir les occurrences de grands feux dévastateurs au Canada au moyen d'un jeu de covariables météorologiques, topographiques et démographiques définies sur une grille spatiale. Notre cadre de modélisation vise à mieux se préparer en cas d'incendie et à prendre des décisions plus efficaces quant au partage des ressources entre les organismes provinciaux, territoriaux et fédéraux, ainsi qu'à améliorer la capacité du Canada à affronter les grands feux de friches. Dans cet article, nous nous intéresserons particulièrement à la prévision des feux de friches en Colombie-Britannique, au Canada, afin de présenter notre méthodologie.

[Tuesday June 13/Mardi 13 juin, 13:45-14:00]

Jabed H. Tomal (University of Toronto Scarborough), **Jan J.H. Ciborowski** (University of Windsor), **Karen Fung** (University of Windsor)

Piecewise Linear Quantile Regression for Estimation of Ecological Breakpoints

Régression quantile linéaire par parties pour l'estimation de points de rupture écologiques

The relationships between ecological variables are usually estimated by fitting statistical models which go through the conditional means of the variables. For example, the piecewise linear regression model - which goes through the conditional mean of the response variable given the predictor(s) - is used to obtain relationships and breakpoints in variables to answer relevant ecological questions. In contrast, the piecewise linear quantile regression model - which goes through the quantiles of the condi-

Les relations entre variables écologiques sont généralement estimées par ajustement de modèles statistiques qui traitent les moyennes conditionnelles des variables. Ainsi, le modèle de régression linéaire par parties - qui traite la moyenne conditionnelle de la variable réponse selon le(s) prédicteur(s) - permet de déterminer les relations et les points de rupture des variables pour répondre à des questions écologiques pertinentes. Par contraste, le modèle de régression quantile linéaire par parties - qui traite les quantiles de la distribution conditionnelle de la variable réponse selon le(s) prédicteur(s)

tional distribution of the response variable given the predictor(s) - provides much richer information in terms of estimating relationships, break-points and confidence bands. The methods are illustrated with two examples from the ecological literature – relating an index of wetlands’ fish community ‘health’ to the amount of human activity in wetlands’ adjacent watersheds; and relating the biomass of Cyanobacteria to the Total Phosphorus concentration in Canadian lakes.

– offre des informations beaucoup plus riches pour l’estimation des relations, des points de rupture et des intervalles de confiance. Nous illustrons ces méthodes par deux exemples tirés de la littérature écologique – d’abord, les relations entre un indice de la santé de la communauté des poissons des milieux humides et la quantité d’activités humaines dans les bassins versants avoisinants; et ensuite les relations entre la biomasse de cyanobactéries et la concentration de phosphore totale dans les lacs canadiens.

[Tuesday June 13/Mardi 13 juin, 14:00-14:15]

Matias Salibian-Barrera (University of British Columbia), **Daniel Dinsdale** (University of British Columbia)

Geostatistical Inference under Preferential Sampling

Inférence géostatistique sous échantillonnage préférentiel

Preferential sampling in geostatistics occurs when the process that determines the sampling locations may depend on the spatial process that is being modeled. In this case inference cannot generally be performed conditionally on the locations. Here we revisit the seminal work of Diggle et al. (2010, JRSS-C, 59, 191-232) and show that, under certain conditions satisfied by most commonly used geostatistical models, previously proposed Monte Carlo estimates for the likelihood function may not behave as expected. We also show that alternative numerical methods to approximate the likelihood function yield noticeably better results than Monte Carlo-based methods. Although we illustrate these results with the relatively simple models in Diggle et al. (2010), the advantage of our approach is particularly important for more complex preferential sampling models.

On parle d’échantillonnage préférentiel en géostatistique lorsque le processus qui détermine les emplacements sélectionnés dépend du processus spatial à modéliser. Dans de tels cas, règle générale il n’est pas possible de produire des inférences conditionnellement aux emplacements choisis. Nous revenons ici sur les travaux précurseurs de Diggle et al. (2010, JRSS-C, 59, 191-232) et montrons que, dans certaines conditions qui sont réunies dans les modèles géostatistiques les plus couramment employés, les estimations de Monte-Carlo de la fonction de vraisemblance proposées précédemment peuvent ne pas se comporter comme prévu. Nous montrons également que les méthodes numériques alternatives d’approximation de la fonction de vraisemblance donnent des résultats bien supérieurs aux méthodes de Monte-Carlo. Même si nous illustrons nos résultats à l’aide des modèles relativement simples de Diggle et al. (2010), l’avantage de notre approche est particulièrement notable en présence de modèles préférentiels plus complexes.

[Tuesday June 13/Mardi 13 juin, 14:15-14:30]

Michelle Carey (McGill University), **James O. Ramsay** (McGill University)

Parameter Estimation for PDEs Defined over Irregular Domains

Estimation des paramètres pour des équations aux dérivées partielles définies dans des domaines irréguliers

Spatial temporal data are abundant in many scientific fields, some examples include; temperature readings from multiple weather stations and the spread of an infectious disease over a particular region taken over time. In many instances the spatial data are accompanied by mathematical models expressed in terms of partial differential equations (PDEs). A PDE determines the theoretical aspects of the behaviour of the physical, chemical or biological phenomena considered. The parameters of the

Les données spatiales temporelles abondent dans de nombreux domaines scientifiques, notamment les relevés de température à partir de plusieurs stations météorologiques et la propagation d’une maladie infectieuse dans une région particulière au fil du temps. Dans de nombreux exemples, les données spatiales sont accompagnées de modèles mathématiques exprimés en termes d’équations aux dérivées partielles. Une équation aux dérivées partielles détermine les aspects théoriques du comportement des phénomènes physiques, chimiques ou biologiques considéré. Les paramètres

PDE are typically unknown and must be inferred from expert knowledge of the phenomena considered. I will introduce a methodology for attaining data driven estimates of the PDE parameters by extending the profiling with parameter cascading procedure outlined in Ramsay et al. (2007). We also incorporate splines on triangulations Lai and Schumaker (2007) to account for attributes of the geometry of the problem such as irregular shaped domains and internal boundary features.

d'une équation aux dérivées partielles sont en général inconnus et doivent être déduits des connaissances approfondies des phénomènes considérés. Je présenterai une méthodologie permettant d'obtenir des estimations fondées sur les données des paramètres d'équation aux dérivées partielles en étendant le profilage au moyen d'une procédure de paramètres en cascade décrite dans les travaux de Ramsay et coll. (2007). Nous intégrerons également les travaux de Lai et Schumaker (2007) concernant les splines définies sur les triangulations pour tenir compte des attributs de géométrie du problème, comme les domaines irréguliers et les délimitations internes.

Time Series and Longitudinal Data

Séries chronologiques et données longitudinales

Chair/Président: Louis-Paul Rivest

Room/Salle: E2 155 (EITC)

Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 13:30-13:45]

Gregory Rice (University of Waterloo), **Lajos Horvath** (University of Utah)

Empirical Eigenvalue Based Testing for Structural Breaks in Linear Factor Models

Test empirique fondé sur la valeur propre des ruptures structurelles dans les modèles factoriels linéaires

Testing for stability in linear factor models has become an important topic in both the statistics and econometrics research communities. We develop a general test for structural stability of linear factor models that is based on monitoring for changes in the largest eigenvalue of the sample covariance matrix. The asymptotic distribution of the proposed test statistic is established under the null hypothesis that the mean and covariance structure of the cross sectional units remain stable during the observation period. We show that the test is consistent assuming common breaks in the mean or factor loadings. These results are investigated by means of a Monte Carlo simulation study, and their usefulness is demonstrated with an application to U.S. treasury yield curve data, in which some interesting features of the 2007-2008 subprime crisis are illuminated.

Les tests de stabilité des modèles factoriels linéaires sont devenus un sujet important en statistique et en économétrie. Nous mettons au point un test général pour la stabilité structurelle des modèles factoriels linéaires, fondé sur le suivi de l'évolution de la plus grande valeur propre de la matrice de covariance de l'échantillon. Nous déterminons la distribution asymptotique de la statistique de test proposée, sous l'hypothèse nulle selon laquelle la moyenne et la structure de covariance des unités transversales restent stables pendant la période d'observation. Nous montrons que le test est convergent pour des ruptures communes de la moyenne ou des coefficients de saturation. Nous étudions ces résultats par une simulation de Monte-Carlo et illustrons leur efficacité par une application aux données de la courbe de rendement du département du Trésor des États-Unis, en mettant en lumière certaines caractéristiques intéressantes de la crise des prêts hypothécaires à risque de 2007-2008.

[Tuesday June 13/Mardi 13 juin, 13:45-14:00]

Md. Erfanul Hoque (University of Manitoba), **Mahmoud Torabi** (University of Manitoba)

Modeling the Random Effects Covariance Matrix for Longitudinal Data with Covariate Measurement Error

Modélisation de la matrice de covariance des effets aléatoires pour données longitudinales avec erreur de mesure des covariables

Longitudinal data occur frequently in practice such as medical studies and life sciences. Generalized linear mixed models (GLMMs) are commonly used to analyze such data. It is typically assumed that the random effects covariance matrix is constant across the subject (and among subjects) in these models. In many situations, however, this correlation structure may differ among subjects and ignoring this heterogeneity can cause the biased estimate of model parameters. Covariates measured with error also happen frequently in the longitudinal data set-up. Ignoring this issue in the data may also produce bias in model parameters estimate and lead to wrong conclusions. In this work, we propose an approach to properly model the ran-

Les données longitudinales sont fréquentes en pratique dans les études médicales et en sciences de la vie. On utilise communément des modèles mixtes linéaires généralisés (GLMM) pour les analyser. On considère généralement, dans ces modèles, que la matrice de covariance des effets aléatoires est constante pour le sujet (et entre sujets). Mais bien souvent, cette structure de corrélation varie entre les sujets, et si on ignore cette hétérogénéité, l'estimation des paramètres du modèle peut être biaisée. Il est également fréquent que les covariables soient sujettes à une erreur de mesure dans les données longitudinales. Négliger cette question peut également produire un biais dans l'estimation des paramètres du modèles et mener à des conclusions erronées. Dans cette présentation, nous proposons une approche qui permet de modéliser adéquatement la matrice de covariance des effets aléa-

Time Series and Longitudinal Data

Séries chronologiques et données longitudinales

dom effects covariance matrix based on covariates in the class of GLMMs where covariates are subject to measurement error. The resulting parameters from the decomposition of random effects covariance matrix have a sensible interpretation and can easily be modeled without the concern of positive definiteness of the resulting estimator. Performance of the proposed approach is evaluated through simulation studies as well as a real longitudinal data application from Manitoba Follow-up study.

toires en fonction des covariables, dans la classe de GLMM où les covariables sont sujettes à une erreur de mesure. Les paramètres résultant de la décomposition de la matrice de covariance des effets aléatoires se laissent bien interpréter et sont faciles à modéliser sans devoir se soucier de la définitivité positive de l'estimateur résultant. Nous évaluons la performance de l'approche proposée par des études de simulation et une application à des données longitudinales réelles tirées de l'étude de suivi à long terme du Manitoba (Manitoba Follow-up study).

[Tuesday June 13/Mardi 13 juin, 14:00-14:15]

Lichen Chen (University of Waterloo), **Adam Kolkiewicz** (University of Waterloo), **Tony Wirjanto** (University of Waterloo)

Empirical Likelihood Multiscale Inference of the Scaling Property of Time Series Models

Inférence multi-échelle de la vraisemblance empirique de la propriété d'échelle des modèles de séries chronologiques

Over the years, financial time series have become available at increasingly higher frequencies. It is well-known that time series sampled at different frequencies carry information relevant to different components in the original time series, e.g. fast- and slow-moving components. We are interested in a research question of how to take advantage of samples at multiple frequencies in order to improve statistical inference in the presence of multiscale phenomenon. Multiscale phenomenon in financial time series have been reported by researches in a wide range of fields, ranging from physicists, applied mathematicians, to econometricians. In this talk, we introduce an empirical likelihood (EL) based test for model scaling property using multiple frequency samples. The method is used to test whether or not a given model can be consistently used across different time scales. For application, we apply it to models for capturing multiscale phenomenon in financial asset's volatility.

Au fil des ans, les séries chronologiques financières sont devenues disponibles à des fréquences toujours croissantes. Il est bien connu que les séries chronologiques échantillonnées à des fréquences différentes portent des informations pertinentes pour différentes composantes de la série chronologique initiale, par exemple, les composantes à déplacement rapide et lent. Nous nous intéressons à une question de recherche sur la façon de tirer profit des échantillons à fréquences multiples afin d'améliorer l'inférence statistique en présence de phénomène multi-échelle. Le phénomène multi-échelle dans les séries chronologiques financières a été rapporté par des chercheurs dans un large éventail de domaines, allant des physiciens, mathématiciens appliqués, aux économétriciens. Dans cet exposé, nous introduisons un test basé sur la vraisemblance empirique (VE) pour la propriété d'échelle du modèle à l'aide d'échantillons à fréquence multiple. La méthode est utilisée pour tester si un modèle donné peut ou non être utilisé de manière cohérente à travers différentes échelles de temps. Nous appliquons la méthode à des modèles capturant le phénomène multi-échelle dans la volatilité de l'actif financier.

[Tuesday June 13/Mardi 13 juin, 14:15-14:30]

Yuanhao Lai (Western University), **A. Ian McLeod** (Western University)

A Nonparametric Test for Detecting Periodicity in Time Series

Un test non paramétrique pour détecter la périodicité dans les séries chronologiques

Traditional methods of detecting periodicity in time series depend on the assumption of identical independent Gaussian noises and are sensitive to outliers. We then propose an innovative nonparametric robust periodicity test which is highly resistant to

Les méthodes traditionnelles de détection de la périodicité dans les séries chronologiques reposent sur l'hypothèse de bruits gaussiens indépendants identiques et sont sensibles aux valeurs aberrantes. Nous proposons alors un test innovateur de périodicité robuste et non paramétrique, très résis-

Time Series and Longitudinal Data

Séries chronologiques et données longitudinales

outliers and is distribution free. It also works when the second moment of the error is not finite while traditional methods fail. We derive an asymptotic distribution for the test statistic and show that this method is efficient under the assumption of Gaussian noises. A response surface regression approach is also implemented to approximate the finite distribution of the test statistic. In addition, the implementation of this approach is dramatically sped up by utilizing the parallel computation through the Shared Hierarchical Academic Research Computer Network.

tant aux valeurs aberrantes et sans aucune hypothèse quant à la distribution. Il fonctionne également lorsque le deuxième moment de l'erreur n'est pas fini alors que les méthodes traditionnelles échouent. Nous dérivons une distribution asymptotique pour la statistique du test et montrons que cette méthode est efficace sous l'hypothèse de bruits gaussiens. Une approche de régression de surface de réponse est également mise en œuvre pour approximer la distribution finie de la statistique du test. De plus, la mise en œuvre de cette approche est considérablement accélérée en utilisant le calcul parallèle à travers le réseau informatique de recherche universitaire hiérarchique partagé.

Recent Developments in Joint Modelling of Mixed Types of Outcomes Développements récents dans la modélisation conjointe de types de résultats mixtes

Chair/Président: Xin (Cindy) Feng
Organizer/Responsable: Xin (Cindy) Feng
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-16:00]

Charmaine B. Dean (University of Western Ontario)

Joint Analysis of Longitudinal Zero-heavy Panel Count Outcomes with Application to Understanding Desistance from Criminal Activity

Analyse conjointe de variables de dénombrement longitudinales à surreprésentation de zéros avec application à la compréhension du renoncement de la criminalité

Regression models for zero-inflated count data often need to accommodate within-subject correlation and between-individual heterogeneity; frequently random effects models are utilized for incorporating complex correlation structures. In cases where several longitudinal zero-heavy count outcomes are jointly considered, excess zeros may arise from several distinct sources. This is the case for a study of the patterns and mechanisms of the process of desistance from criminal activity. Methodological challenges in the analysis of longitudinal criminal behaviour data include the need to develop methods for multivariate longitudinal discrete data, incorporating modulating exposure variables and several possible sources of zero-inflation. There are additional complications such as some outcomes being prohibited during time in a secure facility; as well as intervention carry-over effects for some outcomes, and that the underlying process generating events may resolve for some individuals.

Les modèles de régression pour les données de dénombrement à surreprésentation de zéros doivent souvent prendre en considération la corrélation à l'intérieur même du sujet ainsi que l'hétérogénéité entre individus. Les modèles à effets aléatoires sont d'ailleurs souvent utilisés pour incorporer des structures de corrélation complexes. Lorsque plusieurs variables dépendantes longitudinales de dénombrement à surreprésentation de zéros sont étudiés conjointement, un excès de zéros peut provenir de plusieurs sources distinctes. C'est le cas de l'étude des structures et des mécanismes du processus de renoncement des activités criminelles. L'analyse des données longitudinales du comportement criminel illustre le besoin de développer des méthodes pour les données longitudinales discrètes multivariées, permettant l'incorporation de variables d'exposition modulantes ainsi que de prendre en compte plusieurs sources possibles de surreprésentation de zéros. Il existe des complications supplémentaires, comme par exemple la collecte de certaines variables qui était interdite dans le passé dans les endroits sécurisés, ainsi que des effets de reports d'interventions pour certaines variables, et des éléments générateurs du processus sous-jacents qui peuvent se résoudre pour certains individus.

[Tuesday June 13/Mardi 13 juin, 16:00-16:30]

Dongsheng Tu, Hui Song (Dalian University of Technology), **Yingwei Peng** (Queen's University)

Joint Modeling of Longitudinal Proportional Measurements and Survival Times with a Cure Fraction

Modélisation conjointe de mesures proportionnelles longitudinales et de temps de survie avec fraction de guérison

In cancer clinical trials and other medical studies, both longitudinal measurements and data on a time to an event (survival time) are often collected from the same patients. Joint analyses of these data would improve the efficiency of the statistical inferences. We propose a new joint model for the longitudinal proportional measurements which are restricted in a finite interval and survival times with a potential cure fraction. A penalized joint likeli-

Lors d'études cliniques sur le cancer et autres études médicales, les mesures longitudinales et les données sur le temps d'événements (temps de survie) sont souvent recueillies chez les mêmes patients. Des analyses conjointes de ces données amélioreraient l'efficacité des inférences statistiques. Nous proposons un nouveau modèle conjoint pour les mesures proportionnelles longitudinales qui sont restreintes dans un intervalle fini et pour les temps de survie avec une fraction de guérison potentielle. Une vraisemblance conjointe pénali-

Recent Developments in Joint Modelling of Mixed Types of Outcomes Développements récents dans la modélisation conjointe de types de résultats mixtes

hood is derived based on the Laplace approximation and a semiparametric procedure based on this likelihood is developed to estimate the parameters in the joint model. A simulation study is performed to evaluate the statistical properties of the proposed procedures. The proposed model is applied to data from a clinical trial on early breast cancer.

sée est dérivée en se basant sur l'approximation de Laplace et une procédure semi paramétrique basée sur cette vraisemblance est développée pour estimer les paramètres du modèle conjoint. Une étude de simulation est réalisée pour évaluer les propriétés statistiques des procédures proposées. Le modèle proposé est appliqué aux données d'un essai clinique sur le cancer du sein précoce.

[Tuesday June 13/Mardi 13 juin, 16:30-17:00]

Elizabeth Juarez-Colunga (University of Colorado Denver), **Brandie Wagner** (University of Colorado Denver), **Edith Zemanick** (University of Colorado Denver)

Joint Analysis of a Longitudinal Binary Outcome and Recurrent Events with Application to Cystic Fibrosis Outcomes
Analyse conjointe d'une variable longitudinale binaire et d'événements récurrents avec une application sur la fibrose kystique

In cystic fibrosis, chronic *Pseudomonas aeruginosa* (Pa) infection is associated with worse clinical outcomes including more frequent pulmonary exacerbations (PE). PE are themselves a leading cause of morbidity in CF and important endpoints in CF clinical trials. This talk discusses joint models that address the challenge of understanding the co-dependence of progression of Pa infection and recurrent PE over time. Using data from the ongoing Early *Pseudomonas* Infection Control study, we propose a joint model built within the frameworks of hidden Markov chain models to model progression of Pa, and dynamic recurrent event models to model the recurrence of PE. We address additional challenges in the motivating study including missing and misalignment of covariates, the dynamic aspect of the recurrence of PE, and the latent aspect of the different Pa states of infection.

Chez les personnes atteintes de la fibrose kystique, les infections à *Pseudomonas aeruginosa* (Pa) chroniques sont associées à de très mauvais résultats cliniques dont des exacerbations pulmonaires (PE) plus fréquentes. Les PE sont elles-mêmes une des causes principales de morbidité chez les personnes atteintes de la fibrose kystique et représentent d'importants paramètres lors d'essais cliniques sur cette maladie. Cet exposé discute des modèles conjoints qui permettent une compréhension de la co-dépendance entre la progression de l'infection Pa et les PE récurrents à travers le temps. À l'aide de données provenant de l'étude en cours Early *Pseudomonas* Infection Control, nous proposons un modèle conjoint construit à l'intérieur du cadre des modèles de chaînes de Markov cachées qui permet de modéliser la progression de l'infection Pa. Nous proposons aussi des modèles dynamiques d'événements récurrents pour modéliser la récurrence des PE. D'autres défis sont abordés, comme les valeurs manquantes et le désalignement des covariables, l'aspect dynamique de la récurrence des PE et l'aspect latent des différents stades de l'infection Pa.

New Vistas in Applied Probability
Nouvelles perspectives en probabilité appliquée

Chair/Président: Gail Ivanoff
Organizer/Responsable: Gail Ivanoff
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-16:00]

Yaniv Plan (University of British Columbia), **Chris Liaw** (University of British Columbia), **Abbas Mehrabian** (University of British Columbia), **Roman Vershynin** (University of Michigan)

A Simple Tool for Bounding the Deviation of Random Matrices on Geometric Sets, with Applications

Un outil simple pour borner la déviation de matrices aléatoires sur des ensembles géométriques; avec applications

Let A be an isotropic, sub-gaussian m by n matrix. We give a simple sufficient condition for a random sub-Gaussian matrix to be well conditioned when restricted to a subset of R^n . We also prove a local version of this theorem, which allows for unbounded sets. These theorems have various applications, such as a general theory of compressed sensing.

Supposons que A est une matrice m par n isotropique et sous-gaussienne. Nous fournissons une condition simple suffisante pour obtenir une matrice sous-gaussienne aléatoire bien conditionnée lorsqu'elle est restreinte à un sous-ensemble R^n . Nous prouvons également une version locale de ce théorème qui permet des ensembles non bornés. Ces théorèmes servent à diverses applications, comme la théorie générale de l'acquisition comprimée.

[Tuesday June 13/Mardi 13 juin, 16:00-16:30]

Clarence Simard (Université du Québec à Montréal)

Approximation of Squared Integrable Random Variables with Nonlinear Stochastic Integrals

Approximation d'une variable aléatoire de carré intégrable par une intégrale stochastique non linéaire.

Global quadratic hedging is the problem of finding a trading strategy which minimizes the expected squared error between the value of the portfolio and the value of a contingent claim. Under some conditions, the value of the portfolio is given by a stochastic integral and the value of the contingent claim by a random variable. In this context, the problem of global quadratic hedging can be seen as the projection of an L^2 random variable in the linear space of stochastic integrals respect to a semimartingale. More recently, to incorporate the effect of illiquidity, stock prices are given by semimartingales with space parameters. In this case, the value of the portfolio is given by a stochastic integral which is no more linear with respect to the integrand. During this talk, we will look at some conditions which allow to find an L^2 approximation of random variables with nonlinear stochastic integrals.

Le problème de couverture quadratique a pour objectif de déterminer la stratégie de transactions qui minimise le carré de l'erreur entre la valeur du portefeuille et la valeur du droit contingent. Sous certaines conditions, la valeur du portefeuille est donnée par une intégrale stochastique et la valeur du droit contingent par une variable aléatoire. Dans ce contexte, le problème de couverture quadratique peut être vu comme la projection d'une variable aléatoire dans L^2 sur l'espace des intégrales stochastiques par rapport à une semimartingale. Récemment, pour représenter l'effet de l'illiquidité, on a utilisé des semimartingales avec paramètre spatial pour définir le prix des actions. Dans ce cas, la valeur du portefeuille est donnée par une intégrale stochastique qui n'est plus une fonction linéaire de l'intégrande. Pour cette présentation, nous allons étudier des conditions qui permettent de trouver une approximation d'une variable aléatoire dans L^2 avec une intégrale stochastique non linéaire.

[Tuesday June 13/Mardi 13 juin, 16:30-17:00]

Yi Shen (University of Waterloo)

Probabilistic Symmetries and Random Locations

Symétries probabilistes et emplacements aléatoires

New Vistas in Applied Probability

Nouvelles perspectives en probabilité appliquée

In this talk we briefly discuss different types of symmetries in probability, including stationarity, stationarity of the increments, isotropy, self-similarity, exchangeability, and their combinations. Each of these symmetries is naturally related to an operator in the path space, in the sense that the symmetry can be expressed as the invariance of the distribution of the stochastic processes with respect to the corresponding operator. In particular, we consider the random locations of the stochastic processes, such as the hitting times or the location of the path supremum over a fixed interval. On one hand, we see how the probabilistic symmetries can imply the properties of the distributions of these random locations; on the other hand, we also discuss how the distributions of the random locations can be used to characterize the probabilistic symmetries.

Cet exposé porte sur différents types de symétries en probabilité, y compris la stationnarité, la stationnarité des incréments, l'isotropie, l'autosimilarité, l'échangeabilité, et leurs combinaisons. Chacune de ces symétries est naturellement liée à un opérateur dans un espace de trajectoires, ce qui veut dire que la symétrie peut être exprimée comme l'invariance de la distribution des processus stochastiques en ce qui concerne l'opérateur correspondant. Nous abordons en particulier les emplacements aléatoires des processus stochastiques, comme les temps d'entrée ou l'emplacement du chemin supremum sur un intervalle fixe. D'une part, nous voyons comment les symétries probabilistes peuvent impliquer les propriétés des distributions de ces emplacements aléatoires; d'autre part, nous abordons la façon dont nous pouvons utiliser les distributions de ces emplacements aléatoires pour caractériser les symétries probabilistes.

How to Stay Sane Pre-tenure: Addressing the Challenges of New Investigators Comment rester sain d'esprit jusqu'à l'agrégation : les défis des nouveaux chercheurs

Chair/Président: Nathaniel T. Stevens
Organizer/Responsable: Nathaniel T. Stevens
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-17:00]

Nathaniel T. Stevens (University of San Francisco), **Hua Shen** (University of Calgary), **Mireille E. Schnitzer** (Université de Montréal), **Jingjing Wu** (University of Calgary), **Erica E.M. Moodie** (McGill University)

Panel discussion

Discussion de groupe

In this panel discussion, four panelists at different career stages will begin by addressing the audience with prepared points prior to a 50-minute question-and-answer period facilitated by the session chair. Likely points of discussion include: a post-doc or not? navigating applications, interviews and job offers; applying for funding (NSERC, CIHR, NSF, NIH etc.); applying for tenure; and work-life balance.

Dans cette discussion de groupe, quatre participants à différentes étapes de leur carrière présenteront au public des points préselectionnés avant une période de 50 minutes de questions et de réponses animée par le président ou la présidence de la séance. Les points de discussion comprennent ce qui suit : faire ou ne pas faire un post-doctorat ? Naviguer dans les demandes d'emploi, les entrevues et les offres d'emplois ; faire des demandes de subventions (Instituts de recherche en santé du Canada, du Conseil de recherches en sciences naturelles et en génie, Fondation nationale des sciences, National Institutes of Health des États-Unis, etc.), postuler à un poste permanent ; concilier sa vie professionnelle et sa vie privée.

Statistical Methods for Omics Data

Méthodes statistiques pour les données “-omiques”

Chair/Président: Pingzhao Hu
Organizer/Responsable: Pingzhao Hu
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-16:00]

Wei Xu

Statistical Methodology on Human Microbiome Data Analysis

Méthodologie statistique sur l'analyse de données sur le microbiome

The technological development in genomic sequencing has enabled researchers to unveil the wide variability of bacteria presented within different locations of the body. It is necessary to better understand both environmental and host genetic factors impact the composition of the microbiome to improve disease management. Several analytic approaches are introduced to summarize and assess the single or multiple OTUs using different computational algorithms. Motivated by the multivariate nature of microbiome data with hierarchical taxonomic clusters, counts that are often skewed and zero inflated, we propose a Bayesian latent variable methodology to jointly model multiple operational taxonomic units within a single taxonomic cluster. This novel method can incorporate both negative binomial and zero-inflated negative binomial responses, and can account for serial and familial correlations. This method can help discover both genetic and environmental factors that influence the microbiome.

Les progrès technologiques dans le séquençage du génome ont permis aux chercheurs de révéler la grande variabilité des bactéries présentes dans différentes parties du corps. Il est nécessaire de mieux comprendre comment les facteurs environnementaux et génétiques de l'hôte ont une incidence sur composition du microbiome pour améliorer la gestion des maladies. Plusieurs approches analytiques sont présentées dans le but de résumer et évaluer un seul OTU ou plusieurs OTUs grâce à différents algorithmes computationnels. Motivés par la nature multivariée des données sur le microbiome avec des classifications hiérarchiques taxonomiques, et par le fait que ces données de dénombrement sont souvent asymétriques et à surreprésentation de zéros, nous proposons une méthodologie bayésienne à variable latente pour modéliser conjointement plusieurs unités taxonomiques opérationnelles au sein d'une seule classification taxonomique. Cette nouvelle méthode peut intégrer des variables réponses binomiales négatives à surreprésentation de zéros et peut prendre en compte des corrélations temporelles et familiales. Cette méthode peut aider à découvrir les facteurs génétiques et environnementaux qui influencent le microbiome.

[Tuesday June 13/Mardi 13 juin, 16:00-16:30]

Elif Fidan Acar (University of Manitoba)

Aspects of Genetic Meta-Analysis under Data Uncertainty

Aspects de la meta-analyse de la génétique dans le cadre de l'incertitude entourant les données

Genotype imputation is a technique extensively used in genome-wide association studies (GWAS), as well as in their meta-analysis. In this talk, we outline strategies to account for imputation accuracy when including imputation-based GWAS results in a meta-analysis, considering both fixed-effect and random-effects models. While adding studies to the meta-analysis typically boosts the power to detect genetic associations, inclusion of imputation-based studies may adversely affect power due to increased genotype uncertainty. We address this trade-off via a reweighing scheme

L'imputation des génotypes est une technique largement utilisée dans les études d'association pangénomiques, ainsi que dans leur meta-analyses. Dans cet exposé, nous décrivons les stratégies visant à prendre en compte la précision de l'imputation lorsqu'on inclut des résultats provenant d'études d'association pangénomiques basées sur l'imputation dans une meta-analyse, tout en considérant les modèles à effets fixes et à effets aléatoires. Alors que le fait d'ajouter des études à la meta-analyse augmente en général la puissance de détection des associations génétiques, l'inclusion d'études basées sur de l'imputation pourrait avoir une incidence négative sur la puissance en raison de l'incertitude accrue du génotype.

Statistical Methods for Omics Data

Méthodes statistiques pour les données “-omiques”

that controls the contribution of imputation-based studies in meta-analyses. The proposed method achieves a better detection power relative to traditional approaches, and improves the validity and reliability of imputation-based meta-analysis results.

Nous abordons ce compromis au moyen d'un système de pondération qui contrôle la contribution des études basées sur de l'imputation. La méthode proposée réussit à donner une meilleure puissance de détection par rapport aux approches traditionnelles, et elle améliore la validité et la fiabilité des résultats de meta-analyses basées sur de l'imputation.

[Tuesday June 13/Mardi 13 juin, 16:30-17:00]

David R. Bickel (University of Ottawa)

Correcting False Discovery Rates for their Bias Toward False Positives

Correction des taux de fausses découvertes pour leur biais à l'égard des faux positifs

Conventional methods of adjusting p values for multiple comparisons control a family-wise error rate (FWER) such as a genome-wise error rate. The recognition that they lead to excessive false negative rates in genomics applications has led to widespread use of false discovery rates (FDRs) in place of the conventional adjustments. While this is an improvement, the way FDRs are used in the analysis of genomics data leads to the opposite problem, excessive false positive rates. In this sense, the FDR overcorrects for the excessive conservatism (bias toward false negatives) of the FWER-controlling methods of adjusting p values. Estimators of the local FDR (LFDR) are much less biased but have not been widely adopted due to their high variance and lack of availability in software. To address both issues, we propose estimating the LFDR by correcting an estimated FDR or the level at which an FDR is controlled. Preprint: <http://hdl.handle.net/10393/34277>

Les méthodes conventionnelles d'ajustement des valeurs-p pour des comparaisons multiples contrôlent un taux d'erreur global, comme le taux d'erreur dans l'ensemble du génome par exemple. La reconnaissance du fait que ces procédures mènent à des taux de faux négatifs excessifs dans les applications génomiques a conduit à une utilisation généralisée du taux de fausses découvertes à la place des ajustements conventionnels. Même si c'est une amélioration, la façon dont les taux de fausses découvertes sont utilisés dans l'analyse des données génomiques engendre le problème inverse, à savoir des taux de faux positifs excessifs. Le taux de fausses découvertes sur-corrige donc le conservatisme excessif (biais à l'égard des faux négatifs) de la méthode de contrôle du taux d'erreur global. Les estimateurs des taux de fausses découvertes locaux sont beaucoup moins biaisés, mais n'ont pas été largement adoptés en raison de leur variance élevée et de leur faible disponibilité dans les logiciels. Pour remédier à ces problèmes, nous proposons d'estimer le taux de fausses découvertes local en corrigeant le taux de fausses découvertes ou le niveau auquel un taux de fausses découvertes est contrôlé. Preprint : <http://hdl.handle.net/10393/34277>.

Brewing Insight from Beer Data: In Honour of Gosset's Birthday

La sagesse est dans la bière et ses données : en l'honneur de l'anniversaire de Gosset

Chair/Président: Jason Loeppky and David Campbell
Organizer/Responsable: Jason Loeppky and David Campbell
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-16:00]

Pulindu Ratnasekera (Simon Fraser University)

Geographic Topic Modelling to Predict Regional Beer Success

Modélisation géographique pour prévoir le succès d'une bière selon la région

This study explores online beer reviews to identify which types and flavours of beer are popular among consumers in different geographic regions. The exploration of beer reviews was carried out with Latent Dirichlet Allocation, a Natural Language Processing approach to cluster text from beer reviews into flavour and sentiment topics. Combining text and topic data from reviews with numerical ratings of beers in categories of flavour, scent, and colour we can determine flavour preferences. When matching breweries with their geographical locations, we explore beer flavour and type preference trends across the country. This analysis could be used as a recommendation system for predicting regional beer success from flavour and beer type combinations.

Cette étude passe en revue des critiques en ligne de bières pour en identifier les types et les saveurs populaires chez les consommateurs dans diverses régions. L'étude des commentaires critiques sur les bières a été menée avec l'allocation de Dirichlet latente, une approche du traitement du langage naturel, afin de regrouper les textes en fonction des saveurs et des émotions. En combinant les données textuelles à un classement numérique des bières dans les catégories de la saveur, du parfum et de la couleur, il est possible de déterminer les préférences en matière de saveur. En situant les brasseries dans leurs emplacements géographiques, nous pouvons découvrir quelles sont les tendances en matière de préférences de saveur et de type de bière dans l'ensemble du pays. Cette analyse peut servir de système de recommandations pour prévoir le succès d'une bière dans une région donnée en combinant les données sur la saveur et le type de bière.

[Tuesday June 13/Mardi 13 juin, 16:00-16:30]

Joanna Zhao (Simon Fraser University), **Derek Qiu** (Simon Fraser University)

How to Brew your Beer Tour with Data Science

Préparer une tournée des brasseries grâce à la science des données

Motivated by the emerging craft brewery scene in Vancouver B.C., we set our goal to design a beer tour to visit some of the amazing hotspots in town. Using Hadley Wickham's tidyverse package, we automate data extraction to pinpoint exact locations of breweries, and proceed to extract distances and travelling times between them from GoogleMaps API. A travelling salesman algorithm is then applied to optimize for the best tour path, which can be tailored for any individual's preferences. To improve the usability of our tour planning, we extend our workflow into an interactive and customizable web application built with Shiny. Our app shows the location of all the craft breweries in Vancouver and their respective menus, with average rating scores of all menu options. Users can simply point and click to express their preferences and ob-

Motivés par la fabrication émergente de bières artisanales à Vancouver, en Colombie-Britannique, notre objectif était de concevoir une tournée de certains des endroits populaires à cet égard. À l'aide du paquet tidyverse d'Hadley Wickham, nous avons automatisé une extraction des données pour déterminer les emplacements exacts de ces brasseries et avons tiré de GoogleMaps API les distances et les durées de trajet entre ces établissements. Un algorithme de voyageur de commerce est ensuite appliqué pour optimiser le meilleur circuit de tournée qui peut être personnalisé au choix de chacun. Pour améliorer l'utilisabilité de notre planification de tournée, nous avons intégré notre flux de travail dans une application Web interactive et personnalisable développée avec Shiny. Notre application affiche l'emplacement de toutes les brasseries artisanales dans Vancouver et leur menu respectif, avec la cote moyenne de tous les choix proposés. L'utilisateur peut simplement pointer et cliquer pour exprimer ses

Brewing Insight from Beer Data: In Honour of Gosset's Birthday

La sagesse est dans la bière et ses données : en l'honneur de l'anniversaire de Gosset

tain a personalized path. A utility chart derived from user inputs that displays the marginal utility gained from each stop is also provided.

préférences et obtenir un trajet personnalisé. Un tableau de services tiré des commentaires des utilisateurs et affichant l'utilité marginale de chaque arrêt est aussi fourni.

[Tuesday June 13/Mardi 13 juin, 16:30-17:00]

James A. Hanley (McGill University)

Gosset at Guinness: How 'Student' Derived (and Checked) the Distribution of his z-statistic

W.S. Gosset, « Student », chez Guinness : comment il a obtenu (et vérifié) la distribution de sa statistique Z

William Gosset ("Student") published "The Probable Error of a Mean" in 1908. Despite its seminal nature, modern-day statistics textbooks give him, and this article, short shrift. Few of today's students – or their teachers – are aware of the "z" statistic whose sampling distribution he actually derived, the mathematical derivation, his simulations to check his work, the material used in the simulations, the table he produced, the "one-line" missing proof supplied by the 22-year old Fisher (still a student himself) or the subsequent switch, in collaboration with Fisher, from the z to the t-statistic. I remind readers of these. I hope the next generation of statisticians come to know more about the man than simply that "he worked for the Guinness brewery," and appreciate that not all statistical distributions are derived in a single pass. Students would do well to use his paper as a model when writing their first statistical article. Extra material: www.epi.mcgill.ca/hanley/Student/

William « Student » Gosset a publié « The Probable Error of a Mean » (L'erreur probable de la moyenne) en 1908. Malgré son esprit fécond, les manuels de statistique d'aujourd'hui font peu de cas de lui, ni de son article. De nos jours, rares sont les étudiants – ou leurs professeurs – qui connaissent la statistique Z dont W. Gosset a obtenu la distribution d'échantillonnage, la dérivation mathématique, ou ses simulations pour évaluer ses travaux, le matériel utilisé pour ses simulations, le tableau qu'il a conçu, la preuve manquante d'« une ligne » fournie par l'étudiant de 22 ans Ronald Fisher, ni le passage subséquent de la statistique Z à la T, en collaboration avec ce même Fisher. Nous ferons un rappel de ces faits, en espérant que la prochaine génération de statisticiens connaisse mieux cet homme que simplement comme « le gars qui a travaillé à la brasserie Guinness » et convenir que toutes les distributions statistiques ne sont pas dérivées en une seule fois. Il serait bon que les étudiants s'inspirent de ses écrits pour la rédaction de leur premier article sur la statistique. Autre matériel : www.epi.mcgill.ca/hanley/Student/

Applied Survey Research with Impact

Recherche appliquée sur les sondages avec un impact

Chair/Président: Matthias Schonlau
Organizer/Responsable: Matthias Schonlau
Room/Salle: E2 150 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-16:00]

Kristen Himelein (World Bank)

Innovations in Sampling for Insecure and Data-Poor Environments

Innovations dans l'échantillonnage pour les environnements non sécuritaires et pauvres en données

The standard challenges faced in survey design – completeness of sample frame, sample selection, interviewer quality, etc. – can manifest themselves in unique ways in the developing world. Sample frames are incomplete or non-existent. Interviewer education and capacity are limited. Field conditions are often challenging, and the possibility (or probability) of unforeseen events must be built into survey designs. The presentation covers a set of illustrative recent case studies of the challenges faced in sub-Saharan Africa. Topics include: using satellite imagery as a sampling frame in Kinshasa; a proposed alternative to random walk for second stage selection in conflict areas of Mogadishu; the application of a random geographic cluster sample methodology to a survey of pastoralist and nomadic populations in eastern Ethiopia; and the implementation of a high frequency cell phone survey to measure the socioeconomic impacts of Ebola in Sierra Leone and Liberia. Focus will be on technical and implementation challenges faced, how they were addressed while maintaining a generalizable probability design, and lessons learned for future applications.

Les problèmes courants auxquels on fait face dans les plans d'enquête – exhaustivité de la base d'échantillonnage, sélection de l'échantillon, qualité des intervieweurs, etc. – peuvent survenir de façon particulière dans les pays en développement. La formation des intervieweurs et leurs capacités sont limitées. Les conditions en milieu réel posent souvent des défis, et le risque (ou la probabilité) d'imprévu doit être intégré dans les plans d'échantillonnage. Cet exposé présente une série d'études de cas récentes qui illustre les défis en Afrique subsaharienne. Nous aborderons les sujets suivants : l'utilisation de l'imagerie satellitaire comme base d'échantillonnage à Kinshasa ; la proposition d'une solution de rechange au trajet aléatoire pour la sélection à la deuxième étape dans les zones de conflit de Mogadishu ; l'application d'une méthodologie géographique d'échantillonnage aléatoire en grappes à une étude des populations pastorales et nomades dans l'est de l'Éthiopie ; la mise en œuvre d'une étude à fréquence élevée sur des téléphones cellulaires afin de mesurer les répercussions socio-économiques de l'Ébola au Sierra Leone et au Libéria. Nous nous intéresserons principalement aux difficultés techniques et de mise en œuvre, à la façon d'y répondre tout en maintenant un plan à probabilités généralisable, et aux leçons retenues pour les applications futures.

[Tuesday June 13/Mardi 13 juin, 16:00-16:30]

Kim P. Huynh (Bank of Canada)

Measuring Cash and Payment Card Usage at the Point-of-Sale for Consumers and Merchants

Mesure de l'utilisation des cartes de débit et de paiement au point de vente pour les consommateurs et les marchands

Measuring cash holdings and usage is particularly difficult due to the anonymous nature of cash. The Bank of Canada has undertaken some surveys of consumer and merchants to address this issue. However, there are several challenges involved such as recruitment of respondents and low response rates. We discuss some methodology employed to address these concerns.

Il est particulièrement difficile de mesurer les liquidités et leur utilisation en raison de la nature anonyme de l'argent comptant. La Banque du Canada a mené des sondages auprès des consommateurs et de marchands pour faire le point sur cette question. Parmi plusieurs difficultés, il y a le recrutement de répondants et le faible taux de réponse. Nous traitons de certaines méthodologies utilisées pour résoudre ces problèmes.

Applied Survey Research with Impact

Recherche appliquée sur les sondages avec un impact

[Tuesday June 13/Mardi 13 juin, 16:30-17:00]

Robert Clark (University of Wollongong, New Zealand), **Steven Johnston** (New Zealand Ministry of Health)

How the Design of the New Zealand Health Survey Contributes to Informed Decision Making in Public Health Policy
Comment la conception de l'Enquête sur la santé de Nouvelle-Zélande contribue à des décisions éclairées en matière de politique de santé publique

The New Zealand Health Survey monitors the health of the nation's adult and child populations. It produces a consistent core set of indicators, and must also reflect emerging policy priorities, and provide regular information on national, demographic, ethnic and regional subgroups. To achieve these goals, the survey is continuously in the field, but reports annually. Its continuous nature allows a set of core questions to be combined with a flexible programme of modules. Data are pooled over different periods to balance timeliness and precision. The sample design oversamples key subpopulations using a novel method which reflects the strengths and weaknesses of the data available for design. Using these mechanisms, the survey supports policy directions including the Smokefree 2025 goal, the Childhood Obesity Plan, the Rheumatic Fever Prevention Programme, and the continuous improvement of the responsiveness and equity of health services.

L'Enquête sur la santé de Nouvelle-Zélande surveille l'état de santé de la population des adultes et des enfants au pays. Tout en produisant un ensemble cohérent d'indicateurs de base, cette enquête doit aussi refléter les priorités émergentes en matière de politique et fournir régulièrement des renseignements sur les sous-groupes nationaux, démographiques, ethniques et régionaux dans le pays. Pour réaliser ces objectifs, l'enquête se poursuit de façon continue sur le terrain, mais le rapport d'enquête est annuel. Cette continuité permet de combiner un ensemble de questions de base à un programme flexible de modules. Les données sont regroupées selon différentes périodes afin d'équilibrer leur actualité et leur précision. Le mode d'échantillonnage sur-échantillonne des sous-populations clés à l'aide d'une nouvelle méthode qui reflète les points forts et les faiblesses des données disponibles pour concevoir l'enquête. Grâce à ces mécanismes, l'enquête appuie l'orientation des politiques, y compris l'objectif de fin du tabagisme (Smokefree) en 2025, la lutte contre l'obésité chez les enfants (the Childhood Obesity Plan), le programme de prévention de la fièvre rhumatismale (Rheumatic Fever Prevention Program) et les améliorations continues du temps de réponse et de l'équité des services de santé.

Causality and Design Causalité et conception

Chair/Président: Mireille E. Schnitzer
Room/Salle: E2 165 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-15:45]

Po Yang (University of Manitoba), **William Li** (University of Minnesota)

Foldover Designs with Column Permutations

Plans de repliement avec permutation de colonnes

Foldover is a follow-up technique used in the design of experiments. Traditional foldover designs are obtained by changing the signs of some columns of an initial design. We further consider performing a column permutation. We investigate when a column permutation results in a combined foldover design that has better G-aberration than the corresponding traditional combined foldover design. Properties of such foldover designs are studied.

Le repliement est une technique de suivi utilisée pour la conception d'expériences. On obtient les plans de repliement traditionnels en changeant les signes de certaines colonnes dans le plan original. Nous proposons d'ajouter une permutation de colonnes et tentons de voir quand une telle permutation donne un plan de repliement combiné avec une meilleure aberration G que le plan de repliement combiné traditionnel correspondant. Nous étudions ici les propriétés de tels plans de repliement.

[Tuesday June 13/Mardi 13 juin, 15:45-16:00]

Katherine Dagnault (University of Toronto), **Olli Saarela** (University of Toronto, Dalla Lana School of Public Health)

A Doubly Robust Estimator for Indirectly Standardized Mortality Ratios

Un estimateur doublement robuste pour les ratios de mortalité indirectement standardisés

Routinely collected administrative and clinical data are increasingly being utilized for comparing quality of care outcomes between hospitals. This problem can be considered in a causal inference framework, as such comparisons have to be adjusted for hospital-specific patient case-mix, which can be done using either an outcome or an assignment model, but is subject to misspecification. It is often of interest to compare the performance of the hospitals against the average level of care in the health care system, using indirectly standardized mortality ratios. A doubly robust estimator makes use of both outcome and assignment model in the case-mix adjustment, requiring only one of these to be correctly specified for valid inferences. We present the causal estimand in indirect standardization in terms of potential outcome variables, propose a doubly robust estimator, and study its properties.

Les données administratives et cliniques collectées de façon routinière sont de plus en plus utilisées pour comparer les issues de qualité des soins entre les hôpitaux. Ce problème peut être considéré dans un cadre d'inférence causale, car ces comparaisons doivent être ajustées pour les groupes clients de patients dans chaque hôpital, ce qui peut être fait en utilisant un modèle d'issue ou d'affectation, mais il est sujet à une mauvaise spécification. Il est souvent intéressant de comparer la performance des hôpitaux par rapport au niveau moyen de soins dans le système de santé, en utilisant des taux de mortalité indirectement standardisés. Un estimateur doublement robuste utilise à la fois le modèle d'issue et d'affectation dans l'ajustement des groupes clients, nécessitant que seulement l'un d'entre eux soit correctement spécifié pour des inférences valides. Nous présentons l'estimateur causal dans la standardisation indirecte en termes de variables d'issues potentielles, proposons un estimateur doublement robuste et étudions ses propriétés.

[Tuesday June 13/Mardi 13 juin, 16:00-16:15]

Shomoita Alam (McGill University), **Erica E.M. Moodie** (McGill University), **David A. Stephens** (McGill University)

Should a Propensity Score be Super? A Comparison of Estimation Methods

Un score de propension devrait-il être super ? Comparaison de méthodes d'estimation

The propensity score (PS) is a tool to eliminate imbalance in the distribution of confounding variables between treatment groups. Previous work (Pirracchio et al. 2015) suggested that Super Learner (SL), an ensemble method, outperforms logistic regression (LR) in non-linear settings. We investigated wider range of settings of varying complexities to compare the performances of logistic regression, generalized boosted models (GBM) and SL. We estimated the average treatment effect using PS regression, PS matching and inverse probability of treatment weighting (IPTW). We found that SL and LR are comparable in terms of bias, covariate balance and mean squared error, and both outperform GBM, however SL is computationally very expensive. We also found that PS regression is superior to either IPTW or matching. We used two real data examples to demonstrate the performances of the estimators to support our findings, and find LR provides the best balance in two empirical analyses.

Le score de propension (SP) est un outil employé pour éliminer le déséquilibre dans la distribution des variables de confusion entre les groupes de traitement. Des travaux antérieurs (Pirracchio et al., 2015) ont suggéré que Super Learner (SL), une méthode d'ensemble, performe mieux que la régression logistique (RL) dans des contextes non linéaires. Nous avons étudié un éventail plus large de situations de diverses complexités pour comparer les performances de la régression logistique, des modèles amplifiés généralisés (MAG) et SL. Nous avons estimé l'effet moyen du traitement à l'aide de la régression SP, par appariement sur le SP et par pondération par probabilité inverse du traitement. Nous avons trouvé que SL et RL sont comparables en termes de biais, d'équilibre des covariables et d'erreur quadratique moyenne, et tous les deux surpassent MAG, mais SL est très coûteuse en termes de calcul. Nous avons également trouvé que la régression SP est supérieure à la pondération ou égale. Nous avons utilisé deux exemples de données réelles pour démontrer les performances des estimateurs appuyant nos résultats et avons trouvé que la RL fournit le meilleur équilibre dans les deux analyses empiriques.

[Tuesday June 13/Mardi 13 juin, 16:15-16:30]

Lin Liu (University of California, San Diego), **Loki Natarajan** (University of California, San Diego), **Karen Messer** (University of California, San Diego)

Estimation in Model Selection after Imputation

Estimation dans la sélection du modèle après imputation

Missing data is often handled using multiple imputation and the parameters are estimated using Rubin's Rules. If model selection is performed after multiple imputation, we propose bootstrap imputation followed by model selection, then model averaging over the bootstrap distribution. The estimated variance is efficiently computed from the same bootstrap sample as in Efron (2014) and confidence intervals are computed using the normal distribution. We call this bootstrap imputation with Efron's Rules. We compare Efron's Rules estimator to the Rubin's Rules estimator. We also compare a single imputation followed by model selection (the Impute-Select estimator) with bootstrap percentile confidence intervals. Simulation studies show that the model averaged estimators perform better than the Impute-Select. The confidence intervals based on Rubin's Rules can have severe undercoverage. The Efron's Rules estimator improves the performance of normal theory confidence intervals.

Les données manquantes sont souvent traitées par l'imputation multiple et les paramètres sont estimés selon les règles de Rubin. Si la sélection du modèle est faite après une imputation multiple, nous proposons une imputation bootstrap suivie par la sélection du modèle et ensuite de combiner les modèles par rapport à la distribution bootstrap. La variance estimée est calculée efficacement à partir du même échantillon bootstrap que dans Efron (2014), et les intervalles de confiance sont calculés en utilisant la loi normale. Nous appelons ceci l'imputation selon les règles d'Efron. Nous comparons l'estimateur selon les règles d'Efron à l'estimateur selon les règles de Rubin. Nous comparons aussi une imputation unique suivie de la sélection du modèle (l'estimateur imputer-sélectionner) où les intervalles de confiance sont obtenues par la méthode des percentiles bootstrap. Des simulations démontrent que les estimateurs issus des modèles combinés performant mieux que ceux de type imputer-sélectionner. Les intervalles de confiance fondés sur les règles de Rubin peuvent présenter une sous-couverture importante. L'estimateur selon les règles d'Efron améliore la performance des intervalles de confiance fondée sur la loi normale.

Causality and Design Causalité et conception

[Tuesday June 13/Mardi 13 juin, 16:30-16:45]

Rui Hu (MacEwan University), **Doug Wiens** (University of Alberta)

Robust Discrimination Designs over Hellinger Neighbourhoods

Plans de discrimination robuste sur les voisinages de Hellinger

To aid in the discrimination between two, possibly nonlinear, regression models, we study the construction of experimental designs. Considering that each of these two models might be only approximately specified, robust “maximin” designs are proposed. The rough idea is as follows. We impose neighbourhood structures on each regression response, to describe the uncertainty in the specifications of the true underlying models. We determine the least favourable – in terms of Kullback-Leibler divergence – members of these neighbourhoods. Optimal designs are those maximizing this minimum divergence. Sequential, adaptive approaches to this maximization are studied. Asymptotic optimality is established.

Pour aider à la discrimination entre deux modèles de régression, possiblement non linéaires, nous étudions l'élaboration de plans expérimentaux. Considérant que chacun de ces deux modèles peut être seulement spécifié approximativement, des plans « maximin » sont proposés. L'idée générale est la suivante. Nous proposons des structures de voisinage sur chaque réponse de régression pour décrire l'incertitude dans les spécifications des véritables modèles sous-jacents. Nous déterminons les membres de ces voisinages les moins favorables en termes de divergence Kullback-Leibler. Les plans optimaux sont ceux qui maximisent cette divergence minimale. Des approches adaptatives et séquentielles pour cette maximisation sont étudiées. L'optimalité asymptotique est établie.

[Tuesday June 13/Mardi 13 juin, 16:45-17:00]

Saumen Mandal (University of Manitoba), **Monsur Chowdhury** (University of Winnipeg)

Optimal Designs for Estimating Parameters of the Bradley-Terry Model for Paired Comparisons

Plans optimaux pour l'estimation des paramètres du modèle de Bradley-Terry pour des comparaisons par paires

The Bradley-Terry model for paired comparisons has been broadly applied in many areas such as statistics, sports and machine learning. In this work we determine the maximum likelihood estimates of the parameters of the latent variable models such as Bradley-Terry model by using the theory of optimal design. We consider the parameters of the Bradley-Terry model in terms of a set of another parameters which we consider as weights in optimal design. These weights are positive and sum to one. In order to solve this problem, we first consider the likelihood function as our criterion function, and then we determine the optimality conditions in terms of point to point directional derivatives of the likelihood function. We then determine the maximum likelihood estimates using a class of algorithms, indexed by a function which depends on the derivatives of the likelihood function. Finally, we apply this problem to a data set from American League Baseball Teams.

Le modèle de Bradley-Terry pour des comparaisons par paires a largement été appliqué dans de nombreux domaines, comme les statistiques, le sport et l'apprentissage machine. Dans cet article, nous déterminons les estimations des paramètres de modèles de variables latentes par maximum de vraisemblance, comme le modèle de Bradley-Terry, en utilisant la théorie des plans optimaux. Nous considérons les paramètres du modèle de Bradley-Terry selon un autre ensemble de paramètres que nous considérons comme des poids du plan optimal. Ces poids sont positifs et ont pour somme l'unité. Afin de résoudre ce problème, nous considérons tout d'abord les fonctions de vraisemblance comme nos fonctions de critère. Ensuite, nous déterminons les conditions d'optimalité selon les dérivées directionnelles de point à l'autre de la fonction de vraisemblance. Puis, nous déterminons les estimations par maximum de vraisemblance au moyen d'une classe d'algorithmes indexée par une fonction qui dépend des dérivées de la fonction de vraisemblance. Enfin, nous appliquons un jeu de données provenant des équipes de la ligue américaine de baseball.

New Methods in Classification Nouvelles méthodes en classification

Chair/Président: Paul D. McNicholas
Room/Salle: E2 160 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-15:45]

Hyukjun (Jay) Gweon (Waterloo University), **Matthias Schonlau** (University of Waterloo), **Stefan Steiner** (University of Waterloo)

An Overview of Existing and a Novel Approaches to Multi-label Classification

Aperçu d'une nouvelle approche et d'approches existantes de classification multi-étiquettes

Multi-label classification is a supervised learning problem where an observation may be associated with multiple binary (outcome) labels simultaneously. We give an overview over common approaches to multi-label classification and also introduce a new approach as an extension of the nearest neighbor principle. Experiments on benchmark multi-label data sets show that the proposed method on average outperforms other commonly used approaches in terms of classification performance.

La classification multi-étiquettes est un problème d'apprentissage supervisé dans lequel l'observation peut être simultanément associée à plusieurs étiquettes (résultats) binaires. Nous donnons un aperçu des approches courantes de classification multi-étiquettes et présentons également une nouvelle approche qui sert d'extension au principe du plus proche voisin. Des expériences sur des jeux de données de référence multi-étiquettes montrent que la méthode proposée est généralement meilleure que les approches couramment utilisées en termes d'efficacité de classification.

[Tuesday June 13/Mardi 13 juin, 15:45-16:00]

Armin Hatefi (University of Toronto), **Nancy Reid** (University of Toronto), **Mohammad Jafari Jozani** (University of Manitoba), **Omer Ozturk** (Ohio State University)

Estimation and Classification for Finite Mixture Models with Order Statistics

Estimation et classification pour les modèles de mélange fini avec des statistiques d'ordre

We study the problems of maximum likelihood estimation and classification based on different collections of order statistics from finite mixture models. We consider these problems under both labeled and unlabeled collections of order statistics. New missing mechanisms and expectation-maximization algorithms are developed to exploit the structure of observed data in estimation procedures. Various model-based classification criteria are then developed for observed and unobserved data from the underlying FMM. Through simulation studies, we evaluate the performance of estimation and classification methodologies. Finally, the proposed methods are used for a real data analysis.

Nous étudions les problèmes d'estimation par le maximum de vraisemblance et de classification selon différents jeux de statistiques d'ordre issus de modèles de mélange fini. Nous considérons ces problèmes à la fois pour des jeux indexés et non indexés de statistiques d'ordre. Nous développons de nouveaux mécanismes de données manquantes et des algorithmes d'espérance-maximisation pour exploiter la structure des données observées dans les procédures d'estimation. Différents critères de classification basés sur le modèle sont ensuite élaborés pour les données observées et non observées du MMF sous-jacent. Nous évaluons la performance des méthodologies d'estimation et de classification à l'aide d'études de simulation. Enfin, les méthodes proposées sont utilisées pour analyser de vraies données.

[Tuesday June 13/Mardi 13 juin, 16:00-16:15]

Wanhua Su (MacEwan University)

Two Innovative Applications of Classification and Regression Trees

Deux applications innovantes des arbres de classification et de régression

Classification and regression trees (CART) are machine learning methods for constructing prediction

Les arbres de classification et de régression CART sont des méthodes d'apprentissage machine qui permettent la

New Methods in Classification

Nouvelles méthodes en classification

models by recursively partitioning the data space and fitting a simple model within each partition. Two innovative applications of CART were proposed: discretization and interaction detection in fitting regression models. Discretization is to convert a continuous variable to a categorical which is necessary for statistical methods such as analysis of variance (ANOVA), chi-square tests, fitting a multinomial model, and etc. Simulation studies and medical applications are used to illustrate the effectiveness of classification and regression trees in transferring continuous data to categorical and extracting the interactions structure between explanatory variables.

construction de modèles de prévision par partitionnement récursif de l'espace de données et l'ajustement d'un modèle simple dans chaque partition. Nous proposons deux applications innovantes des arbres CART : la discrétisation et la détection d'interactions pour l'ajustement des modèles de régression. La discrétisation permet de convertir une variable continue en une variable catégorielle, ce qui est nécessaire pour les méthodes statistiques comme l'analyse de variance (ANOVA), les tests du chi-carré, l'ajustement d'un modèle multinomial, etc. Nous utilisons des études par simulation et des applications médicales pour illustrer l'efficacité des arbres CART pour convertir des données continues en données catégorielles et extraire la structure d'interactions entre les variables explicatives.

[Tuesday June 13/Mardi 13 juin, 16:15-16:30]

Pingzhao Hu (University of Manitoba), **Md. Mohaiminul Islam** (University of Manitoba), **Yang Wang** (University of Manitoba)

A deep learning-based integrative analysis framework for molecular classification of breast cancer

Analyse intégrative fondée sur le deep learning pour la classification moléculaire du cancer du sein

Classification of molecular subtypes of breast cancer using deep learning poses an exceptional challenge as we have to train a deep neural network (DNN) with hundreds of thousands of parameters using a small dataset with thousands of samples and tens of thousands of features. It is expected that integration of knowledge from multiple data sources measured on the same individuals can overcome the challenge. In this paper, we proposed a number of DNN models to integrate different omics data sets collected from the same set of breast cancer patients for predicting their molecular subtypes. We compared the results of our DNNs with traditional machine learning models such as Support Vector Machine (SVM) and Random Forest (RF). We demonstrated that our DNN models are superior to SVM and RF.

La classification des sous-types moléculaires du cancer de sein en utilisant le deep learning pose un véritable défi puisque nous devons former un réseau de neurones profond (DNN) avec des centaines de milliers de paramètres à l'aide d'un petit jeu de données comprenant des milliers d'échantillons et des dizaines de milliers de caractéristiques. Il est supposé que l'intégration des connaissances provenant de multiples sources de données mesurées sur les mêmes individus peut surmonter cette difficulté. Dans cet article, nous proposons des modèles DNN pour intégrer différents ensemble de données omiques recueillis à partir du même ensemble de patients atteints du cancer du sein pour prédire leurs sous-types moléculaires. Nous comparons les résultats de nos DNNs avec des modèles traditionnels d'apprentissage machine tels que la machine à vecteurs de support (SVM) et la forêt aléatoire (RF). Nous démontrons que nos modèles DNN sont supérieurs aux SVM et RF.

[Tuesday June 13/Mardi 13 juin, 16:30-16:45]

Jian Yang (University of Calgary), **Jingjing Wu** (University of Calgary), **Haocheng Li** (Tom Baker Cancer Centre and University of Calgary), **Xiaolan Feng** (Tom Baker Cancer Centre and University of Calgary)

Minimum Profile Hellinger Distance Estimation for Two-Class Location Models

Estimation de la distance du profil minimal d'Hellinger pour les modèles d'emplacement à deux classes

Minimum Hellinger Distance estimator is obtained by minimizing the Hellinger distance between an assumed parametric model and a nonparametric estimator of a model density. The method receives increasing attention over the past decades due to

L'estimateur de la distance minimale d'Hellinger est obtenu en minimisant la distance d'Hellinger entre un modèle que l'on suppose paramétrique et un estimateur non paramétrique d'un modèle de densité. La méthode reçoit une attention croissante au cours des dernières décennies en raison

New Methods in Classification

Nouvelles méthodes en classification

their asymptotic efficiency and excellent robustness against small deviation from the model. In this work, we propose to use a Minimum Profile Hellinger Distance estimation (MPHD) for the parameters in two-class symmetric location models. Asymptotic normality and the property of robustness of the estimator are discussed and a comparison with MLE is carried out through Monte Carlo simulation studies. This estimation is applied to a breast cancer dataset.

de son efficacité asymptotique et de son excellente robustesse par rapport à de petits écarts face au modèle. Dans ce travail, nous proposons d'utiliser une estimation de la distance du profil minimal d'Hellinger (DPMH) pour les paramètres dans les modèles d'emplacement symétrique à deux classes. Nous discutons la normalité asymptotique et la propriété de robustesse de l'estimateur et nous la comparons à l'EMV par des études de simulation Monte-Carlo. Cette estimation est appliquée à un jeu de données sur le cancer du sein.

[Tuesday June 13/Mardi 13 juin, 16:45-17:00]

Anita Brobbey (University of Calgary), **Tyler Williamson** (University of Calgary), **Samuel Wiebe** (University of Calgary), **Lisa M. Lix** (University of Manitoba), **Tolulope T. Sajobi** (University of Calgary)

Comparison of Multivariate Classifiers in the Presence of Covariates in Repeated Measures Designs

Comparaison de classificateurs multivariés en présence de covariables dans les modèles de mesures répétées

Classification models based on multivariate outcomes are commonly used for discriminating between groups in repeated measures (RM) designs. However the majority of these models rely on the assumption of multivariate normality and rarely incorporate covariate information, which are routinely collected in RM designs. Using Monte Carlo simulation methods, we investigate the impact of time-varying and time-invariant covariates on the misclassification error rate of classical linear discriminant analysis (LDA), LDA based on linear mixed effect regression, a Bayes classifier, logistic regression, and quadratic inference function classifiers. We provide recommendations about using these classifiers in RM designs with multiple covariates.

Les modèles de classification basés sur des réponses multivariées sont couramment utilisés pour discriminer entre les groupes dans des modèles de mesures répétées (MR). Cependant, la majorité de ces modèles reposent sur l'hypothèse d'une normalité multivariée et incorporent rarement l'information des covariables, qui sont recueillies systématiquement dans les modèles de MR. À l'aide de méthodes de simulation de Monte-Carlo, nous étudions l'impact des covariables variables dans le temps et invariables dans le temps sur le taux d'erreur de classement de l'analyse discriminante linéaire (ADL) classique, l'ADL basée sur la régression linéaire mixte, un classificateur de Bayes, la régression logistique et des classificateurs de fonctions quadratiques d'inférence. Nous formulons des recommandations sur l'utilisation de ces classificateurs dans les devis de MR comportant plusieurs covariables.

Insuring, Investing, and Paying Up Assurer, investir et payer

Chair/Président: Cary Chi-Liang Tsai
Room/Salle: E2 155 (EITC)
Abstract/Résumé

[Tuesday June 13/Mardi 13 juin, 15:30-15:45]

Jean-François Bégin (Simon Fraser University)

Deflation Risk and Implications for Life Insurers

Risque de déflation et conséquences pour les assureurs-vie

Life insurers are exposed to deflation risk: falling prices could lead to insufficient investment returns, and inflation-indexed protections could make insurers vulnerable to deflation. In this spirit, this study proposes a market-based methodology for measuring deflation risk based on a discrete framework: the latter accounts for the real interest rate, the inflation index level, its conditional variance, and the expected inflation rate. US inflation data are then used to estimate the model and show the importance of deflation risk. Specifically, the distribution of a fictitious life insurer's future payments is investigated. We find that the proposed inflation model yields higher risk measures than the ones obtained using competing models, stressing the need for dynamic and market-consistent inflation modelling in the life insurance industry.

Les assureurs-vie sont exposés au risque de déflation : une baisse générale du niveau des prix pourrait mener à des rendements insuffisants. De plus, certaines protections offertes par les assureurs pourraient les rendre plus vulnérables à un taux d'inflation négatif. Cette étude propose une méthode basée sur le prix marchant pour mesurer le risque de déflation : le modèle discret proposé tient en compte le taux d'intérêt réel, le niveau d'indice de l'inflation, sa variance conditionnelle et le taux d'inflation prévu. Les données de l'inflation américaines sont ensuite utilisées pour estimer le modèle et pour montrer l'importance du risque de déflation. Plus précisément, la distribution des versements futurs d'un assureur-vie fictif est étudiée. Nous trouvons que le modèle d'inflation proposé donne des mesures de risques plus élevées que celles obtenues à l'aide de modèles concurrents.

[Tuesday June 13/Mardi 13 juin, 15:45-16:00]

Valéry Dongmo Jiongo (Bank of Canada)

Outlier Robust Estimation of the Total Private Cost of Payment Methods for Large Businesses

Estimation robuste des valeurs aberrantes du coût total privé des méthodes de paiement pour les grandes entreprises

The Bank of Canada conducted the 2015 Retailer Survey on the Cost of Payment Methods for two reasons: to estimate the total private and social costs to Canadian retailers of cash and cards and to analyse the efficiency of payment methods. We focus specifically on the private costs to large businesses. The presence of outliers in the sample suggests the need for a robust estimation method. Three robust versions of Royall's (1970) best linear unbiased predictor (BLUP) are computed; these estimators are the naive, the conditional bias (Beaumont, Haziza and Ruiz-Gazen 2013 [BHR]) and the Chambers (1986). BHR propose an adaptive method of choosing the tuning constant. We derive the tuning constant for the Chambers (1986) estimator by minimising a parametric bootstrap es-

La Banque du Canada a mené l'Enquête auprès des détaillants de 2015 sur le coût des méthodes de paiement pour deux raisons : estimer le total des coûts privés et sociaux des détaillants canadiens en espèces et en cartes et analyser l'efficacité des modes de paiement. Nous nous concentrons particulièrement sur les coûts privés pour les grandes entreprises. La présence de valeurs aberrantes dans l'échantillon suggère la nécessité d'une méthode d'estimation robuste. Trois versions robustes du meilleur prédicteur linéaire non biaisé de Royall (1970) sont calculées ; ces estimateurs sont naïfs, le biais conditionnel (Beaumont, Haziza et Ruiz-Gazen 2013 [BHR]) et Chambers (1986). BHR propose une méthode adaptative de choix de la constante de réglage. Nous dérivons la constante de réglage pour l'estimateur de Chambers (1986) en minimisant un estimateur bootstrap paramétrique de l'erreur quadratique moyenne (EQM). Les résultats

Insuring, Investing, and Paying Up Assurer, investir et payer

estimator of the mean squared error (MSE). The results of the total private costs to retailers for cash and cards are compared; the robust Chambers estimator and the conditional bias estimator are also compared.

des coûts privés totaux aux détaillants pour les espèces et les cartes sont comparés; l'estimateur robuste de Chambers et l'estimateur de biais conditionnel sont également comparés.

[Tuesday June 13/Mardi 13 juin, 16:00-16:15]

Suxin Wang (Simon Fraser University), **Yi Lu** (Simon Fraser University), **Barbara Sanders** (Simon Fraser University)

Optimal Investment Strategies and Inter-Generational Risk-Sharing for Target Benefit Pension Plans

Stratégies d'investissement optimales et partage du risque entre générations pour régimes de retraite à prestations cibles

We consider a stochastic model for a target benefit pension fund in continuous time, where the plan member's contributions are set in advance while the pension payments depend on the financial situation of the plan, implying risk-sharing between different generations. The pension fund is invested in both a risk-free asset and a risky asset. In particular, stochastic salary and the correlation between the salary movement and the financial market fluctuation are considered. Using the stochastic optimal control approach, we derive closed-form solutions for optimal investment strategies as well as optimal benefit payment adjustments, which minimize the combination of benefit risk (deviating from the target benefit) and inter-generational transfers. Numerical analysis is presented to illustrate the sensitivity of the optimal strategies to parameters of the financial market and salary rates. We also consider how the optimal benefit changes with respect to different target levels.

Nous considérons un modèle stochastique pour un régime de retraite à prestations cibles en temps continu où les cotisations du membre du régime sont prédéfinies tandis que les paiements de retraite dépendent de la situation financière du régime, ce qui implique un partage du risque entre les générations. Le fonds de pension est investi à la fois dans un actif sans risque et dans un actif risqué. En particulier, la rémunération stochastique et la corrélation entre le mouvement salarial et les fluctuations du marché financier sont considérées. En utilisant l'approche de contrôle stochastique optimal, nous dérivons des solutions à formes fermées pour des stratégies de placement optimales ainsi que pour des ajustements de versements de prestations optimaux, ce qui minimise la combinaison risque bénéfice (s'écartant de la prestation cible) et les transferts entre générations. Une analyse numérique est présentée pour illustrer la sensibilité des stratégies optimales aux paramètres du marché financier et aux taux de rémunération. Nous examinons aussi la façon dont les prestations optimales changent par rapport aux différents niveaux cibles.

[Tuesday June 13/Mardi 13 juin, 16:15-16:30]

Francois Watier (Université du Québec à Montréal), **Yassin El Kasmi** (Université du Québec à Montréal)

Constrained Mean-Variance Portfolio in a Market with Stochastic Volatility

Portefeuille à variance moyenne sous contraintes dans un marché à volatilité stochastique

In this talk, we will study a specific stochastic control problem known as the mean-variance investment problem where the portfolio strategy is subject to a no short-selling constraint. More specifically, under a Black-Scholes type financial market modelization where the stock's volatility follows a Heston diffusion process, we will construct an optimal portfolio through carefully chosen adjoint processes or backward stochastic differential equations.

Cet exposé porte sur un problème de contrôle stochastique précis appelé problème d'investissement de moyenne-variance où la stratégie d'investissement est assujettie à une contrainte d'interdiction de vente à découvert. Selon une modélisation du marché financier de type Black-Scholes où la volatilité des actions suit un processus de diffusion Heston, nous élaborons plus précisément un portefeuille optimal à l'aide de processus adjoints soigneusement choisis ou d'équations différentielles stochastiques rétrogrades.

[Tuesday June 13/Mardi 13 juin, 16:30-16:45]

Etienne Marceau (Université de Laval), **Hélène Cossette** (Université Laval)

Insuring, Investing, and Paying Up Assurer, investir et payer

Portfolio of Exchangeable Risks: Aggregation with Partial Information on Dependence
Portefeuille de risques échangeables : agrégation avec information partielle sur la dépendance

In this paper, we consider the computation of risk measures, such as the VaR and the TVaR, for a portfolio of exchangeable risks assuming that the marginal distributions are known but that the dependence structure is partially known. In our approach, we compute lower and upper bounds on risk measures, such as the VaR and the TVaR, for the aggregate risk of the portfolio. Our approach is based on some known moments and on stochastic orders. Numerical examples are provided to illustrate our proposed approach.

Dans cet article, nous examinons le calcul de mesures du risque, telles que le VaR et le TVaR, pour un portefeuille de risques échangeables en présumant que les distributions marginales sont connues mais que la structure de dépendance est seulement connue partiellement. Dans notre approche, nous calculons les bornes inférieures et supérieures des mesures du risque, telles que le VaR et le TVaR, pour le risque agrégé du portefeuille. Notre approche est fondée sur certains moments connus et sur des ordres stochastiques. Des exemples numériques sont présentés pour illustrer l'approche proposée.

[Tuesday June 13/Mardi 13 juin, 16:45-17:00]

Yi Lu (Simon Fraser University), **Shuanming Li** (University of Melbourne)

On the Moments and the Distribution of Aggregate Discounted Claims in a Markovian Environment
Moments et répartition des déclarations de sinistres totales actualisés dans un environnement markovien

This paper studies the moments and the distribution of the aggregate discounted claims in a Markovian environment, where the claim arrivals, claim amounts and forces of interest (for discounting) are influenced by an underlying Markov process representing external environment changes. Specifically, we assume that claims occur according to a Markovian arrival process (MAP). A recursive formula is derived for the moments of the aggregate discounted claims occurred in certain states. We also study two types of covariances, the covariance of the aggregate discounted claims occurred in any two subsets of the state space and the covariance of aggregate discounted claims occurred at two different times. The distribution of the aggregate discounted claims occurred in certain states by any specific time is also investigated. For a two-state Markov-modulated model, numerical results are presented.

Cet article traite des moments et de la répartition des déclarations de sinistres actualisés totales dans un environnement markovien, dans lequel l'arrivée et les montants des déclarations de sinistre, ainsi que les intensités d'intérêt (pour l'actualisation) sont influencés par un processus markovien sous-jacent représentant les changements environnementaux externes. Plus précisément, nous assumons que les déclarations de sinistre surviendront en fonction d'un processus d'arrivée de Markov. Une formule récursive est obtenue pour les moments de déclarations de sinistres actualisés totales qui ont été effectuées dans certains états. Nous étudions également deux types de covariables, la covariable des déclarations de sinistres actualisés qui se sont produites dans deux sous-ensembles d'espace d'états et la covariance des déclarations de sinistres actualisés totales qui sont survenues à deux moments différents. De plus, nous analysons la répartition des déclarations de sinistres actualisés totales qui se sont produites dans certains états, à un moment précis. Enfin, nous présentons des résultats numériques pour un modèle modulé de Markov à deux états.

CRM-SSC Prize in Statistics Address
Allocution de la récipiendaire du Prix CRM-SSC en statistique

Chair/Président: Hugh Chipman
Organizer/Responsable: Alejandro Murua
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 08:40-09:45]

Lei Sun (University of Toronto)

The Power of Simple Statistical Techniques in the Era of Big and Complex Data: Some Recent Examples from Genetic Association Studies

La puissance des techniques statistiques simples à l'ère des données volumineuses et complexes : quelques exemples récents d'études d'association génétique

In many scientific studies, different statistical tests are proposed with competing claims about method performance. Discussions of optimality rely on assumptions, and an optimal test is often impossible in multi-dimensional parameter space. For example, in studying association between multiple genetic variants and a complex human trait, we show that most existing methods belong to linear or quadratic class of tests, each being powerful only in sub-spaces. In another, location- or scale-test is preferred depending on the presence of main or interaction effect. In both cases, we show that the two classes of tests are asymptotically independent of each other under the global null hypothesis. Thus, we can use Fisher's method to derive a new class of robust tests and obtain asymptotic distribution; this is desirable when analyzing big data. The talk is based on work from Derkach, Lawless, Sun (2014, Statistical Science), Soave et al. (2015, AJHG), and Soave and Sun (2017, Biometrics).

Dans de nombreuses études scientifiques, les auteurs proposent divers tests statistiques dont ils vantent la performance. Mais toute discussion quant à l'optimalité de telle ou telle méthode repose sur des hypothèses qu'il est impossible de tester en espace de paramètres multidimensionnel. Ainsi, s'agissant de l'association entre plusieurs variantes génétiques et une caractéristique humaine complexe, nous montrons que la plupart des méthodes existantes appartiennent à la classe de tests linéaires ou quadratiques, qui ne sont puissantes que dans des sous-espaces. Dans d'autres cas, on préférera plutôt un test de position ou un test d'échelle, en la présence d'un effet principal ou d'un effet d'interaction. Dans les deux cas, nous montrons que les deux classes de tests sont asymptotiquement indépendantes sous l'hypothèse nulle globale. Nous pouvons donc utiliser la méthode de Fisher pour dériver une nouvelle classe de tests robustes et obtenir une distribution asymptotique, ce qui est souhaitable pour l'analyse de données volumineuses. Cette présentation s'appuie sur les travaux de Derkach, Lawless, Sun (2014, Statistical Science), Soave et al. (2015, AJHG) et Soave et Sun (2017, Biometrics).

Showcase: Graduate Student Research in Actuarial Sciences
Vitrine : Recherches d'étudiants gradués en sciences actuarielles

Chair/Président: Mathieu Boudreault
Organizer/Responsable: Mathieu Boudreault
Room/Salle: E2 150 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:42]

Hongcan Lin (University of Waterloo), **David Saunders** (University of Waterloo), **Chengguo Weng** (University of Waterloo)

Optimal Investment Strategies for Participating Contracts

Stratégies d'investissement optimales pour contrats avec participation

Participating contracts are popular insurance policies, in which the payoff to a policyholder is linked to the performance of a portfolio managed by the insurer. We consider the portfolio selection problem of an insurer that offers participating contracts and has an S-shaped utility function. Applying the martingale approach, closed-form solutions are obtained. The resulting optimal strategies are compared with portfolio insurance hedging strategies (CPPI and OBPI). We also study numerical solutions of the portfolio selection problem with constraints on the portfolio weights.

Les contrats avec participation sont des polices d'assurance populaires dont le rendement dépend du rendement d'un portefeuille géré par l'assureur. Nous étudions le problème de sélection de portefeuille pour un assureur qui propose des contrats avec participation avec une fonction d'utilité en forme de S. Nous obtenons des solutions sous forme fermée à l'aide de l'approche martingale. Nous comparons les stratégies optimales résultantes aux stratégies de couverture d'assurance de portefeuille (CPPI et OBPI). Nous étudions aussi des solutions numériques du problème de sélection de portefeuille avec contraintes de pondération.

[Wednesday June 14/Mercredi 14 juin, 10:42-11:04]

Ihsan Chaoubi (Université Laval), **Hélène Cossette** (Université Laval), **Etienne Marceau** (Université Laval)

Archimedean Copulas through Multivariate Gamma Distributions

Copules archimédiennes par distributions gamma multivariées

Inspired by Marshall-Olkin's approach, multivariate distributions can be constructed through the use of exponential mixtures. In this paper, we propose an alternative hierarchical Archimedean copula, obtained from multivariate survival functions of multivariate mixed exponential distributions. The key element of our construction is that the latter are defined with a vector of mixing random variables, which follows a multivariate gamma distribution such as Kibble's bivariate gamma distribution. After presenting the construction technique, properties of this new family of copulas are investigated, simulation algorithms are provided and illustrative examples are given. Risk aggregation and capital allocation under this newly proposed dependence structure are also examined.

Inspirés par l'approche de Marshall-Olkin, nous construisons des distributions multivariées à l'aide de mélanges exponentiels. Dans cette présentation, nous proposons une copule archimédienne hiérarchique alternative, obtenue à partir de fonctions de survie multivariées de distributions exponentielles mixtes multivariées. L'élément clé de notre construction est que ces dernières sont définies à l'aide d'un vecteur de mélange de variables aléatoires, qui suit une distribution gamma multivariée comme la distribution gamma bivariée de Kibble. Après avoir présenté la technique de construction, nous explorons les propriétés de cette nouvelle famille de copules et proposons des algorithmes de simulation et des exemples illustratifs. Nous étudions aussi l'agrégation des risques et l'allocation du capital sous cette nouvelle structure de dépendance.

[Wednesday June 14/Mercredi 14 juin, 11:04-11:26]

Lidan (Lucy) Zhang (Simon Fraser University), **Yi Lu** (Simon Fraser University)

A Multi-State Model for a Life Insurance Product with Integrated Health Rewards Program

Showcase: Graduate Student Research in Actuarial Sciences

Vitrine : Recherches d'étudiants gradués en sciences actuarielles

Modèle multi-états pour produits d'assurance-vie avec programme de récompenses santé intégré

With the prevalence of chronic diseases that account for a significant portion of deaths, a new approach to life insurance has emerged to address this issue. The new approach integrates health rewards programs with life insurance products; the insured are classified by fitness statuses according to their level of participation and would get premium reductions at the superior statuses. We introduce a Markov chain process to model the dynamic transition of the fitness statuses, which are linked to corresponding levels of mortality risks reduction. We then embed this transition process into a stochastic multi-state model to describe the new life insurance product. Formulas are given for calculating its benefit, premium, reserve and surplus. These results are compared with those of the traditional life insurance. Numerical examples are given for illustration.

Avec la prévalence de maladies chroniques qui contribuent à une proportion importante des décès, une nouvelle approche de l'assurance-vie se dessine. Cette approche intègre des programmes de récompenses santé aux produits d'assurance-vie; les assurés sont classés par état de forme physique, en fonction de leur niveau de participation, et bénéficient de réductions de primes pour les états supérieurs. Nous introduisons un processus de chaînes de Markov pour modéliser la transition dynamique des états de forme physique, qui sont liés à des niveaux correspondants de réduction du risque de mortalité. Nous intégrons ensuite ce processus de transition dans un modèle multi-états stochastique pour décrire le nouveau produit d'assurance-vie. Nous proposons des formules pour calculer les prestations, les primes, la réserve et l'excédent du produit. Nous comparons ces résultats à ceux des produits d'assurance-vie traditionnels. Nous donnons des exemples numériques à titre d'illustration.

[Wednesday June 14/Mercredi 14 juin, 11:26-11:50]

Ruixi Zhang (University of Western Ontario), **Kristina Sendova** (University of Western Ontario)

On Total Dividends under a Threshold Strategy

Total des dividendes sous une stratégie de seuil

We consider the total dividends paid prior to ruin under a threshold strategy. We investigate a class of Sparre-Andersen risk processes and its associated class of delayed risk processes with rational-distributed inter-claim times. With the presence of a constant dividend barrier, we present a structure of integro-differential equation, which is applicable to numerous quantities such as the Gerber-Shiu function, the maximum surplus, etc. Finally, we study the fraction of time of dividends payment prior to ruin. This is of interest to an investor who wants to know how often to expect dividends.

Nous examinons le total des dividendes payés jusqu'à la ruine sous une stratégie de seuil. Nous étudions une classe de processus de risque de Sparre-Andersen et la classe associée de processus de risques différés avec des temps entre réclamations rationnellement distribués. En présence d'une barrière de dividendes constante, nous présentons une structure d'équations intégrales et différentielles, qui est applicable à diverses quantités comme la fonction Gerber-Shiu, l'excédent maximum, etc. Enfin, nous étudions la fraction de temps de versement de dividendes jusqu'à la ruine. Cela peut intéresser les investisseurs qui veulent savoir à quelle fréquence ils recevront des dividendes.

Technology in the Classroom La technologie en salle de classe

Chair/Président: Anne-Sophie Charest
Organizer/Responsable: Anne-Sophie Charest
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:50]

Jim Stallard (University of Calgary), **Scott Robison** (University of Calgary)

A More Dynamic Lab Experience: Implementation of Weekly Automated and Teaching Assistant Supported Lab Exercises to Increase Student Engagement through Small-Step Learning
Expérience de laboratoire plus dynamique : mise en œuvre d'exercices en laboratoire hebdomadaires automatisés et soutenus par un assistant à l'enseignement afin d'augmenter l'engagement des étudiants par l'apprentissage par petites étapes

The lab component provides an intimate and stimulating setting for a student to interact with lectured content. The lab is designed to supplement the lecture material and mature student comprehension. However, too often labs have mutated into rarely attended communal homework periods that actually diminish every pedagogical underpinning associated with higher education. The focus is to present the findings of a dynamic statistics lab experience, designed to refresh and re-establish the noble venture of supplemental learning. A multi-staged weekly seminar that compels preparation, requires participation, enforces comprehension, and evaluates retention. This approach utilizes small-step-learning to replace the anxiety of cumulative lab examination. We will also display the open-source software integration of WebWork and The R Project for Statistical Programming to alleviate additional resource overhead.

Le laboratoire offre un cadre intime et stimulant à l'étudiant et lui permet d'interagir avec le contenu enseigné. Le laboratoire est conçu pour être le complément du matériel d'enseignement et favoriser la compréhension de l'étudiant. Cependant, les laboratoires se sont trop souvent transformés en pièces de travail communes rarement occupées, ce qui du coup diminue toute pédagogie de l'enseignement supérieur. Nous présentons principalement les conclusions statistiques d'une expérience de laboratoire dynamique, conçue pour rafraîchir et rétablir la noble entreprise de la formation complémentaire. Un séminaire hebdomadaire en plusieurs étapes qui oblige à faire des préparations, nécessite une participation, impose la compréhension et évalue la rétention. Cette approche utilise l'apprentissage en petites étapes pour réduire l'anxiété générée par l'accumulation de tests en séances de laboratoire. Nous présenterons également l'intégration du logiciel ouvert de WebWork dans le projet R de programmation statistique afin de réduire les coûts indirects de ressources supplémentaires.

[Wednesday June 14/Mercredi 14 juin, 10:50-11:20]

Jenna G. Tichon (University of Manitoba)

Educreations: Video Tools for Statistical Teaching
Educreations : outils vidéo pour l'enseignement des statistiques

Educreations is a free downloadable app for the iPad that allows instructors to make narrated videos as they write out problems with a stylus pen on the iPad screen. The created videos can then be uploaded for student access on a website. The ability for students to access videos on their own time and review questions other than their own notes can be a valuable instructional tool for both distance education and regular in-class students. In this session, I will be talking about ways the program can be used effectively inside and outside of

Educreations est une application téléchargeable gratuite pour les iPad. Elle permet aux enseignants de faire des vidéos commentées à mesure qu'ils écrivent des problèmes avec un stylet sur l'écran d'un iPad. Les vidéos créées peuvent être téléchargées sur un site Web pour diffusion auprès des étudiants. Cette application, permettant aux étudiants d'accéder aux vidéos quand ils le souhaitent et d'examiner des sujets en dehors de leurs notes de cours, peut représenter un outil d'enseignement de valeur à la fois pour ceux qui suivent des cours à distance et en classe. Pendant cette séance, j'examinerai les façons dont le programme peut être utilisé efficacement

Technology in the Classroom

La technologie en salle de classe

the classroom, basics of creating videos and how to make them accessible to your students. Though not necessary, feel free to bring along an iPad if you would like to follow along with the instructions.

en classe ou à distance, les éléments de base pour créer des vidéos et la façon de les rendre accessibles aux étudiants. Même si ce n'est pas nécessaire, sentez-vous à l'aise d'amener un iPad si vous souhaitez suivre les instructions pendant la séance.

[Wednesday June 14/Mercredi 14 juin, 11:20-11:50]

Thierry Duchesne (Université Laval)

Automated Homework Generation and Grading with R

Génération et correction automatisées de devoir avec R

In this talk I will show how one can use R to simulate personalized assignments and to automatically grade them. This is particularly useful if one wants to make sure that students in large introductory statistics courses do their homework individually. We have successfully implemented this strategy in our introductory statistics course for engineers that is taken by about 400 students. In this course students are assigned two homework for which they have to analyze their own individual randomly generated datasets. A few R scripts generate the datasets and take care of the grading.

Dans cet exposé, je vais montrer comment on peut utiliser R pour simuler des devoirs personnalisés et les corriger automatiquement. Cela est particulièrement utile si l'on veut s'assurer que les élèves des cours de statistique offerts à de grands groupes font leurs devoirs individuellement. Nous avons mis en œuvre avec succès cette stratégie dans notre cours d'introduction à la statistique pour les ingénieurs, cours suivi par environ 400 étudiants. Dans ce cours, les élèves reçoivent deux devoirs pour lesquels ils doivent analyser leurs propres jeux de données générés au hasard. Quelques scripts R génèrent les jeux de données et s'occupent de la correction.

Imputation of Multivariate Missing Data Imputation de données manquantes multivariées

Chair/Président: Matthias Schonlau
Organizer/Responsable: Caren Hasler
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:50]

Jae Kwang Kim (Iowa State University), **Shu Yang** (North Carolina State University)

Predictive Mean Matching Imputation in Survey Sampling

L'imputation par appariement d'après la moyenne prévisionnelle en échantillonnage

Predictive mean matching imputation is popular for handling item nonresponse in survey sampling. We study the asymptotic properties of predictive mean matching for mean estimation. Moreover, the conventional bootstrap inference for matching estimators with fixed matches has been shown to be invalid due to the nonsmoothness nature of the matching estimator. We propose asymptotically valid replication variance estimation for predictive mean matching estimators. The key strategy is to construct replicates of the estimator directly based on linear terms of the estimator, instead of individual records of variables. Extension to nearest neighbor imputation is also discussed. A simulation study confirms that the new procedure provides confidence intervals with correct coverage properties.

L'imputation par appariement d'après la moyenne prévisionnelle est populaire dans le traitement de la non-réponse partielle en échantillonnage. Nous étudions les propriétés asymptotiques de l'appariement d'après la moyenne prévisionnelle pour l'estimation de la moyenne. De plus, il a été démontré que l'inférence bootstrap des estimateurs d'appariement avec des appariements fixés n'était pas valide en raison de la nature non lisse de l'estimateur d'appariement. Nous proposons une estimation de la variance par répliques asymptotiquement valide pour les estimateurs d'appariement d'après la moyenne prévisionnelle. La principale stratégie consiste à construire des répliques de l'estimateur directement en fonction des termes linéaires de l'estimateur plutôt que des enregistrements au sein des variables concernées. Nous discutons également de l'extension à l'imputation par le plus proche voisin. Une étude de simulation confirme que la nouvelle procédure offre des intervalles de confiance avec des propriétés de couverture adéquates.

[Wednesday June 14/Mercredi 14 juin, 10:50-11:20]

Louis-Paul Rivest (Université Laval), **Caren Hasler** (University of Toronto), **Radu Craiu** (University of Toronto)

Vine Copulas for Imputation of Monotone Non-response

Les copules en vigne pour l'imputation d'une non-réponse monotone

Monotone patterns of non-response may occur in longitudinal studies. When the measured variables are dependent it is beneficial to use their joint statistical model to impute the missing values. We propose to use vine copulas to factorize the density of the observed variables into a cascade of bivariate copulas that yield a flexible model of their joint distribution. The structure of the vine depends on the non-response pattern. We build on the work of Aas et al. (2009) and propose a method to select the model, to estimate the parameters of the bivariate copulas of the selected model, and to impute using the constructed model. The imputed values are drawn from the conditional distribution of the

La non-réponse monotone est fréquente dans les études longitudinales. Lorsque les variables à l'étude sont dépendantes il est utile d'utiliser leur distribution conjointe pour imputer les valeurs manquantes. Nous proposons d'utiliser des copules vignes pour factoriser la densité des variables observées en une cascade de copules bivariées, ce qui donne une distribution très flexible pour leur loi conjointe. La structure de la vigne dépend du type de non-réponse. En nous inspirant des travaux de Aas et al. (2009), nous proposons des méthodes pour estimer les paramètres des copules bivariées retenues et pour faire l'imputation à partir du modèle final. Les valeurs imputées sont générées à partir de la distribution conditionnelle des variables manquantes sachant celles qui ont été observées. Un exemple numérique à l'aide de données

Imputation of Multivariate Missing Data Imputation de données manquantes multivariées

missing values, given the observed data. An example using the United Kingdom Labour Force Survey quarterly data illustrates the proposed methodology.

trimestrielles d'emploi au Royaume-Uni illustre la procédure.

[Wednesday June 14/Mercredi 14 juin, 11:20-11:50]

Yves Tillé (Université de Neuchâtel), **Audrey-Anne Vallée** (Université de Neuchâtel)

Balanced Imputation for Swiss Cheese Nonresponse

Imputation équilibrée pour la non-réponse en fromage suisse

Swiss cheese nonresponse refers to the case where all the variables can contain missing values in a general pattern even if this wording is abuse since most of the Swiss cheeses do not have holes. However, in the case of Swiss cheese nonresponse, it is not possible to consider that some variables are known for every units and can thus be used as auxiliary variables. We propose a new technique of random donor imputation. The method is based on the establishment of consistency principles. A nonrespondent and its donor should be close to each other. The same donor must be used for all missing variables of a unit. Moreover we impose that, if we were imputing the known variables of a nonrespondent by the values of the donor, the totals of these variables must remain the same. The procedure is based on the calculation of imputation probabilities. Next, the donors are selected randomly in such a way that the constraints are satisfied.

La non-réponse en fromage suisse se réfère au cas où toutes les variables peuvent contenir des valeurs manquantes, même si cette appellation est un peu abusive puisque la plupart des fromages suisses n'ont pas de trous. Cependant, dans le cas de la non-réponse en fromage suisse, il n'est pas possible de considérer que certaines variables sont connues pour chaque unité et peuvent donc être utilisées comme variables auxiliaires. Nous proposons une nouvelle technique d'imputation aléatoire par donneurs. La méthode repose sur l'établissement de principes de cohérence. Un non-répondant et son donneur doivent être proches les uns des autres. Le même donneur doit être utilisé pour toutes les variables manquantes d'une unité. De plus, nous imposons que, si l'on impute les variables connues d'un non-répondant par les valeurs du donneur, les totaux de ces variables doivent rester identiques. La procédure est basée sur le calcul de probabilités d'imputation. Ensuite, les donneurs sont choisis au hasard de telle sorte que les contraintes sont satisfaites.

Optimal Design of Experiments and Inference

Conception optimale d'expériences et inférence

Chair/Président: Yanqing Yi
Organizer/Responsable: Yanqing Yi
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:50]

Michael Wallace (University of Waterloo), Erica E.M. Moodie (McGill University), David A. Stephens (McGill University)

Dynamic Treatment Regimes via Reward Ignorant Modelling

Régimes de traitement dynamiques par modélisation aveugle au résultat

Personalized medicine optimizes patient outcome by tailoring treatments to patient-level characteristics. This approach is formalized by dynamic treatment regimes (DTRs): decision rules that take patient information as input and output recommended treatment decisions. The DTR literature has seen the development of increasingly sophisticated causal inference techniques, which attempt to address the limitations of our typically observational datasets. We note, however, that in practice most patients should receive optimal or near-optimal treatment, and so the outcome used as part of a typical DTR analysis does not provide much extra information. In light of this, we propose reward ignorant modelling: ignoring the outcome and eliciting an optimal DTR by regressing the observed treatment on relevant covariates, as in a more standard analysis. We present some early results investigating this concept, and analysis of data from the International Warfarin Pharmacogenetics Consortium.

La médecine personnalisée optimise le sort du patient en personnalisant le traitement en fonction de caractéristiques propres au patient. Cette approche est formalisée par les régimes de traitement dynamiques (RTD) : des règles de décision qui utilisent l'information patient comme données et produisent des décisions de traitement recommandé. Les recherches sur les RTD ont produit des techniques d'inférence causale de plus en plus sophistiquées en réponse aux limites de nos ensembles de données d'observation. Nous notons, toutefois, que dans la pratique la plupart des patients devraient recevoir un traitement optimal ou quasi-optimal, si bien que l'issue utilisée dans le cadre de l'analyse de RTD typique ne produit pas beaucoup d'informations supplémentaires. Dans ce contexte, nous proposons une modélisation aveugle à l'issue : nous ne tenons pas compte de l'issue, mais obtenons plutôt un RTD optimal par régression du traitement observé sur les co-variables pertinentes, comme dans une analyse plus standard. Nous présentons quelques résultats préliminaires sur ce concept et une analyse de données tirées de l'International Warfarin Pharmacogenetics Consortium.

[Wednesday June 14/Mercredi 14 juin, 10:50-11:20]

Nancy Flournoy (University of Missouri)

Issues with Inference in Adaptive Optimal Design

Problèmes d'inférence concernant les plans adaptatifs optimaux

Informative designs are increasingly common in experiments with sequential accrual of subjects, such as when observed responses are used to estimate an optimal design for the next cohort, making early stopping decisions for safety or efficacy, or re-estimating sample size. Typically, estimates following informative designs are not normally distributed asymptotically, even assuming a normal model with unknown mean. Instead of ignoring ancillary processes, inference can be based on fully unconditional probabilities or conditional on an ancillary statis-

Les plans informatifs sont de plus en plus courants dans les expériences avec constitution séquentielle de l'échantillon, par exemple lorsque les réponses observées sont utilisées pour estimer un plan optimal pour la cohorte suivante, pour décider de l'arrêt anticipé pour raisons de sécurité ou d'efficacité, ou encore pour ré-estimer la taille d'échantillon. Généralement, les estimations qui suivent des plans informatifs ne suivent pas une distribution asymptotiquement normale, même dans un modèle normal avec moyenne inconnue. Plutôt que d'ignorer les processus auxiliaires, on peut fonder l'inférence sur des probabilités non conditionnelles ou

Optimal Design of Experiments and Inference Conception optimale d'expériences et inférence

tic. Indeed, norming parameter estimates by measures of information other than Fisher information will often produce asymptotic normal random variables. We present these alternatives explicitly for a two-stage adaptive optimal design in which data accumulated in the first stage are used to select the design for a second stage. Lane & Flournoy JPS (2012). Lane, Yao & Flournoy JSPI (2014). Lane Wang & Flournoy mODa 11 (2016).

conditionnelles à une statistique auxiliaire. En effet, normer des estimations de paramètres par des mesures d'information autres que l'information de Fisher produit souvent des variables aléatoires asymptotiquement normales. Nous présentons ces alternatives explicitement pour un plan adaptatif optimal à deux étapes où les données accumulées dans la première étape sont utilisées pour sélectionner le plan de la deuxième étape. Lane & Flournoy JPS (2012). Lane, Yao & Flournoy JSPI (2014). Lane Wang & Flournoy mODa 11 (2016).

[Wednesday June 14/Mercredi 14 juin, 11:20-11:50]

Xiaojian Xu (Brock University), **Sanjoy Kumar Sinha** (Carleton University)

Model-Robust Designs for Generalized Linear Mixed Models with Possible Misspecification

Plans robustes vis-à-vis du modèle pour modèles mixtes linéaires généralisés avec erreurs de spécification éventuelles

Generalized linear mixed models (GLMMs) are commonly used for analyzing clustered or correlated data with non-continuous responses including longitudinal data or repeated measurements. In this paper, we discuss the impact of different types of model misspecification on the estimation of regression coefficients in a GLMM and explore the construction methods of model-robust designs for GLMMs against such possible misspecification. We first derive the asymptotic distribution of the maximum likelihood estimators of the fixed effect parameters when imprecision appears in the assumed linear predictor. Then, we investigate the techniques for sequentially designing experiments where the values of the explanatory variables in a GLMM can be chosen in an optimal and robust way. Design problems in a general setting of GLMMs are addressed yet with an emphasis of their application to longitudinal data. Finally, the performance of our proposed designs is assessed by some simulation studies.

Les modèles mixtes linéaires généralisés (GLMMs) sont communément utilisés pour l'analyse de données en grappe ou corrélées avec réponses non continues, par exemple les données longitudinales ou les mesures répétées. Dans cette présentation, nous discutons de l'impact de différents types d'erreurs de spécification de modèles sur l'estimation des coefficients de régression dans un modèle GLMM et nous explorons les méthodes de construction de plans de GLMM robustes vis-à-vis du modèle pour compenser ce risque d'erreur de spécification. Nous commençons par dériver la distribution asymptotique des estimateurs du maximum de vraisemblance des paramètres des effets fixes, lorsque l'imprécision apparaît dans le prédicteur linéaire présumé. Ensuite, nous examinons les techniques de conception séquentielle d'expériences pour lesquelles les valeurs des variables explicatives d'un GLMM peuvent être choisies de manière optimale et robuste. Nous étudions les problèmes de conception dans le contexte général des GLMM, mais en mettant l'accent sur leur application aux données longitudinales. Enfin, nous évaluons la performance des plans proposés par des études de simulation.

Data Fusion, Network Meta-Analysis, and Causal Inference Fusion de données, méta-analyse en réseau et inférence causale

Chair/Président: Mireille E. Schnitzer and Russell Steele

Organizer/Responsable: Mireille E. Schnitzer

Room/Salle: E3 270 (EITC)

Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:42]

Elias Bareinboim (Purdue University)

Causal Inference from Big Data: Theoretical Foundations and the Data-Fusion Problem

Inférence causale à partir de données volumineuses : les fondements théoriques et le problème de la fusion des données

In this paper, we summarize some of the latest results in the field of causal inference that are related to big data. In particular, we address the problem of data-fusion – piecing together multiple datasets collected under heterogeneous conditions (i.e., different populations, regimes, and sampling methods) so as to obtain valid answers to queries of interest. The availability of multiple heterogeneous datasets presents new opportunities to big data analysts since the knowledge that can be acquired from combined data would not be possible from any individual source alone. However, the biases that emerge in heterogeneous environments require new analytical tools. Some of these biases, including confounding, sampling selection, and cross-population biases, have been addressed in isolation, largely in restricted parametric models. We here present a general, non-parametric framework for handling these biases and, ultimately, a theoretical solution to the problem of data-fusion in causal inference tasks.

Dans cet article, nous résumons quelques résultats récents dans le domaine de l'inférence causale qui concernent les données volumineuses. Nous adressons le problème de fusion de données – nous regroupons plusieurs ensembles de données recueillies dans des conditions hétérogènes (populations, régimes et méthodes d'échantillonnage différents) dans le but d'obtenir des réponses valides à nos questions d'intérêt. La disponibilité des ensembles de données hétérogènes offre de nouvelles opportunités aux analystes de données volumineuses étant donné que les connaissances qui peuvent être acquises à partir de données combinées ne seraient pas disponibles à partir d'une source individuelle. Par contre, les biais qui émergent dans des environnements hétérogènes nécessitent de nouveaux outils analytiques. Certains de ces biais, dont la confusion, la sélection d'échantillons et les biais inter-populations, ont été adressés de manière isolée, en grande partie dans des modèles paramétriques restreints. Nous présentons ici un cadre général et non paramétrique pour traiter ces biais et, ultimement, une solution théorique au problème de la fusion de données dans les travaux d'inférence causale.

[Wednesday June 14/Mercredi 14 juin, 10:42-11:04]

Russell Steele (McGill University), **Mireille E. Schnitzer** (Université de Montréal), **Ian Shrier** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University)

Breaking the Myth of Breaking Randomization: A Causal Examination of Arm-Based Meta-Analysis

Détruire le mythe de la détérioration la randomisation : un examen causal de la méta-analyse à volets

In the analysis of multi-arm randomized trials, methods for pooling data across trials belong to one of two broad classes. The first class of methods consists of contrast-based estimators that estimate the contrast in treatment effect for each pair of treatment levels and then pool across the estimated contrasts. The second class encompasses arm-based methods that contrasts marginal estimates for each treatment arm. Leading researchers have assailed arm-based methods under the broad criticism of

Pour l'analyse des essais aléatoires à volets multiples, les méthodes de regroupement des données provenant de divers essais appartiennent à l'une de deux grandes catégories. La première fait appel à des estimateurs fondés sur les contrastes qui évaluent les contrastes de l'effet d'un traitement pour chaque paire de niveaux de traitement et regroupent ensuite les contrastes estimés. La seconde englobe des méthodes à volets qui comparent les estimations marginales pour chaque volet d'un traitement. Des chercheurs émérites ont pourfendu les méthodes à volets, critiquant surtout la « détérioration de

Data Fusion, Network Meta-Analysis, and Causal Inference Fusion de données, méta-analyse en réseau et inférence causale

“breaking randomization”, implying biased estimation for causal effects of treatment. However, no one has established a formal causal definition of “breaking randomization”, nor a critical examination of the amount of bias that would result. In this talk, I characterize the conditions under which the arm-based methods be biased for population causal effects and discuss the advantages that arm-based methods have over contrast-based methods with regards to precision.

la randomisation », sous-entendant une estimation biaisée des effets de causalité d’un traitement. Personne n’a cependant donné une définition causale formelle de cette même « détérioration de la randomisation », ni effectué d’examen critique de la quantité de biais qui en découlerait. Cette allocution porte sur la caractérisation des conditions sous lesquelles le biais des méthodes à volets s’appliquerait aux effets de causalité sur une population donnée et également sur les avantages à l’égard de la précision de celles-ci, comparativement à celles fondées sur les contrastes.

[Wednesday June 14/Mercredi 14 juin, 11:04-11:26]

Qinshu Lian (University of Minnesota-Twin Cities), **Haitao Chu** (University of Minnesota-Twin cities)

A Bayesian HSROC Model for Network Meta-analysis of Diagnostic Tests

Modèle bayésien de résumé hiérarchique ROC pour la méta-analyse en réseau de tests diagnostiques

When evaluating the accuracy of diagnostic tests, three designs are commonly used: (1) the crossover design; (2) the randomized design; and (3) the non-comparative design. Existing methods on meta-analysis of diagnostic tests mainly consider the simple cases when the reference test in all or none of the studies can be considered as a gold standard test, and when all studies use either a randomized or non-comparative design. Yet the proliferation of diagnostic instruments and diversity of study designs being used have boosted the demand to develop more general methods. We extend the Bayesian hierarchical summary receiver operating characteristic model to network meta-analysis of diagnostic tests to simultaneously compare multiple tests under a missing data framework. Our model accounts for the potential correlations between multiple tests within a study and the heterogeneity across studies. It also allows different studies to perform different subsets of diagnostic tests. Our model is evaluated through simulations and illustrated using real data from deep vein thrombosis tests.

L’évaluation de l’exactitude des tests diagnostiques fait appel couramment à trois plans d’étude : (1) plan croisé ; (2) plan d’expérience aléatoire ; et (3) étude non comparative. Les méthodes actuelles de méta-analyse des tests diagnostiques sont appliquées à des cas simples, lorsque le test de référence dans l’ensemble ou aucune des études peut être vu comme le test-étalon d’or et que toutes les études font appel soit au plan d’expérience aléatoire, soit à l’étude non comparative. Pourtant, la prolifération d’instruments diagnostiques et la diversité des plans d’étude utilisés ont stimulé la demande pour l’élaboration de méthodes plus générales. Nous étendons le modèle bayésien du résumé hiérarchique ROC à une méta-analyse en réseau de tests diagnostiques pour une comparaison simultanée de tests multiples dans un cadre de données manquantes. Notre modèle prend compte des corrélations potentielles entre les tests multiples dans le cadre de l’étude et de leur hétérogénéité dans l’ensemble des études. Il permet également que diverses études fassent appel à différents sous-ensembles de tests diagnostiques. Notre modèle est évalué à l’aide de simulations et illustré par des données réelles tirées de tests pour la thrombose veineuse profonde.

[Wednesday June 14/Mercredi 14 juin, 11:26-11:50]

Christopher Schmid (Brown University), **Youdan Wang** (Brown University)

Hierarchical Models for Combining N-of-1 Trials

Modèles hiérarchiques pour combiner des essais individuels (N-de-1)

N-of-1 trials are single-patient multiple-crossover studies for determining the relative effectiveness of treatments for an individual participant. A series of N-of-1 trials assessing the same scientific question may be combined to make inferences about the average efficacy of the treatment as well as to borrow

Les essais individuels (N-de-1) sont des études à croisements multiples avec un seul patient afin de déterminer l’efficacité relative d’un traitement pour un patient donné. Des essais N-de-1 évaluant la même question scientifique peuvent être combinés pour en tirer des inférences sur l’efficacité moyenne d’un traitement et également un renforcement de

strength across the series to make improved inferences about individuals. Series that include more than two treatments may enable a network model that can simultaneously estimate and compare the different treatments. Such models are complex because each trial contributes data in the form of a time series with changing treatments. The data are therefore both highly correlated and potentially contaminated by carryover. We will use data from a series of 100 N-of-1 trials in an ongoing study assessing different treatments for chronic pain to illustrate different models that may be used to represent such data.

données empruntées dans l'ensemble de la série pour tirer de meilleures inférences au sujet des patients. Une série comportant plus de deux traitements permettra un modèle de réseau qui peut simultanément évaluer et comparer les différents traitements. Ce sont des modèles complexes, car chaque essai fournit des données sous forme de série temporelle avec des traitements changeants. Par conséquent, les données sont fortement corrélées et potentiellement contaminées par un report. Nous nous sommes servis des données d'une série de 100 essais N-de-1 aux fins d'une étude en cours visant à évaluer divers traitements de la douleur chronique afin d'illustrer divers modèles pouvant être utilisés pour représenter des données de cet ordre.

Biological Data in Action

Données biologiques en action

Chair/Président: Dongsheng Tu

Room/Salle: E2 165 (EITC)

Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:35]

Farhad Shokoohi (McGill University), **Celia M.T. Greenwood** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University), **David A. Stephens** (McGill University), **Aurélié Labbe** (HEC Montréal)

A Novel Hidden Markov Model Approach for Differentially Methylated Region Identification in DNA Methylation Data

Une nouvelle approche de modèle de Markov caché pour l'identification de régions différemment méthylées pour des données de méthylation de l'ADN

A new hidden Markov model (HMM)-based method is proposed to identify differentially-methylated CpG sites and regions in data arising from genome sequencing in conjunction with bisulphite treatment (BS-seq). Features of BS-Seq data include variable read depth, unevenly distributed sites, and non-smooth patterns of correlations between sites. Shortcomings of available methods include the inability to compare more than two groups or to examine associations with continuous covariates, and difficulties in handling missing and low read depth data. The proposed method is flexible in choosing HMM order, uses weights to account for read depth, and introduces a “noise state” to allow some sites to not follow the general HMM pattern. A comprehensive comparison of this new method with several existing methods will be presented, based on real and simulated data.

Une nouvelle méthode basée sur un modèle de Markov caché (HMM) est proposée pour identifier les sites et régions CpG différemment méthylés dans le cadre de données de méthylation obtenues par la technique du séquençage en conjonction avec un traitement au bisulfite (BS-seq). Les caractéristiques des données BS-Seq sont i) la profondeur du séquençage, ii) des sites CpG répartis de façon non uniforme sur le génome, et iii) une corrélation entre les sites. Les inconvénients des méthodes disponibles dans la littérature sont entre autres l'impossibilité de comparer la méthylation de plus de deux groupes, l'impossibilité d'examiner des associations avec des covariables continues, et des difficultés pour traiter les données manquantes et de faible profondeur de séquençage. La méthode proposée est souple dans le choix de l'ordre HMM, utilise des poids pour tenir compte de la profondeur du séquençage et introduit un «état de bruit» pour permettre à certains sites de ne pas suivre le modèle HMM général. Une comparaison complète de cette nouvelle méthode avec plusieurs méthodes existantes sera présentée, basée sur des données réelles et simulées.

[Wednesday June 14/Mercredi 14 juin, 10:35-10:50]

Julien St-Pierre (Université du Québec à Montréal), **Karim Oualkacha** (Université du Québec à Montréal)

A Multivariate Rare-Variant Association Test for Non-Normally Distributed Phenotypes

Test d'association multivarié entre variants rares et phénotypes de distributions non normales

Recently, several region-based multivariate tests have been proposed to identify rare variant (RV) associations when multiple correlated continuous phenotypes are observed. However, most of such methods assume multivariate normality (MN) distribution for the phenotypes. If the MN assumption is violated, modelling phenotypic dependence using the Pearson correlation may not be suitable and can lead to inflated type I error. We address this issue using copulas to model the phenotypic dependence in a bivariate case. This allows us

Récemment, de nombreux tests multivariés d'association pour variants rares (VRs) ont été proposés quand plusieurs phénotypes continus corrélés sont observés. Toutefois, la plupart de ces méthodes présument des distributions normales multivariées (NM) pour les phénotypes. Lorsque l'hypothèse de NM est violée, la modélisation de la dépendance phénotypique par la corrélation de Pearson peut être inadéquate et causer une inflation de l'erreur de type I. Dans ce projet, nous modélisons la dépendance phénotypique à l'aide de copules et nous présumons des distributions phénotypiques marginales provenant de familles exponentielles. Nous développons un

Biological Data in Action

Données biologiques en action

to assume marginal phenotypic distribution from any exponential family. We derive a copula-based variance-component score test for association between RVs and bivariate phenotypes and we provide an analytic test p-value calculation. We illustrate the performance of the proposed methodology using simulation studies and analyzing real data of the PETALE project from Sainte-Justine Hospital research center on patients who survived childhood leukemia.

test du score pour une composante de variance afin de tester pour l'association entre VRs et phénotypes bivariés. Nous proposons aussi une méthode analytique pour calculer la valeur p de notre test. Nous illustrons la performance de la méthode proposée à l'aide de simulations et en analysant des données réelles du projet PETALE du CHU Sainte-Justine.

[Wednesday June 14/Mercredi 14 juin, 10:50-11:05]

Angelo J. Canty (McMaster University), **Keelin Greenlaw** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University), **Celia M.T. Greenwood** (Lady Davis Institute, Jewish General Hospital, Montreal and McGill University)

Measurement and Association Testing of 5-hydroxymethylcytosine (5hmC) Methylation

Mesure et test d'association de la méthylation 5-hydroxyméthylcytosine (5hmC)

Testing association between DNA methylation and health phenotypes is an important area of research. Illumina arrays exploit resistance of methylated DNA to the effects of bisulfite treatment to allow genome-wide inference of the methylation-phenotype relationship. This method does not distinguish between two methylations types: 5mC and 5hmC. The latter form has been shown to be abundant in the human brain and so may be important for neurological phenotypes. A new method called oxidative bisulfite treatment allows for isolation of 5mC methylation and so, in theory, we can find the 5hmC methylation signal by taking the difference of signals from two parallel runs; one using the regular bisulfite treatment and one using the oxidative bisulfite treated DNA. Both of these measures are very noisy and so the difference is often negative. We examine statistical methods to evaluate the amount of 5hmC methylation signal from two parallel runs and the association of 5hmC with a phenotype.

Les tests d'association entre la méthylation de l'ADN et les phénotypes de santé sont un domaine de recherche important. Les bio-puces d'Illumina exploitent la résistance de l'ADN méthylé aux effets du traitement au bisulfite pour permettre une inférence à l'échelle du génome de la relation méthylation-phénotype. Mais cette méthode ne fait pas la distinction entre les deux types de méthylation : 5mC et 5hmC. Cette dernière forme s'avère abondante dans le cerveau humain, si bien qu'elle peut avoir de l'importance pour les phénotypes neurologiques. Une nouvelle méthode, appelée traitement au bisulfite oxydatif, permet d'isoler la méthylation 5mC si bien qu'en théorie, on peut identifier le signal de méthylation 5hmC en calculant la différence entre les signaux de deux expériences parallèles : l'une qui utilise un traitement au bisulfite ordinaire, l'autre utilisant l'ADN traité au bisulfite oxydatif. Mais ces deux mesures sont très bruyantes et la différence est souvent négative. Nous étudions des méthodes statistiques qui permettent d'évaluer le niveau de signal de méthylation 5hmC sur deux expériences parallèles et l'association du 5hmC avec un phénotype.

[Wednesday June 14/Mercredi 14 juin, 11:05-11:20]

Camila P.E. de Souza (University of British Columbia), **Mirela Andronesu** (University of British Columbia), **Sohrab Shah** (University of British Columbia)

A Probabilistic-Based Approach for Clustering Sparse Single-Cell DNA Methylation Data

Approche probabiliste pour le regroupement de données de méthylation d'ADN d'une seule cellule

Recent advances in technology and laboratory protocols have allowed the generation of high-throughput sequencing data from a single cell, including DNA methylation data. One of the main

Les progrès récents en matière de technologie et de protocole de laboratoires ont permis de générer des données de séquençage à haut débit pour une seule cellule, y compris des données de méthylation d'ADN. L'un des principaux objec-

Biological Data in Action

Données biologiques en action

goals in this field of research is to cluster the data from different cells together according to their DNA methylation profiles. Because the amount of DNA material per cell is limited, the generated single-cell data are usually sparse, i.e., they contain a large amount of missing data. To address this problem we propose a probabilistic-based algorithm to infer the missing data and cluster the cells simultaneously by borrowing statistical strength across cells and neighbouring sites. We investigate the properties of our proposed method under different scenarios via simulation studies. We apply our algorithm to publicly available single-cell DNA methylation data sets.

tifs de ce domaine de recherche est de regrouper les données de différentes cellules en fonction de leur profil de méthylation d'ADN. Étant donné que la quantité d'ADN par cellule est limitée, les données générées pour une seule cellule sont généralement éparpillées, c'est-à-dire qu'elles contiennent beaucoup de valeurs manquantes. Pour répondre à ce problème, nous proposons un algorithme probabiliste pour inférer les données manquantes et regrouper les cellules de façon simultanée, en exploitant l'information statistique entre cellules et entre sites voisins. Nous étudions les propriétés de notre méthode sous divers scénarios par des études de simulation. Nous appliquons notre algorithme à des ensembles de données accessibles au public sur la méthylation de l'ADN d'une cellule.

[Wednesday June 14/Mercredi 14 juin, 11:20-11:35]

Elaheh Torkashvand (University of Waterloo), **Joel Dubin** (University of Waterloo), **Gregory Rice** (University of Waterloo)

Modeling T-cell Trajectory Data using Hidden Markov Models

Modélisation de données de trajectoires de cellules T en utilisant des modèles de Markov cachés

Recent technological advances have allowed for the real-time tracking of immune cells, such as cytotoxic T-cells, as they search their environment for pathogens. Understanding the way these cells move is of interest to cell biologists who aim to improve their search efficiency via treatment. In the literature, several different types of stochastic models have been considered for T-cell movements, including Levy flight, Brownian motion, and Persistent random walks. The data that we consider is based on observing the motion of cytotoxic T-cells in a synthetic collagen matrix. Based on our observations, cells often switch between several modes of motion, and these are apparently affected by the attendance of cells nearby in the collagen. We therefore propose a hidden Markov model with separate, but correlated random effects between different Cartesian axes for the increments of the cells' trajectory. We assess the performance of the proposed model using simulation studies.

Les progrès technologiques récents ont permis le suivi en temps réel des cellules immunitaires, comme les cellules T cytotoxiques, qui recherchent des agents pathogènes dans leur environnement. Les biologistes cellulaires cherchent à comprendre la façon dont ces cellules se déplacent dans le but d'améliorer l'efficacité de leurs recherches par le biais du traitement. Dans la littérature, plusieurs types de modèles stochastiques ont été utilisés pour les mouvements des cellules T, y compris le vol Levy, le mouvement brownien et les marches aléatoires persistantes. Nous considérons des données basées sur l'observation du mouvement des cellules T cytotoxiques dans une matrice synthétique de collagène. Selon nos observations, les cellules alternent souvent entre plusieurs modes de mouvement et ces modes sont apparemment affectés par la présence de cellules voisines dans le collagène. Nous proposons donc un modèle de Markov caché avec des effets aléatoires séparés mais corrélés entre différents axes cartésiens pour les augmentations de la trajectoire des cellules. Nous évaluons la performance du modèle proposé à l'aide d'études de simulation.

[Wednesday June 14/Mercredi 14 juin, 11:35-11:50]

Karim Oualkacha (Université du Québec à Montréal), **Mohamed Ouhourane** (Université du Québec à Montréal), **Yi Yang** (McGill University)

Block-Wise Descent Algorithms for Group Variable-selection in Quantile Regression

Algorithmes de descente par bloc pour la sélection d'un groupe de variables dans la régression quantile

Quantile regression (QR) models are attractive in several fields due to their capability to provide

Les modèles de régression quantile sont intéressants dans nombreux domaines en raison de leur capacité de fournir une

a rich description of a set of predictor effects on an outcome without making assumptions on the shape of the outcome distribution. In this work we consider the problem of selecting grouped variables in linear quantile regression models. We introduce a Group variable-Selection framework for Quantile Regression (GSQR) with most appealing group-penalties: the group lasso penalty, the group non-convex penalties (MCP and SCAD), the local approximation of the non-convex penalties and the sparse group lasso penalty. We propose a smooth block-wise descent algorithm that approximates the QR check function by a smooth pseudo-quantile check function, which is differentiable in zero. Then, using a maximization-minimization trick to update each group predictors, we derive a simple and efficient group-wise descent algorithm. We illustrate the GSQR approach performance using simulations in high dimensional settings and we analyze a breast cancer gene expression dataset.

description détaillée d'un ensemble d'effets des prédicteurs sur une réponse sans faire des suppositions sur la forme de la distribution des réponses. Dans cette présentation, nous nous intéressons au problème de sélection de variables groupées dans des modèles de régression quantile linéaire. Nous présentons un cadre de sélection de variables groupées pour la régression quantile au moyen des pénalités de groupe les plus intéressantes : la pénalité Group-Lasso, les pénalités non-convexes (pénalité concave minimax, MCP ; et pénalité lisse coupée de la déviation absolue, SCAD), l'approximation locale des pénalités non-convexes et la pénalité Group-Lasso d'un groupe rare. Nous proposons un algorithme de descente par bloc qui lisse la fonction de vérification de la régression quantile au moyen d'une fonction de vérification pseudo-quantile lisse, qui est dérivable en zéro. Ensuite, au moyen d'une méthode de maximisation-minimisation permettant de mettre à jour les prédicteurs de chaque groupe, nous obtenons un algorithme de descente par groupe simple et efficace. Nous illustrons l'approche de régression quantile avec pénalités de groupe au moyen de simulations dans des cadres à haute dimension et nous analysons un jeu de données d'expression génétique du cancer du sein.

Computational Methods Méthodes computationnelles

Chair/Président: Lawrence C. McCandless
Room/Salle: E2 160 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:35]

François Alastair Marshall

Mode Detection in a Non-Stationary Environment

Détection de modes dans un environnement non stationnaire

Multitaper spectrum estimation techniques can be used to test for the presence of modes in a space physics dataset. The model assumes the presence of random-amplitude periodic components. Random modulators only account for a single form of non-stationary process, and other forms will be sought which reside in the signal. Accounting for these processes reduces the effects of test bias due to departures from the model assumption of stationarity. To detect and identify non-stationary processes, it is standard to use a map of two-frequency coherence, where the process is manifest as a discrete set of high-coherence curves. However, when the dynamic range of the two-frequency spectrum is high, the coherence statistic is the subject of severe bias contamination. The proposed alternative is a test statistic which makes use of circularity coefficients, a set of coefficients which contains the effects of the data tapers on coherence over frequency when bias is high.

On peut utiliser les techniques d'estimation spectrale de type «multitaper» pour détecter la présence de modes dans un ensemble de données en physique cosmique. Le modèle suppose la présence de composantes périodiques d'amplitude aléatoire. Comme les modulateurs aléatoires ne comptent que pour une seule forme de processus non stationnaire, nous recherchons d'autres formes résidant dans le signal. La prise en compte de ces processus réduit les effets de la partialité du test en raison des écarts par rapport à l'hypothèse de stationnarité du modèle. Pour détecter et identifier des processus non stationnaires, on utilise normalement une carte de cohérence à deux fréquences, sur laquelle le processus se manifeste comme un ensemble discret de courbes à forte cohérence. Cependant, en présence d'une longue portée dynamique du spectre à deux fréquences, la statistique de cohérence est assujettie à une sérieuse contamination partielle. La solution de rechange proposée est une statistique de test se servant de coefficients de circularité, un ensemble de coefficients montrant les effets des filtres (tapers) de données sur la cohérence plutôt que sur la fréquence, quand la partialité est élevée.

[Wednesday June 14/Mercredi 14 juin, 10:35-10:50]

Samira Soleymani (Western University), **A. Ian McLeod** (Western University)

Box-Cox Time Series

Séries temporelles de Box-Cox

A new algorithm is developed for Box-Cox estimation, forecasting and simulation of time series models. The algorithm is implemented in R. Some illustrative applications are presented.

Nous mettons au point un nouvel algorithme pour l'estimation, la prévision et la simulation de modèles de séries temporelles de Box-Cox. L'algorithme est implémenté dans R. Nous présentons des exemples d'applications.

[Wednesday June 14/Mercredi 14 juin, 10:50-11:05]

Yilei Wu (University of Waterloo), **Yingli Qin** (University of Waterloo), **Mu Zhu** (University of Waterloo)

An Extended Compound Symmetry Model for High-dimensional Covariance Matrix Estimation

Un modèle élargi de symétrie composée pour l'estimation d'une matrice de covariance à haute dimension

We study high-dimensional covariance matrix estimation under the assumption of the low-rank and diagonal matrix decomposition. The covariance matrix estimator is decomposed into a low-rank ma-

Nous étudions l'estimation d'une matrice de covariance à haute dimension sous l'hypothèse de la décomposition en une matrice diagonale et en une matrice de faible rang. L'estimateur de la matrice de covariance est décomposé en une

Computational Methods Méthodes computationnelles

trix L and a diagonal matrix D , and the rank of L is either fixed to be small or suppressed by a penalty function. Under moderate conditions on the population covariance matrix and the penalty function, our estimator enjoys some consistency results. An algorithm which iteratively updates L and D is applied to solve for the estimator. Some simulations and real data analysis are presented to show the performance with finite sample size.

matrice L de faible rang et en une matrice diagonale D et le rang de L est soit fixé pour être petit ou supprimé par une fonction de pénalité. Sous des conditions modérées pour la matrice de covariance de la population et la fonction de pénalité, notre estimateur démontre des résultats cohérents. Un algorithme qui met à jour L et D selon un mode itératif est appliqué pour calculer l'estimateur. Quelques simulations et analyses de données réelles sont présentées pour démontrer la performance avec un échantillon de taille finie.

[Wednesday June 14/Mercredi 14 juin, 11:05-11:20]

Waleed Almutiry (Calgary University), **Rob Deardon** (Calgary University)

Incorporating Contact Network Uncertainty in Individual-Level Models of Infectious Disease using Approximate Bayesian Computation

Intégrer l'incertitude d'un réseau de contact dans les modèles de maladies infectieuses au niveau des individus à l'aide d'un calcul bayésien approximatif

Infection disease transmission between individuals in a heterogeneous population is often best modelled through a contact network. However, such contact network data are often unobserved. This form of missing data can be accounted for in a Bayesian data augmented framework using Markov chain Monte Carlo (MCMC). However, fitting models in an MCMC a framework can be highly computationally intensive. We investigate the fitting of network-based infectious disease models with completely unknown contact networks using both a full Bayesian MCMC framework and approximate Bayesian computation population Monte-Carlo (ABC-PMC) methods. This is done in the context of both simulated data, and data from the UK 2001 foot-and-mouth disease epidemic. We show that ABC-PMC is able to obtain reasonable approximations of the underlying infectious disease model, with huge savings in computation time.

Le meilleur moyen de modéliser la transmission des maladies infectieuses entre des individus d'une population hétérogène est souvent par un réseau de contacts. Cependant, ces données de réseau de contact ne sont habituellement pas observées. Cette forme de données manquantes peut être prise en compte dans un cadre bayésien de données augmentées utilisant des chaînes de Markov de Monte-Carlo (MCMC). Cependant, ajuster des modèles dans un cadre MCMC peut nécessiter d'intenses calculs. Nous étudions l'ajustement de modèles de maladies infectieuses basés sur des réseaux de contact en présence de réseaux de contact totalement inconnus à l'aide d'un cadre MCMC bayésien complet et à l'aide de méthodes de calcul bayésiens approximatifs d'une population de Monte-Carlo (ABC-PMC). Cela se fait au moyen de données simulées et des données de l'épidémie de fièvre aphteuse du Royaume-Uni de 2001. Nous montrons que le ABC-PMC est en mesure d'obtenir des approximations raisonnables du modèle de maladie infectieuse sous-jacent, avec des économies énormes au niveau du temps de calcul.

[Wednesday June 14/Mercredi 14 juin, 11:20-11:35]

Luc Villandré (McGill University), **Auréli Labbe** (HEC Montréal), **Bluma Brenner** (McGill AIDS Centre, Lady Davis Institute, Jewish General Hospital), **Michel Roger** (Université de Montréal), **David A. Stephens** (McGill University)

DM-PhyClus: A Bayesian Phylogenetic Clustering Algorithm Incorporating Dirichlet-Multinomial Priors

DM-PhyClus : Une approche bayésienne de groupement phylogénétique basée sur des a priori Dirichlet-multinomiaux

Conventional phylogenetic clustering approaches rely on arbitrary cutpoints applied a posteriori to phylogenetic estimates. Although in practice, Bayesian and bootstrap-based clustering tend to lead to similar estimates, they often produce con-

Les approches phylogénétiques conventionnelles de groupement reposent sur des critères arbitraires appliqués a posteriori aux estimations phylogénétiques. Bien qu'en pratique, les méthodes bayésiennes de groupement et celles employant plutôt le bootstrap résultent souvent en des estimations si-

flicting measures of confidence in clusters. Further, bootstrap support for phylogenetic clusters does not benefit from a straightforward interpretation. The current study proposes a new Bayesian phylogenetic clustering algorithm, which we refer to as DM-PhyClus, that identifies sets of sequences resulting from quick transmission chains, thus yielding easily-interpretable clusters, without using any ad hoc distance or confidence requirement. Simulations reveal that DM-PhyClus can outperform conventional clustering methods in terms of mean cluster recovery. We apply DM-PhyClus to a sample of real HIV-1 sequences, revealing a set of clusters whose inference is in line with the conclusions of a previous thorough analysis.

milaires, elles ont tendance à produire des mesures contradictoires pour la confiance en les groupes. D'autre part, la mesure de confiance dérivée du bootstrap est difficile à interpréter. La présente étude propose un nouvel algorithme bayésien de regroupement, dénommé DM-PhyClus, qui identifie directement les ensembles de séquences résultant d'une cascade d'événements de transmission, produisant ainsi des groupes faciles à interpréter, sans invoquer de critères ad hoc de distance ou de confiance. Les simulations réalisées révèlent que DM-PhyClus peut supplanter les méthodes traditionnelles de groupement en termes de la moyenne du taux de découverte. Nous appliquons DM-PhyClus à un échantillon véritable de séquences génétiques du VIH-1, révélant du coup un ensemble de groupes dont l'inférence confirme en bonne partie les conclusions d'une étude exhaustive précédente.

Chair/Président: René Ferland

Room/Salle: E2 155 (EITC)

Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:35]

Kaiqiong Zhao (McGill University), **Elif Fidan Acar** (University of Manitoba)

Conditional Dependence Models under Covariate Measurement Error

Modèles de dépendance conditionnelle avec erreur de mesure des covariables

In many applications, covariates are subject to measurement error. While there is a vast literature on measurement error problems in regression settings, very little is known about the impact of covariate measurement error on the dependence parameter estimation in multivariate models. We address the latter problem using a conditional copula model, and show that the dependence parameter estimates can be significantly biased if the covariate measurement error is ignored in the analysis. We identify the underlying bias pattern from the direction and magnitude of marginal effect sizes and introduce an exact bias correction method for the special case of the Gaussian copula. For more general conditional copula models, a likelihood-based correction method is proposed, in which the likelihood function is computed via Monte-Carlo integration. The consistency and asymptotic normality of the bias-corrected estimators are established. Numerical studies confirm that the proposed bias-correction methods yield an on target estimation. We demonstrate the proposed correction methods in a subset of the SWAN (Study of Women's Health Across the Nation) data.

Dans plusieurs applications, les covariables sont soumises à des erreurs de mesures. Bien qu'il y ait une littérature abondante sur les problèmes d'erreurs de mesure lors de régressions, on sait très peu sur l'impact de l'erreur de mesure de la covariable sur l'estimation du paramètre de dépendance dans des modèles multivariés. Nous abordons ce dernier problème en utilisant un modèle de copule conditionnel et nous démontrons que les estimations du paramètre de dépendance peuvent être considérablement biaisées si l'erreur de mesure des covariables est ignorée dans l'analyse. Nous identifions la structure du biais sous-jacent par la direction et l'ampleur des tailles d'effets marginales et nous introduisons une méthode exacte de correction de biais pour le cas spécial de la copule gaussienne. Pour des modèles de copules conditionnels plus généraux, une méthode de correction fondée sur la vraisemblance est proposée où la fonction de vraisemblance est calculée par intégration Monte Carlo. La cohérence et la normalité asymptotique des estimateurs corrigés pour le biais sont établies. Des études numériques confirment que les méthodes corrigées pour le biais produisent une estimation exacte. Nous démontrons les méthodes de correction proposées sur un sous-ensemble de données provenant du SWAN (Study of Women's Health Across the Nation).

[Wednesday June 14/Mercredi 14 juin, 10:35-10:50]

Janaki Thakshila Koralage (Memorial University of Newfoundland), **Zhaozhi Fan** (Memorial University of Newfoundland)

Localized Quantile Regression of Realized Volatility

Régression quantile localisée de la volatilité réalisée

It is challenging to accurately assess the volatility of financial assets with high-frequency data. Heteroscedasticity of the realized volatility is one of the factors that causes the difficulty. In this study, we propose a localized quantile regression approach. The proposed approach sequentially identifies homogeneous intervals and then applies a quantile regression model to each homogeneous interval. Hence the quantile regression model has

Il est difficile d'évaluer correctement la volatilité des actifs financiers avec des données à hautes fréquences. L'hétéroscédasticité des volatilités réalisées est un des facteurs qui cause cette difficulté. Dans cette étude, nous proposons une approche de régression quantile localisée. Cette approche identifie séquentiellement les intervalles homogènes et applique ensuite un modèle de régression quantile à chaque intervalle homogène. Le modèle de régression quantile a donc des coefficients dépendants du temps. Le modèle de régression quan-

Finance, Time Series, and Extremes

Finance, séries chronologiques et extrêmes

time-dependent coefficients. The quantile regression model does not require distributional assumptions. Direct interpretation of the results at selected quantiles might be of more interests to practitioners in the area of finance. The simulation study shows that localized quantile regression model fits the realized volatility more closely and is also more predictive.

tile ne nécessite pas d'hypothèses sur la distribution. Une interprétation directe des résultats de certains quantiles sélectionnés peut être plus intéressante pour les praticiens dans le domaine de la finance. La simulation démontre que le modèle localisé de régression quantile s'ajuste plus étroitement à la volatilité réalisée et est plus prédictive.

[Wednesday June 14/Mercredi 14 juin, 10:50-11:05]

Gary Sneddon (Mount Saint Vincent University), **Tariqul Hasan** (University of New Brunswick), **Renjun Ma** (University of New Brunswick)

Modelling Bivariate Count Time Series with Extra Zeros

Modélisation de séries chronologiques de dénombrement bivariées avec surreprésentation de zéros

Bivariate time series counts with excessive zeros frequently occur in biological, social and environmental sciences. To deal with such data we propose a model that includes a serial correlated random effect series shared by both responses and the use of the compound Poisson distribution to capture excessive zeros. We will discuss parameter estimation for this model and implement the model in the analysis of musculoskeletal workplace injuries.

Les dénombrements de séries chronologiques bivariées avec surreprésentation de zéros sont utilisés fréquemment en sciences biologiques, sociales et environnementales. Pour traiter ces données, nous proposons un modèle qui comprend une série à effets aléatoires autocorrélée partagée par les deux réponses et l'utilisation d'une distribution Poisson composée afin de tenir compte de la surreprésentation de zéros. Nous discutons de l'estimation des paramètres pour ce modèle et le mettons en application dans l'analyse des blessures musculosquelettiques au travail.

[Wednesday June 14/Mercredi 14 juin, 11:05-11:20]

Nathalie Akakpo (Université Pierre et Marie Curie, Paris)

A Convex Regression Approach to Pickands Dependence Function Estimation

Estimation de la fonction de dépendance de Pickands par une approche de type régression convexe

Due to its close connection to extreme value copulas, the Pickands dependence function is a most useful tool to study dependence between extremes. This function must at the same time be convex and satisfy boundary constraints, so proposing a bona fide estimator remains a difficult task. In this work, we interpret the Pickands dependence function as some kind of regression function, and adapt a least-squares type convex regression algorithm to this problem. Despite the nonparametric nature of the problem, the algorithm does not require to select some smoothing parameter. Moreover, it can be implemented by using convex quadratic programming. Besides, consistency results for the function as well as its derivative may be derived from results for weighted empirical processes.

En raison de son lien étroit avec les copules des valeurs extrêmes, la fonction de dépendance de Pickands est un outil fondamental pour étudier la dépendance entre les extrêmes. Cette fonction vérifie simultanément une contrainte de convexité et des conditions aux bords. Aussi, proposer un estimateur qui satisfait également toutes ces contraintes de forme reste une question difficile. En interprétant la fonction de dépendance de Pickands comme une fonction de régression, nous proposons d'adapter un algorithme de régression convexe pour l'estimation par moindres carrés de cette fonction. Bien qu'il s'agisse d'un problème d'estimation non paramétrique, cet algorithme ne nécessite pas de sélectionner un paramètre de lissage. De plus, son implémentation est réalisable grâce à des outils d'optimisation quadratique convexe. Par ailleurs, il permet d'envisager des résultats de consistance pour la fonction et sa dérivée, basés sur l'étude de processus empiriques pondérés.

[Wednesday June 14/Mercredi 14 juin, 11:20-11:35]

Debbie J. Dupuis (HEC Montreal), **Marco Bee** (University of Trento), **Luca Trapin** (Scuola Normale Superi-

Finance, Time Series, and Extremes

Finance, séries chronologiques et extrêmes

ore)

Realizing the Extremes: Estimation of Tail-Risk Measures from a High-Frequency Perspective

Réaliser les extrêmes : estimation des mesures de risque de queue à partir d'une perspective à haute fréquence

In this talk, we apply realized volatility forecasting to Extreme Value Theory (EVT). We propose a two-step approach where returns are first pre-whitened with a high-frequency based volatility model, and then an EVT based model is fitted to the tails of the standardized residuals. This realized EVT approach is compared to the conditional EVT of McNeil & Frey (2000). We assess both approaches' ability to filter the dependence in the extremes and to produce stable out-of-sample VaR and ES estimates for one-day and ten-day time horizons. The main finding is that GARCH-type models perform well in filtering the dependence, while the realized EVT approach seems preferable in forecasting, especially at longer time horizons.

Dans cet exposé, nous appliquons la prévision de la volatilité réalisée à la théorie des valeurs extrêmes (EVT). Nous proposons une approche à deux étapes où les rendements sont d'abord pré-blanchis avec une volatilité basée sur les données hautes fréquences, et un modèle basé sur EVT est ensuite ajusté à la queue des résidus standardisés. Cette approche EVT réalisée est comparée à l'EVT conditionnelle de McNeil & Frey (2000). Nous évaluons l'habileté des deux approches à pouvoir filtrer la dépendance dans les extrêmes et de produire des estimations stables hors de l'échantillon de la VaR et de l'ES pour des horizons temporels d'une journée et de dix jours. La principale conclusion est que les modèles de type GARCH fonctionnent bien dans le filtrage de la dépendance, tandis que l'approche EVT réalisée semble préférable dans la prévision, surtout à des horizons plus longs.

[Wednesday June 14/Mercredi 14 juin, 11:35-11:50]

Jooyoung Lee (University of Waterloo), **Richard J. Cook** (University of Waterloo)

Copula Dependence Modeling for Multi-type Recurrent Events using Composite Likelihood

Modélisation de la dépendance des copules pour les différents types d'événements récurrents au moyen de la vraisemblance composite

It is appealing to use marginal models for the analysis of recurrent events in randomized trials. When several types of events arise, interest may lie in the nature of the dependence structure. We adopt multivariate random effect models in which the dependence between the type-specific random effects is accommodated through a Gaussian copula. Such models retain the simple interpretation of marginal treatment effects, separately reflect the heterogeneity in risk for each type of event, and provide insight into the dependence between the different types of events. Inference is proposed based on composite likelihood to avoid high dimensional integration. The relative efficiency of estimators obtained from simultaneous and two-stage estimation is examined. An application to a study of nutritional supplements in malnourished children is given in which the goal is to evaluate the reduction in the rate of several types of infection. Extensions to accommodate interval-censoring are described.

Il est intéressant d'utiliser des modèles marginaux pour l'analyse des événements récurrents dans des essais randomisés. Lorsque plusieurs types d'événements se produisent, il peut être intéressant de se pencher sur la nature de la structure de dépendance. Nous adoptons des modèles multivariés à effets aléatoires dans lesquels la dépendance entre les effets spécifiques aléatoires est créée par la copule gaussienne. Ces modèles retiennent l'interprétation simple des effets de traitement marginaux, font séparément ressortir l'hétérogénéité du risque pour chaque type d'événement, et ils donnent un aperçu de la dépendance entre les différents types d'événements. On propose d'effectuer une inférence en fonction de la vraisemblance composite afin d'éviter une forte intégration dimensionnelle. Nous examinons l'efficacité relative des estimateurs obtenus à partir d'une simulation simultanée et à deux étapes. Nous présentons une application permettant d'étudier les suppléments nutritionnels chez les enfants souffrant de malnutrition. Son objectif est d'évaluer la réduction du taux de plusieurs types d'infection. Nous décrivons des extensions pour créer une censure par intervalles.

Recent Advances in Longitudinal and Incomplete Data Analysis

Progrès récents en analyse de données longitudinales et incomplètes

Chair/Président: Haocheng Li and Hua Shen

Organizer/Responsable: Ying Yan

Room/Salle: E2 110 (EITC)

Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 10:20-10:42]

Michael A. McIsaac (Queen's University), **Lauren Paul** (Queen's University)

Multiple Imputation of Incomplete Accelerometer Data

Imputation multiple des données incomplètes d'un accéléromètre

Accelerometer data possess unique challenges for Multiple Imputation (MI) that require careful consideration when selecting imputation models and methods, and these challenges become more daunting when imputation is performed at the epoch-level. Yet MI is appealing in this setting where accelerometers are not consistently worn by study participants, and missing epochs of data create issues for commonly employed methods of accelerometer data analysis and may introduce bias into physical activity measures. Zero-inflated Poisson and log-normal imputation models are contrasted with simpler forms of imputation and these models are evaluated based on their epoch-level imputation accuracy as well as their ability to recover common physical activity summary measures of interest. The accelerometer data used in this investigation is from The Active Play Study, our ongoing physical activity study involving children and youth in Kingston, Ontario.

Les données d'un accéléromètre posent des défis uniques pour l'imputation multiple (MI) qui nécessite un examen approfondi lors de la sélection des modèles et des méthodes d'imputation. Ces défis deviennent encore plus intimidants lorsque l'imputation se fait au niveau des intervalles de temps où s'effectuent les mesures. Pourtant, MI est attrayante dans ce contexte où les accéléromètres ne sont pas portés en tout temps par les participants de l'étude, et ces données manquantes par intervalles créent des problèmes pour l'utilisation des méthodes d'analyse courantes de données d'accéléromètres et peuvent introduire un biais dans les mesures d'activité physique. Le modèle d'imputation de Poisson à valeurs nulles fréquentes et le modèle d'imputation log-normal sont comparés à des formes d'imputation plus simples et ces modèles sont évalués selon leur précision d'imputation au niveau des intervalles ainsi que selon leur capacité de récupérer des mesures sommaires et courantes de l'activité physique. Les données d'accéléromètre utilisées dans cette étude proviennent de The Active Play Study, notre étude en cours sur l'activité physique impliquant des enfants et des jeunes de Kingston en Ontario.

[Wednesday June 14/Mercredi 14 juin, 10:42-11:04]

Hua Shen (University of Calgary), **Ying Yan** (University of Calgary), **Haocheng Li** (University of Calgary)

Analysis of Heterogeneous Lifetime Data with Measurement Error in Covariate

Analyse de données hétérogènes sur la durée de vie avec erreurs de mesure dans les covariables

In the analysis of lifetime data, interest often lie in modeling the time to a particular event or the occurrence of certain event, identifying the associated risk factors and evaluating their effects on survival time or event occurrence. We may encounter a heterogeneous group where event of interest may not occur or cease to occur after a period of time for some individuals. Mover-stayer models are developed using a binary variable to indicate whether the underlying process has resolved. In addition, data with measurement error often occurs in medical research and public health, and the problems

Lors de l'analyse de données sur la durée de vie, l'objectif est souvent de modéliser le temps jusqu'à un événement particulier ou l'occurrence d'un événement, d'identifier les facteurs de risque associés et d'évaluer leur effets sur la durée de survie ou sur l'occurrence d'un événement. Nous pouvons rencontrer un groupe hétérogène où les événements d'intérêt ne se produisent pas ou cessent de se produire après une certaine période de temps chez certains individus. Les modèles mobiles-stables sont développés en utilisant une variable binaire pour indiquer si le processus sous-jacent a été résolu. De plus, il y a souvent des données avec erreurs de mesure dans la recherche médicale et dans la santé publique et les

Recent Advances in Longitudinal and Incomplete Data Analysis

Progrès récents en analyse de données longitudinales et incomplètes

associated are well known yet often ignored due to technical difficulties. We propose an expectation-maximization (EM) algorithm based method to estimate the parameters and adjust for mismeasurement present in the covariates. Its advantages over the naïve analysis are illustrated in simulation studies. The motivating study in breast cancer is examined to illustrate the proposed method.

problèmes qui y sont associés sont bien connus mais souvent ignorés à cause de problèmes techniques. Nous proposons une méthode fondée sur un algorithme d'espérance-maximisation (EM) pour estimer les paramètres et pour ajuster pour les erreurs de mesure présentes dans les covariables. Ses avantages sur l'analyse naïve sont illustrés par des simulations. Une étude sur le cancer du sein est utilisée pour illustrer la méthode proposée.

[Wednesday June 14/Mercredi 14 juin, 11:04-11:26]

Haocheng Li (University of Calgary), **Grace Yi** (University of Waterloo)

Missing Data Mechanisms for Incomplete Longitudinal Observations in both Responses and Covariates

Mécanismes de données manquantes pour des observations longitudinales incomplètes dans les réponses et les covariables

Missing observations in both responses and covariates are commonly observed in longitudinal studies. When missing data are missing not at random, inferences under the likelihood framework often require joint modelling of response and covariate processes, as well as missing data processes associated with incompleteness of responses and covariates. Specification of these four joint distributions is difficult in both modelling and computation. We discuss three specific missing data mechanisms which have simplified pairwise likelihood formulations. These likelihood functions lead to consistent estimators, and enjoy better robustness and computational convenience. The proposed method is evaluated empirically by simulation studies and applied in a real world dataset.

Des données manquantes dans les réponses et les covariables sont couramment observées dans les études longitudinales. Lorsque le processus qui sous-tend les données manquantes dépend d'informations non observables, les inférences dans le cadre de la vraisemblance nécessitent souvent une modélisation conjointe des processus de réponses et de covariables ainsi que des processus de données manquantes associés aux réponses et des covariables incomplètes. La spécification de ces quatre lois conjointes est difficile à la fois en termes de modélisation et de calcul. Nous traitons de trois mécanismes spécifiques de données manquantes qui ont simplifié la formulation de vraisemblance par paires. Ces fonctions de vraisemblance engendrent des estimateurs convergents, une plus grande robustesse et des calculs plus faciles à effectuer. La méthode proposée est évaluée empiriquement avec des simulations et est appliquée à des données réelles.

[Wednesday June 14/Mercredi 14 juin, 11:26-11:50]

Mireille E. Schnitzer (Université de Montréal), **Steve Ferreira Guerra** (Université de Montréal), **Amélie Forget** (Université de Montréal), **Lucie Blais** (Université de Montréal)

A Self-Selecting Procedure for the Optimal Discretization of the Timeline for Longitudinal Causal Inference Methods with Electronic Health Data

Une procédure de sélection automatique pour la discrétisation optimale de la ligne de temps pour l'application de méthodes longitudinales en inférence causale avec données électroniques sur la santé

In longitudinal observational studies, marginal structural models (MSM) can be used to account for time-dependent confounding. Many estimation approaches, such as Longitudinal Targeted Maximum Likelihood Estimation (LTMLE), require a finite number of time points where variables are measured. In contrast, electronic health data (EHD) are produced by mechanisms that collect health system user information in real-time. In common practice, the longitudinal analysis is preceded by an

Dans le cadre des études d'observations longitudinales, les modèles structuraux marginaux (MSM) peuvent être utilisés pour représenter la confusion dépendante du temps. Plusieurs approches d'estimation, telle que l'estimation par maximum de vraisemblance ciblée longitudinale (LTMLE), nécessitent un nombre fini de points dans le temps où les variables sont mesurées. Par contre, les données électroniques sur la santé (EHD) sont produites par des mécanismes qui recueillent l'information sur les usagers du système de santé en temps réel. La pratique courante est de précéder l'analyse longitudinale

Recent Advances in Longitudinal and Incomplete Data Analysis

Progrès récents en analyse de données longitudinales et incomplètes

arbitrary discretization. In this talk, we describe the causal inference problem when the operating data are defined as a coarsening (discretization) of the observed data. We propose a novel selection procedure that uses cross-validation on an LTMLE loss function to select an "optimal" discretization for estimation of a data-adaptive MSM parameter of interest. We present a simulation study and apply this approach to a study using EHD to evaluate the relative impact of asthma medications during pregnancy.

par une discrétisation arbitraire. Dans cet exposé, nous décrivons le problème en inférence causal lorsque les données opérationnelles sont définies comme étant une discrétisation des données observées. Nous proposons une procédure de sélection novatrice qui utilise une validation croisée sur une fonction de perte LTMLE pour sélectionner une discrétisation « optimale » pour l'estimation d'un paramètre MSM adaptif. Nous présentons une étude de simulation et nous appliquons cette approche à une étude utilisant des EHD pour évaluer l'impact relatif de la médication pour l'asthme durant la grossesse.

Isobel Loutit Lecture
Allocution invitée Isobel Loutit

Chair/Président: François Bellavance
Organizer/Responsable: François Bellavance
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 13:30-14:30]

Shirley E. Mills (Carleton University)

Predictive Analytics in Business

Analytique prévisionnelle en affaires

This talk will present an overview of the history and current advances of predictive analytics in the world of business, with applications drawn from such diverse fields as retail, utilities, manufacturing, public service, financial services, insurance, and sports. Methods will include discussion of such topics as ensemble models, gradient boosting, partial least squares, support vector machines, text mining, link analysis, and association rule mining.

Cette allocution se veut un aperçu de l'histoire de l'analytique prévisionnelle et de ses progrès récents dans le monde des affaires, avec présentation d'applications tirées de domaines aussi diversifiés que la vente au détail, les services publics, la fabrication, la fonction publique, les services financiers, l'assurance et le sport. Nous aborderons notamment des méthodes comme les suivantes : modèle d'ensemble, boosting par descente du gradient, moindres carrés partiels, machine à vecteurs de support, exploration de texte, analyse des liens et règle d'association.

CJS Award Address
Allocution du récipiendaire du Prix de la RCS

Chair/Président: Richard Lockhart
Organizer/Responsable: Richard Lockhart
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 13:30-14:15]

Bryan E. Shepherd (Vanderbilt University School of Medicine), **Chun Li** (Case Western Reserve University),
Qi Liu (Merck)

Probability-Scale Residuals for Continuous, Discrete, and Censored Data

Résidu lié à une échelle de probabilités pour données continues, discrètes et tronquées

We describe a new residual for general regression models defined as $\text{pr}(Y^* < y) - \text{pr}(Y^* > y)$, where y is the observed outcome and Y^* is a random variable from the fitted distribution. This probability-scale residual (PSR) can be written as $E(\text{sign}(y, Y^*))$, whereas the popular observed-minus-expected residual can be thought of as $E(y - Y^*)$. Therefore the PSR is useful in settings where differences are not meaningful or where the expectation of the fitted distribution cannot be calculated. We present several desirable properties of the PSR that make it useful for diagnostics and measuring residual rank correlation. We demonstrate its utility for continuous, ordered discrete, and censored outcomes, and with various models including Cox regression, quantile regression, and ordinal cumulative probability models, for which fully specified distributions are not desirable or needed, and in some cases suitable residuals are not available. The residual is illustrated with simulated and real data.

Nous décrivons une nouvelle forme de résidus pour un modèle général de régression définis par $\text{pr}(Y^* < y) - \text{pr}(Y^* > y)$, où y est la valeur observée et Y^* est une variable aléatoire suivant la distribution prescrite par le modèle ajusté. Lié à une échelle de probabilités, ce résidu peut s'écrire $E(\text{sign}(y, Y^*))$, alors que la définition populaire correspond plutôt à $E(y - Y^*)$. Le résidu proposé est donc utile si la différence entre la valeur observée et espérée de la définition populaire n'a pas de sens interprétable, ou lorsque la valeur espérée selon le modèle n'est pas calculable. Nous présentons de nombreuses propriétés désirables de ces résidus, rendant cette approche utile pour le diagnostic de modèles et le calcul de corrélations dans les résidus. Nous illustrons son usage pour des données continues, ordonnées discrètes et censurées et pour divers modèles comme la régression de Cox, la régression quantile et les modèles ordinaux de probabilités cumulatives. Les distributions implicites de ces modèles n'ont pas besoin d'être complètement définies et, dans certains cas, les résidus habituels sont simplement indisponibles. Nous illustrons notre nouvelle définition des résidus par des simulations et avec un jeu de données réelles.

Pierre Robillard Address
Allocution du récipiendaire du Prix Pierre-Robillard

Chair/Président: Gordon Fick
Organizer/Responsable: Gordon Fick
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 14:15-15:00]

Andy Leung (University of British Columbia), **Victor J. Yohai** (Universidad de Buenos Aires), **Ruben H. Zamar** (University of British Columbia)

Robust Estimation under Cellwise and Casewise Contamination

Estimation robuste en cas de contamination « cellwise » ou « casewise »

In traditional robust statistics, it is generally assumed that the majority of the observations in the data are free of contamination, while only a minority of the observations are contaminated. The contaminated observations are flagged as outliers and down-weighted even if only a single component is contaminated. Some observations may fully depart from the bulk of the data. This situation usually refers to casewise outliers. However, observations can be only partially contaminated. This type of contamination often appears as single outlying cells in a data matrix and therefore, usually refers to cellwise contamination. Under cellwise contamination, a lot of information could be lost through down-weighting the whole observation, especially for high-dimensional data. Recent work has shown that procedures that proceed in such way are not robust. In this talk, we will sketch out our proposal to estimate multivariate location and scatter under this cell-and-casewise contamination.

En statistique robuste traditionnelle, on présume généralement que la majorité des observations sur les données sont libres de contamination et que seule une minorité est contaminée. Ces dernières sont identifiées comme valeurs aberrantes et sous-pondérées même si un seul composant de l'observation est contaminé. Certaines observations s'écartent entièrement de l'essentiel des données : on parle alors de valeurs aberrantes « casewise » (par cas). Toutefois, d'autres observations peuvent n'être que partiellement contaminées. Ce type de contamination apparaît souvent comme de simples cellules aberrantes dans une matrice de données : on parle alors de contamination « cellwise » (par cellule). Si la contamination ne concerne qu'une cellule, on perdrait beaucoup d'informations en sous-pondérant l'ensemble de l'observation, surtout pour des données de haute dimension. De récents travaux ont montré que les procédures typiquement employées ne sont pas robustes. Dans cette présentation, nous ébauchons une méthode d'estimation des paramètres de position et dispersion multivariés en cas de contamination cellwise et casewise.

Recent Advances in Compositional Data Analysis and Applications

Progrès récents en analyse de données de composition et ses applications

Chair/Président: Connie Stewart
Organizer/Responsable: Connie Stewart
Room/Salle: E3 270 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 15:00-15:30]

Holly N. Steeves (Dalhousie University)

Maximum Likelihood Estimation of Marine Predators' Diet Proportions

Estimation du maximum de vraisemblance des proportions alimentaires des prédateurs marins

Diet compositions of marine predators are often of interest for marine ecologists in trophic structure studies where non-lethal sampling has created a need for non-invasive diet estimation techniques. Methods using fatty acids have been developed to obtain dietary estimates that have previously been difficult to acquire. Building on the existing quantitative fatty acid signature analysis, we have constructed a maximum likelihood approach to estimating dietary proportions that includes standard errors and allows for the potential for inclusion of covariates in the model. The model is assessed using simulated as well as real-life data, and results are compared to those of the current approach.

La composition des régimes alimentaires des prédateurs marins suscite souvent l'intérêt des écologistes marins dans les études de structure trophique dans lesquelles l'échantillonnage non létal a créé un besoin de techniques d'estimation alimentaires non invasives. Des méthodes utilisant des acides gras ont été mises sur pied afin d'obtenir des estimations du régime alimentaire qui auparavant étaient difficiles à obtenir. En nous appuyant sur l'analyse quantitative de la signature des acides gras, nous avons élaboré une approche de vraisemblance qui permet d'estimer les proportions alimentaires comprenant des erreurs-types, ainsi que d'inclure potentiellement des covariables dans le modèle. Nous analysons le modèle au moyen de données simulées et réelles, et nous comparons les résultats à ceux de l'approche actuelle.

[Wednesday June 14/Mercredi 14 juin, 15:30-16:00]

Janice Lea Scealy (Australian National University), **Alan Welsh** (Australian National University)

Estimation and Inference in Directional Mixed Models for Compositional Data

Estimation et inférence dans les modèles mixtes directionnels pour données relatives à la composition

We propose a new class of mixed effects model for compositional data based on the Kent distribution for directional data, where the random effects also have Kent distributions. One useful property of the new directional mixed model is that the marginal mean direction has a closed form and is interpretable. The random effects enter the model in a multiplicative way via the product of a set of rotation matrices and the conditional mean direction is a random rotation of the marginal mean direction. For estimation we apply a quasi-likelihood method which results in solving a new set of generalised estimating equations and these are shown to have low bias in typical situations. For inference we use a nonparametric bootstrap method for clustered data which does not rely on estimates of the shape parameters (shape parameters are difficult to estimate in Kent models). The new approach is shown to be more tractable than the traditional approach

Nous proposons une nouvelle catégorie de modèle à effets mixtes pour les données relatives à la composition reposant sur la loi de Kent pour des données directionnelles, où les effets aléatoires ont également des lois de Kent. Le nouveau modèle mixte directionnel présente une propriété utile : la direction de la moyenne marginale est sous une forme analytique fermée et interprétable. Les effets aléatoires sont intégrés dans le modèle de façon multiplicative au moyen du produit d'un jeu de matrices de rotation, et la direction de la moyenne conditionnelle est une rotation aléatoire de la direction de la moyenne marginale. À des fins d'estimation, nous utilisons une méthode de quasi-vraisemblance qui entraîne la résolution d'un nouveau jeu d'équations d'estimation généralisées, et ces équations sont faiblement biaisées dans des situations typiques. Pour l'inférence, nous utilisons une méthode bootstrap non paramétrique pour des données groupées qui ne reposent pas sur des estimations de paramètres de forme (il est difficile d'estimer les paramètres de forme dans les modèles de Kent). La nouvelle approche est beau-

Recent Advances in Compositional Data Analysis and Applications

Progrès récents en analyse de données de composition et ses applications

based on the logratio transformation.

coup plus souple que l'approche traditionnelle fondée sur la transformation logarithmique du ratio.

Recent Advances in Compositional Data Analysis and Applications

Progrès récents en analyse de données de composition et ses applications

[Wednesday June 14/Mercredi 14 juin, 16:00-16:30]

Michail Tsagris (University of Crete), **Connie Stewart** (University of New Brunswick)

A Dirichlet Regression Model for Compositional Data with Zeros

Un modèle de régression de Dirichlet pour des données compositionnelles avec zéros

Compositional data are met in many different fields, such as economics, archaeometry, ecology, geology and political sciences. Regression where the dependent variable is a composition is usually carried out via a log-ratio transformation of the composition or via the Dirichlet distribution. However, when there are zero values in the data these two ways are not readily applicable. Suggestions for this problem exist, but most of them rely on substituting the zero values. In this paper we adjust the Dirichlet distribution when covariates are present, in order to allow for zero values to be present in the data, without modifying any values. To do so, we modify the log-likelihood of the Dirichlet distribution to account for zero values. Examples and simulation studies exhibit the performance of the zero adjusted Dirichlet regression.

Les données compositionnelles se retrouvent dans plusieurs domaines différents, tels qu'en économie, archéométrie, écologie, géologie et sciences politiques. Lorsque la variable dépendante est une composition, la régression est généralement effectuée au moyen d'une transformation log-ratio de la composition ou par la loi de Dirichlet. Par contre, lorsqu'il y a présence de valeurs zéro dans les données ces deux moyens ne sont pas facilement applicables. Il existe des pistes de solutions pour ces problèmes mais la plupart s'appuie sur la substitution des valeurs zéro. Dans cet article nous ajustons la loi de Dirichlet quand des covariables sont présentes pour permettre la présence de valeurs zéro dans les données sans modifier aucune valeur. Pour ce faire, nous modifions la log-vraisemblance de la loi de Dirichlet pour prendre en compte les valeurs zéro. Des exemples et des simulations démontrent la performance de la régression de Dirichlet ajustée pour les zéros.

University of Manitoba Department of Statistics: 50th Anniversary Showcase
Département de statistique de l'Université du Manitoba : vitrine sur 50 ans
d'histoire

Chair/Président: Liquin Wang
Organizer/Responsable: Liquin Wang
Room/Salle: E2 105 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 15:00-15:30]

Jerald F. Lawless (University of Waterloo)

Statistics, Science and Technology

Statistiques, science et technologie

There have been huge advances in measurement, computing and information technology over the 50 years since the University of Manitoba Department of Statistics was founded, and these have transformed the ways that statistics is applied. In this talk I will review some history of statistical applications in science and technology before and then during this period. I will then discuss some current activities and some important issues, including the use of big data, the role of statistics within data science, and the distinction between scientific learning and technological problems involving decision-making.

Depuis la fondation du Département de Statistiques de l'Université du Manitoba il y a 50 ans, il y a eu des avancées spectaculaires dans les technologies de l'informatique, de mesure, et de l'information, ce qui a transformé la façon dont les statistiques sont utilisées. Dans cet exposé, je vais résumer quelques applications historiques des statistiques en science et technologie avant et durant cette période. Je discuterai ensuite de quelques activités en cours et de quelques enjeux importants, dont l'utilisation de données volumineuses, le rôle des statistiques au sein de la science des données et la distinction entre le savoir scientifique et les problèmes technologiques impliquant la prise de décision.

[Wednesday June 14/Mercredi 14 juin, 15:30-16:00]

John F. Brewster (University of Manitoba)

Reflections on Interactions with Industry

Réflexions sur les interactions avec l'industrie

As one of the speakers in this session celebrating the 50th Anniversary of the Department of Statistics at the University of Manitoba, sponsored by the Business and Industrial Statistics Section, I was asked to reflect on some earlier interactions between the Department of Statistics and industrial partners. I will provide examples of these interactions and describe how these interactions arose. I will also discuss the benefits of academic statisticians being involved in such interactions, as well as some of the challenges and lessons learned.

Étant l'un des conférenciers présents pour la célébration du 50e anniversaire du Département de Statistiques de l'Université du Manitoba, parrainé par le Groupe de statistique industrielle et de gestion, on m'a demandé de réfléchir aux premières interactions entre le Département de Statistiques et les partenaires industriels. Je donnerai des exemples de ces interactions et je décrirai comment ces interactions se sont produites. Je discuterai aussi des avantages qu'il y a à avoir des statisticiens académiques engagés dans de telles interactions ainsi que de quelques défis et leçons apprises.

[Wednesday June 14/Mercredi 14 juin, 16:00-16:30]

Brian Macpherson (retired), **Smiley Cheng** (University of Manitoba)

Industrial Quality Improvement Workshops and Applications

Ateliers d'amélioration de la qualité et ses applications

The Department of Statistics at the University of Manitoba for many years received widespread recognition for its activities in the area of quality improvement. Acting on a request from a com-

Pendant de nombreuses années, le Département de Statistiques de l'Université du Manitoba a été largement reconnu pour ses activités dans le domaine de l'amélioration de la qualité. Suite à une demande d'aide de la part d'une com-

University of Manitoba Department of Statistics: 50th Anniversary Showcase Département de statistique de l'Université du Manitoba : vitrine sur 50 ans d'histoire

pany in Winnipeg for help with meeting a “required quality standard” a workshop on the use of statistical tools for quality improvement was developed and delivered. Over the following almost twenty years more than 60 workshops were given to local businesses, organizations, and government agencies. Through these workshops and other actions, the Department gained an enviable reputation and generated significant funds to support graduate students and other activities in the use of statistical techniques for quality improvement. We will describe the nature of these workshops and the resulting benefits to the Department. Examples will be provided of resulting research applications that directly flowed from these activities.

pagnie de Winnipeg pour atteindre un « niveau de qualité requis », un atelier sur l'utilisation d'outils statistiques pour l'amélioration de la qualité a été développé et offert. Au cours des presque 20 années suivantes, plus de 60 ateliers ont été tenus pour des entreprises et organisations locales et pour des agences gouvernementales. À travers ces ateliers et autres actions, le Département a acquis une réputation enviable et a généré des fonds importants pour le soutien d'étudiants aux cycles supérieurs et pour d'autres activités sur l'utilisation de techniques statistiques pour l'amélioration de la qualité. Nous décrirons la nature de ces ateliers et les bénéfices pour le département. Des exemples d'applications de recherche qui ont directement découlés de ces activités seront présentés.

Waterloo at 50: Reflections of Former Students Waterloo à 50 ans

Chair/Président: Mary E. Thompson
Organizer/Responsable: Mary E. Thompson
Room/Salle: E2 110 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 15:00-16:30]

Mary E. Thompson (University of Waterloo), **Vern Farewell** (University of Cambridge), **Wayne Oldford** (University of Waterloo), **Karen Kopciuk** (University of Calgary), **Robert L. Brown** (University of Waterloo)

Panel discussion

Les 50 ans de l'Université de Waterloo : réflexions d'anciens étudiants (table ronde)

The Department of Statistics (later Statistics and Actuarial Science) at the University of Waterloo was founded 50 years ago, in 1967. Four panellists with academic research careers who spent their undergraduate or graduate studies years at Waterloo will reflect on the history of the department, their own student experiences, and how studying at Waterloo influenced their careers.

Le département des Statistiques (appelé plus tard des Statistiques et de la science actuarielle) de l'Université de Waterloo a été mis sur pied il y a 50 ans, en 1967. Quatre panélistes diplômés de premier cycle ou de cycle supérieur de cette institution et qui ont ensuite mené des carrières de chercheurs universitaires feront part de leurs réflexions sur l'histoire de ce département, leur expérience d'étudiants et l'influence sur leur carrière de leurs études à l'Université de Waterloo.

**Recent Developments in Copula Modelling: A Showcase Session of the CANSSI-CRT
on ‘Copula Dependence Modelling: Theory and Applications’**
**Développements récents en modélisation par les copules : vitrine sur le PRC de
l’INCASS sur “Modélisation de la dépendance par les copules : théorie et
applications”**

Chair/Président: Elif Fidan Acar
Organizer/Responsable: Elif Fidan Acar
Room/Salle: E2 130 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 15:00-15:22]

Marie-Hélène Toupin (Université Laval), **Jean-François Quessy** (Université du Québec à Trois-Rivières),
Louis-Paul Rivest (Université Laval)

The Family of Chi-square Copulas and Applications in Spatial Statistics

La famille de copules khi-carré et son utilisation en statistique spatiale

First, some dependence properties of the family of chi-square copulas are presented. This family is very attractive because it generalizes the Gaussian copula and allows for flexible modeling of high dimensional random vectors. This class of dependence structures is then applied in spatial statistics to predict the value of a non-Gaussian stationary random field at a position where it has not been observed. This interpolation method is then compared to the kriging method in a real example and via a simulation study.

Dans un premier temps, les propriétés de dépendance de la famille de copules khi-carré sont présentées. Cette famille est très attrayante, car elle généralise la populaire copule normale et permet la modélisation flexible de vecteurs aléatoires de grande dimension. Cette classe de structures de dépendance est ensuite appliquée dans un contexte de statistique spatiale afin de prévoir la valeur d’un champ aléatoire stationnaire non Gaussien à une position où il n’a pas été observé. Cette méthode d’interpolation est comparée à la méthode du krigage dans un exemple réel et via une étude de simulation.

[Wednesday June 14/Mercredi 14 juin, 15:22-15:45]

Bo Chang (University of British Columbia), **Harry Joe** (University of British Columbia)

Vine Copula Regression

Régression de copules en vigne

Regression analysis is one of the oldest topics in statistics. If explanatory variables and a response variable of interest are simultaneously observed, then fitting a joint multivariate density to all variables would enable prediction via conditional distributions. The vine copulas have proven to be a flexible tool in high-dimensional dependence modeling. We introduce a new regular vine copula based regression model, which avoids issues of variable selection, variable transformation, and heteroscedasticity. It uses general regular vines and we have developed an efficient algorithm to compute the conditional distribution. Our model is able to handle a mix of continuous and discrete variables. That is, the predictors and response variable can be either continuous or discrete; thus we have a unified model for regression and classification problems. We demonstrate the predictive performance of the proposed method using a real data set.

L’analyse de régression est l’un des plus anciens sujets statistiques. Si l’on observe simultanément des variables explicatives et une variable réponse d’intérêt, l’ajustement d’une densité conjointe multivariée à toutes les variables devrait permettre une prévision au moyen de lois conditionnelles. Les copules en vigne se sont avérées être un outil souple dans la modélisation de la dépendance de grande dimension. Nous présentons un nouveau modèle de régression fondé sur les copules en vigne qui évite les problèmes de sélection et de transformation de variables, ainsi que d’hétéroscédasticité. Ce modèle utilise des vignes courantes, et nous avons créé un algorithme efficace pour calculer la loi conditionnelle. Notre modèle peut traiter un mélange de variables continues et discrètes, c’est-à-dire que les prédicteurs et la variable réponse peuvent être continus ou discrets. Ainsi, nous disposons d’un modèle unifié permettant de résoudre les problèmes de régression et de classification. Nous démontrons les propriétés prévisionnelles de la méthode proposée au moyen d’un jeu de données réelles.

[Wednesday June 14/Mercredi 14 juin, 15:45-16:07]

Recent Developments in Copula Modelling: A Showcase Session of the CANSSI-CRT on ‘Copula Dependence Modelling: Theory and Applications’

Développements récents en modélisation par les copules : vitrine sur le PRC de l’INCASS sur “Modélisation de la dépendance par les copules : théorie et applications”

Samuel Perreault (Université Laval), **Thierry Duchesne** (Université Laval), **Johanna Neslehova** (McGill University)

Improved Estimation and Identification of Block-Exchangeable Dependence Structures

Estimation améliorée et identification de structures de dépendance échangeables par blocs

We use concepts from copula modeling and statistical learning to improve the estimation of the Kendall’s tau correlation matrix of a set of variables. We are interested in recovering structural properties of the variables’ joint distribution. Relaxing the full exchangeability assumption, we assume the existence of exchangeable subsets of variables without explicitly specifying them. These block-exchangeable dependence structures are such that their corresponding tau matrix contains many identical entries. We iteratively modify the usual estimator of tau by introducing constraints corresponding to the structural properties discovered, thus producing a sparser representation of the dependence. A first algorithm is used to produce a sequence of decreasingly complex models among which the final model is to be selected. Final selection is provided by a second algorithm. The procedure yields improved estimation of the tau matrix and better representation of the dependence among the variables. The approach is applied to data from the US stock market.

Nous utilisons des concepts populaires en modélisation avec copules et en apprentissage statistique pour améliorer l’estimation de la matrice des tau de Kendall d’un ensemble de variables aléatoires. On s’intéresse particulièrement à recouvrer les propriétés structurales de la distribution conjointe des variables. En relaxant l’hypothèse d’échangeabilité complète entre les variables, nous supposons l’existence de sous-ensembles (blocs) échangeables sans pour autant les définir explicitement. Ces structures sont telles que leurs matrices des tau contiennent plusieurs éléments identiques. On modifie l’estimateur usuel de la matrice des tau itérativement en y introduisant les contraintes associées aux propriétés structurales découvertes. Un premier algorithme est utilisé pour produire une séquence de modèles de plus en plus clairs parmi lesquels le modèle final sera choisi. La sélection finale est faite à l’aide d’un deuxième algorithme. La procédure dans son ensemble fournit un estimateur amélioré de la matrice des tau et une meilleure représentation de la dépendance entre les variables. Les concepts sont expliqués à travers un exemple simple et une application à des données sur 20 titres provenant des marchés boursiers américains.

[Wednesday June 14/Mercredi 14 juin, 16:07-16:30]

Mohamed Belalia (Université du Québec à Trois-Rivières), **Jean-François Quessy** (University of Québec at Trois-Rivières)

Generalized Simulated Method-of-Moments for Copula Parameters of Arbitrary Dimension

Méthode des moments généralisée simulée pour des paramètres de copules de dimension arbitraire

In this talk, we develop a general method for estimating a vector of parameters of an arbitrary dimension in multivariate copula models. The proposed estimator is based on the first p -moments of the multivariate probability integral transformation that one can associate to a given parametric copula model. An unbiased estimator of these p -moments are first described, from which a method-of-moments estimator of the unknown vector of parameters is defined. In order that the method be applicable even when explicit expressions for the theoretical moments are not available, a simulated version of the estimator is developed as well. Interestingly, the latter can be performed as long as one is able to simulate from a given copula model.

Dans cet exposé, nous développons une méthode générale permettant d’estimer un vecteur de paramètres de dimension arbitraire dans des modèles de copules multivariées. L’estimateur proposé est fondé sur les premiers moments p de la transformation intégrale de probabilité que l’on peut associer à un modèle de copules paramétrique donné. Nous décrivons tout d’abord un estimateur sans biais de ces moments p , à partir duquel nous définissons un estimateur de la méthode des moments du vecteur de paramètres inconnu. Afin que cette méthode soit applicable même lorsqu’il n’y a pas d’expressions explicites des moments théoriques, nous développons une version simulée de cet estimateur. Cette méthode peut être appliquée dès lors qu’il est possible de simuler des données à partir du modèle de copules. La convergence et la normalité asymptotique des estimateurs sont formellement

**Recent Developments in Copula Modelling: A Showcase Session of the CANSSI-CRT
on ‘Copula Dependence Modelling: Theory and Applications’
Développements récents en modélisation par les copules : vitrine sur le PRC de
l’INCASS sur “Modélisation de la dépendance par les copules : théorie et
applications”**

The consistency and asymptotic normality of these estimators are formally established under standard and mild conditions. The performance of these estimators in terms of bias and mean-squared errors is investigated through an extensive simulation study.

établies sous des conditions standard et faibles. Nous étudions la performance de ces estimateurs pour ce qui est du biais et des erreurs quadratiques moyennes au moyen d’une vaste étude de simulation.

**Recent Developments in the Analysis of Neuroimaging Data: A Showcase of the
CANSSI-CRT on ‘Joint Analysis of Neuroimaging Data’**
**Développements récents en analyse des données de neuroimagerie : vitrine sur le
PRC de l’INCASS sur “l’analyse conjointe de données de la neuroimagerie”**

Chair/Président: Farouk Nathoo
Organizer/Responsable: Farouk Nathoo
Room/Salle: E2 125 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 15:00-15:30]

Dehan Kong (University of Toronto), **Wei Hu** (University of California Irvine), **Weining Shen** (University of California Irvine), **Hua Zhou** (University of California Los Angeles)

Matrix Linear Discriminant Analysis

Analyse discriminante linéaire de matrices

We propose a novel linear discriminant analysis approach for the classification of high-dimensional matrix-valued data that commonly arise from imaging studies. Motivated by the equivalence of the conventional linear discriminant analysis and the ordinary least squares, we consider an efficient nuclear norm penalized regression that encourages a low-rank structure. Theoretical properties including a non-asymptotic risk bound and a rank consistency result are established. Simulation studies and an application to electroencephalography data show the superior performance of the proposed method over the existing approaches.

Nous proposons une nouvelle approche de l’analyse discriminante linéaire pour la classification de données provenant de matrices de grandes dimensions qui découlent couramment des études en imagerie. Motivés par l’équivalence entre l’analyse discriminante linéaire conventionnelle et la méthode des moindres carrés ordinaires, nous considérons une régression pénalisée à l’aide d’une norme nucléaire efficace favorisant une structure de rang faible. Nous établissons les propriétés théoriques, y compris un risque non-asymptotique ainsi que la convergence d’un rang. Des études de simulation de données et une application à des données électroencéphalographiques indiquent une performance supérieure de la méthode proposée comparativement aux approches existantes.

[Wednesday June 14/Mercredi 14 juin, 15:30-16:00]

Cui Guo (University of Michigan), **Timothy D. Johnson** (University of Michigan)

Scalar on Image Regression with Application to Multiple Sclerosis MRI Lesion Data.

Régression scalaire sur des images avec application à des données d’imagerie par résonance magnétique sur la sclérose en plaques

Multiple sclerosis (MS) is an autoimmune disease that attacks the central nervous system. Magnetic resonance imaging (MRI) plays a central role in the diagnosis and management of MS patients because damage to the myelin is visible on MRI. A research question of interest is whether these MRI images can predict MS subtype. To answer this question we propose a Bayesian scalar-on-image regression model with scalar outcome (MS subtype) and binary image (MRI) covariates. Parameters of these covariates are spatially varying and are fitted using Gaussian random fields. Our proposed model is fitted to a real data set consisting of 228 MS patients with 3 MS subtypes. A Hamiltonian Monte Carlo (HMC) algorithm is proposed to implement full Bayesian statistical inference. To reduce the computational burden, we code the problem to run in parallel on a graphical processing unit (GPU).

La sclérose en plaques (SEP) est une maladie autoimmune qui attaque le système nerveux central. L’imagerie par résonance magnétique (IRM) joue un rôle déterminant dans le diagnostic et la gestion des patients atteints de SEP puisqu’elle permet de voir les lésions à la myéline. La question de recherche est de savoir si ces images IRM peuvent prédire un sous-type de SEP. Pour répondre à cette question, nous proposons un modèle bayésien de régression scalaire pour des images avec une variable dépendante scalaire (sous-type de SEP) et des images binaires (IRM) comme covariables. Les paramètres de ces covariables varient dans l’espace et sont ajustés à l’aide des champs aléatoires gaussiens. Nous proposons un modèle ajusté à un ensemble de données réelles provenant de 228 patients atteints de SEP et trois sous-types de la maladie. Nous proposons un algorithme de Monte-Carlo hamiltonien (HMC) afin d’implémenter l’inférence statistique bayésienne complète. Pour alléger les calculs, le code est parallélisé à l’aide d’un processeur graphique (GPU).

**Recent Developments in the Analysis of Neuroimaging Data: A Showcase of the
CANSSI-CRT on ‘Joint Analysis of Neuroimaging Data’**
**Développements récents en analyse des données de neuroimagerie : vitrine sur le
PRC de l’INCASS sur “l’analyse conjointe de données de la neuroimagerie”**

[Wednesday June 14/Mercredi 14 juin, 16:00-16:30]

Mina Gheiratmand (University of Alberta), **Russell Greiner** (University of Alberta), **Matthew Brown** (University of Alberta), **Andrew Greenshaw** (University of Alberta), **Serdar Dursun** (University of Alberta)

Towards More Reliable Neuroimaging-Based Biomarkers in Mental Illness: The Case of Schizophrenia Discrimination using fMRI Data

Pour une plus grande fiabilité des marqueurs biologiques basés sur l’imagerie cérébrale en santé mentale : le cas de la discrimination en schizophrénie à l’aide de données IRM

Much work in the recent years has shown associations between schizophrenia and abnormal brain functional connectivity, which represents the pattern of interaction between different brain regions. The functional connectivity network can be derived from functional MRI (fMRI) data, which provides an indirect measure of the brain neural activity – during rest or task – at a reasonable spatial and temporal resolution. Despite much work, no reliable “statistical biomarker” has yet been identified that can be used to diagnose schizophrenia or guide treatment. Our work aims at finding reliable neuroimaging-based sets of features or “biomarkers” that are not only statistically significant, but also predictive of the disease in a novel subject and stable across different data subsets. Results of applying this approach to functional connectivity network features extracted from fMRI data of patients diagnosed with schizophrenia and healthy controls from a multi-site fMRI dataset will be presented.

Ces dernières années, de nombreux travaux ont fait état d’une association entre la schizophrénie et une anormalité de la connectivité fonctionnelle cérébrale, connectivité qui représente en fait le mode d’interaction entre différentes régions du cerveau. Le réseau de connectivité fonctionnelle peut être dérivé de données IRM fonctionnelles (IRMf) qui fournissent une mesure indirecte de l’activité neurale du cerveau – au repos ou dans l’exécution de tâches – selon une résolution spatiale et temporelle raisonnable. Malgré la quantité de travaux de recherche, aucun « marqueur biologique statistique » fiable n’a été identifié comme pouvant être utile au diagnostic de la schizophrénie ou pour en orienter le traitement. Nous nous appliquons à trouver des ensembles de caractéristiques ou « marqueurs biologiques » basés sur l’imagerie cérébrale qui, en plus d’être significatifs sur le plan statistique, sont aussi prédictifs de la maladie pour un nouveau sujet, et stables à travers différents sous-ensembles de données. Nous présentons les résultats de l’application de cette approche aux caractéristiques du réseau de connectivité fonctionnelle que nous avons tirées de données IRMf relatives à des patients ayant reçu un diagnostic de schizophrénie et d’un ensemble multisite de données IRMf relatives à des groupes témoins en santé.

Measurement Error, Nonresponse, and Sample Design Erreur de mesure, non-réponse et plan d'échantillonnage

Chair/Président: Zenaida Mateo
Room/Salle: E2 165 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 15:00-15:15]

Edward Kroc (University of British Columbia), **Bruno D. Zumbo** (University of British Columbia)

Calibration of Measurements: An Extension of Classic Measurement Error Theory

Étalonnage des mesures : une extension de la théorie classique de l'erreur de mesure

The classical notion of measurement error typically relies on a mean-zero assumption on the expectation of the errors, conditional on the data. However, in many applied problems in the medical, social and ecological sciences, such an assumption is often unreasonable. In this talk, we will define the notion of a weakly calibrated measurement for an unobservable true quantity based upon a weaker mean-zero assumption that more accurately reflects the structure of applied problems in these disciplines. We will then explore certain attractive features of this measurement error formulation, some not present in the traditional model. Finally, we will indicate how these theoretical considerations can lead to practical inferential gains in the context of some real problems from the medical, social and ecological sciences.

La notion classique d'erreur de mesure repose généralement sur une hypothèse de moyenne égale à zéro pour l'espérance des erreurs, conditionnellement aux données. Cependant, dans de nombreux problèmes pratiques en sciences médicales, sociales et écologiques, une telle hypothèse est souvent déraisonnable. Dans cet exposé, nous définirons la notion de mesure faiblement calibrée pour une quantité réelle non observable basée sur une hypothèse plus faible de moyenne égale à zéro qui reflète plus fidèlement la structure des problèmes pratiques dans ces disciplines. Nous explorerons ensuite certaines caractéristiques intéressantes de cette formulation de l'erreur de mesure, certaines non présentes dans le modèle traditionnel. Enfin, nous indiquerons comment ces considérations théoriques peuvent entraîner des gains inférentiels en pratique dans le contexte de certains problèmes réels en sciences médicales, sociales et écologiques.

[Wednesday June 14/Mercredi 14 juin, 15:15-15:30]

Taraneh Abarin (Memorial University of Newfoundland)

Non-identifiable Interaction Models with Misclassification

Modèles d'interaction non identifiables avec erreurs de classification

In statistical association studies, outcomes are determined by combinations of error-prone and accurately measured variables that interact with each other. As techniques dealing with interaction terms when error-prone variables are involved are quite challenging, in practice, we often see the use of additive models that ignore the interaction effects. However, the consequence of erroneously omitting interactions in those models is recognized as one of the main threats to efficiency of these studies. In fact, in the presence of error-prone variables, detecting the interaction term is more challenging than either the individual effects. In order to improve the accuracy and precision in the assessment these factors, one needs to take into account these errors. On the other hand, including interaction terms often raises the issue of non-identifiability. In this

Dans les études statistiques d'association, les résultats sont déterminés par des combinaisons de variables sujettes à l'erreur et mesurées avec précision qui interagissent les unes avec les autres. Étant donné que les techniques traitant des termes d'interaction en présence de variables sujettes à l'erreur posent des défis, nous voyons souvent en pratique l'utilisation de modèles additifs ignorant les effets d'interaction. Cependant, la conséquence de l'omission erronée d'interactions dans ces modèles est reconnue comme l'une des principales menaces à l'efficacité de ces études. En fait, en présence de variables sujettes à l'erreur, il est plus difficile de détecter le terme d'interaction que les effets individuels. Il faut tenir compte de ces erreurs afin d'améliorer l'exactitude et la précision de l'évaluation de ces facteurs. D'autre part, inclure les termes d'interaction soulève souvent la question de la non-identifiabilité. Dans cette présentation, je discuterai quelques études d'association en mettant principalement l'accent sur

Measurement Error, Nonresponse, and Sample Design Erreur de mesure, non-réponse et plan d'échantillonnage

presentation, I will discuss some association studies with a primary focus on interaction terms, subject to misclassification, and some remedies to deal with the issue of non-identifiable parameters.

les termes d'interaction, sujets à des erreurs de classification, et quelques remèdes pour traiter la question des paramètres non identifiables.

[Wednesday June 14/Mercredi 14 juin, 15:30-15:45]

Audrey-Anne Vallée (Université de Neuchâtel), **Yves Tillé** (Université de Neuchâtel)

Revisiting Variance Decomposition when Independent Samples Intersect

Réexaminer la décomposition de la variance lorsque des échantillons indépendants s'entrecroisent

When a sample is obtained by intersecting two independent samples, the variance and the estimator of variance of the expanded estimator can be decomposed in two different ways according to the sample used in the conditional expectations. Even if both methods give the same result, we show that with one decomposition, it is generally more practical to compute the variance and with the other one, it is more convenient to estimate the variance. These differences in the decompositions are due to simplifications in joint inclusion probabilities and this sheds light on two particular cases: the reverse approach used in the nonresponse case and the estimation of variance in two-stage sampling designs.

Lorsqu'un échantillon est obtenu en croisant deux échantillons indépendants, la variance et l'estimateur de variance de l'estimateur dilaté peuvent être décomposés de deux manières différentes selon l'échantillon utilisé dans les espérances conditionnelles. Même si les deux méthodes donnent le même résultat, nous montrons qu'avec une décomposition, il est généralement plus pratique de calculer la variance et avec l'autre, il est plus commode d'estimer la variance. Ces différences dans les décompositions sont dues à des simplifications des probabilités d'inclusion conjointes, ce qui permet de mettre en lumière deux cas particuliers : l'approche inverse utilisée dans le cas de non-réponse et l'estimation de la variance dans les plans d'échantillonnage à deux degrés.

[Wednesday June 14/Mercredi 14 juin, 15:45-16:00]

Yilin Chen (Carleton University), **Song Cai** (Carleton University)

Combining Multiple Samples through Density Ratio Models for More Efficient Estimation

Combinaison d'échantillons multiples via un modèle de rapport de densité pour une estimation plus efficace

Multiple samples are often available under different settings and more efficient inferences on a particular population may be obtained by suitable use of information from all samples. In this talk, we describe the use of empirical likelihood with Density Ratio Model (DRM) to effectively combine information from multiple samples. We show that more samples outcompetes fewer samples for its dominance in estimation efficiency, as long as DRM assumption is satisfied. Our presentation focuses on the estimation of DRM parameters and demonstrates the efficiency gain of parameter estimation by including more samples under DRM.

De multiple échantillons sont souvent disponibles pour différentes conditions. L'utilisation judicieuse de l'information de plusieurs échantillons permet d'obtenir des inférences plus efficaces concernant une population donnée. Dans cette présentation, nous décrivons l'utilisation de la vraisemblance empirique avec un modèle de rapport de densité (MRD) qui combine efficacement des informations d'échantillons multiples. Nous montrons que l'efficacité de l'estimation augmente avec le nombre d'échantillons, tant que l'hypothèse du modèle MRD est satisfaite. Notre présentation met l'accent sur l'estimation des paramètres sous ce modèle et démontre qu'il permet, grâce à l'inclusion d'un nombre supérieur d'échantillons, un gain d'efficacité.

[Wednesday June 14/Mercredi 14 juin, 16:00-16:15]

Ananthika Nirmalkanna (Memorial University of Newfoundland), **Yildiz Yilmaz** (Memorial University of Newfoundland)

Multivariate Response-Dependent Two-Phase Sampling Designs

Plans d'échantillonnage à deux phases avec dépendance multivariée sur la réponse

In some studies, it may be relatively affordable

Dans certaines études, la mesure de la variable de réponse

Measurement Error, Nonresponse, and Sample Design Erreur de mesure, non-réponse et plan d'échantillonnage

to measure the response variable, while a covariate might be expensive to obtain. In this situation, cost-efficient response-dependent two-phase sampling designs could be considered: in phase 1, we have easily measured variables including the response variable for all individuals in the cohort or in a large random sample from the population, and in phase 2, we obtain expensive variables for a subset of individuals selected according to their response variable obtained in phase 1. We consider the likelihood- and pseudo-likelihood based methods for incomplete data analysis to estimate the regression parameters. We extend the estimation methods for the setting where the second phase sampling depends on multiple response variables. The objective is to compare the efficiency of estimators, to develop an efficient sampling model specification, and to address the change in efficiency compared to the univariate response-dependent sampling.

peut être abordable, tandis que l'obtention d'une covariable peut être coûteuse. Dans ce cas, on peut songer à une solution économique : des plans d'échantillonnage à deux phase avec dépendance sur la réponse. À la première phase, nous avons des variables facilement mesurées, y compris la variable de réponse pour tous les sujets de la cohorte ou pour un vaste échantillon aléatoire de la population ; à la deuxième phase, nous obtenons des variables plus coûteuses à recueillir auprès d'un sous-ensemble de sujets choisis en fonction de leur variable de réponse obtenue à la première phase. Nous examinons les méthodes basées sur la vraisemblance et la pseudo-vraisemblance d'analyse de données incomplètes afin d'évaluer les paramètres de la régression. Nous étendons les méthodes d'estimation à la configuration de l'échantillonnage de la seconde phase qui dépend de multiples variables de réponse. Le but est de comparer l'efficacité des estimateurs, de spécifier un modèle efficace d'échantillonnage et de vérifier le changement d'efficacité, comparativement à un échantillonnage univarié dépendant de la réponse.

[Wednesday June 14/Mercredi 14 juin, 16:15-16:30]

Jianhua Hu (Shanghai University of Finance and Economics, China), **Jian Huang** (University of Iowa), **Feng Qiu** (Zhejiang Agriculture and Forestry University, China)

Response Best Subset Selection Model and Efficient Estimation in Multivariate Linear Regression

Modèle de sélection du meilleur sous-ensemble de réponse et estimation efficace dans une régression linéaire multivariée

In practice, the number of responses is not known prior to data analysis so that responses also need variable selection, for which few research has been found. In this paper, we address a response variable selection approach, propose a novelty response best subset selection (RBS) model and provide an estimation procedure to perform response best subset selection and regression coefficient estimation via penalizing unselected responses variables. Our estimations enjoy the oracle property: model consistency and asymptotic normality of regression coefficient estimators corresponding to the selected response variables. The proposed model and procedure can be extended to the situation where response variables have group effects. Our finite sample-sized sample simulation studies demonstrate that the proposed model and procedure are efficient and complete. We apply our promising approach to study a real data set and the results shed new light on selection of response variables.

Dans la pratique, le nombre de réponses est inconnu avant l'analyse de données, si bien que les réponses nécessitent également une sélection de variables pour laquelle peu d'études ont été faites. Dans cet article, nous examinons une approche de sélection des variables-réponses, nous proposons un nouveau modèle de sélection du meilleur sous-ensemble de réponses et nous fournissons une procédure d'estimation pour effectuer l'estimation de la sélection du meilleur sous-ensemble de réponses et du coefficient de régression par la pénalisation des variables-réponses non sélectionnées. Nos estimations bénéficient de la propriété d'Oracle : elles sont convergentes par rapport au modèle et normalité asymptotique des estimateurs des coefficients de régression correspondant aux variables-réponses sélectionnées. Le modèle et la procédure proposés peuvent être étendus à la situation dans laquelle les variables-réponses ont des effets de groupe. Nos études de simulation sous échantillons de taille finie démontrent que le modèle et la procédure proposés sont efficaces et exhaustifs. Nous appliquons notre approche prometteuse à l'étude d'un jeu de données réelles ; les résultats fournissent une nouvelle perspective sur la sélection des variables-réponses.

Teaching Statistics: New Perspectives Enseignement de la statistique : nouvelles perspectives

Chair/Président: Alison Gibbs
Room/Salle: E2 160 (EITC)
Abstract/Résumé

[Wednesday June 14/Mercredi 14 juin, 15:00-15:15]

Jeffrey D. Picka (University of New Brunswick)

Motivating Introductory Statistics

Favoriser l'apprentissage des statistiques de base

Students will be greatly assisted in learning statistics if they feel that it is useful and relevant to their needs. Ideas from science studies can be used to both provide possible ways of motivating student interest in statistics, and also can be used to suggest why students with good mathematical and computational skills may be completely unprepared for most of the content in an introductory statistics course.

Ce serait une aide précieuse pour les étudiants soucieux de suivre un apprentissage en statistique s'ils jugeaient cette matière utile et pertinente pour eux. Les études scientifiques sont source d'idées qui peuvent être utilisées à la fois pour éveiller l'intérêt des étudiants pour les statistiques et aussi pour savoir pourquoi des étudiants par ailleurs compétents en mathématiques et en informatique peuvent ne pas avoir ce qu'il faut pour assimiler la plus grande partie du contenu d'un cours de base en statistiques.

[Wednesday June 14/Mercredi 14 juin, 15:15-15:30]

Jennifer Thornton (Mount Saint Vincent University)

Flipping the Online Learning Environment

Inverser l'environnement d'apprentissage en ligne

The flipped classroom model is gaining popularity. With its focus on hands-on learning and practical exercise, it makes sense that this would be a useful model for teaching statistics. Course notes read prior to the live class and practice problems completed in a lab setting (small groups, smaller class size than the lectures) are just a couple of ways that we are using the flipped classroom model to teach Introductory level statistics online. This model appeals to more learning styles, and also to students who want to be more self-directed. The challenge becomes making sure they are prepared for tests and exams. The purpose of my presentation is to share some tools that are working for our course and see if others have different things that have worked in their classrooms.

Le modèle de salle de classe inversé gagne en popularité. Avec son accent sur l'apprentissage pratique et l'exercice pratique, il est logique que ce soit un modèle utile pour l'enseignement des statistiques. Les notes de cours lues avant que la classe n'ait lieu et les problèmes pratiques réalisés en laboratoire (petits groupes, taille de classe plus petite que les salles de cours) ne sont que quelques façons d'utiliser le modèle de salle de classe inversé pour enseigner en ligne l'introduction aux statistiques. Ce modèle est compatible avec un plus large éventail de styles d'apprentissage et est attrayant pour les étudiants qui veulent être plus autonomes. Le défi consiste à s'assurer qu'ils sont préparés pour les tests et les examens. Le but de mon exposé est de partager quelques outils qui fonctionnent dans notre cours et de voir si d'autres personnes ont des choses différentes qui ont fonctionné dans leurs salles de classe.

[Wednesday June 14/Mercredi 14 juin, 15:30-15:45]

Anne Michele Millar (Mount Saint Vincent University), **Andrea Perreault** (Mount Saint Vincent University)

Tall Tales vs Tall Tails: Rules of Thumb for Statistical Inference

Invraisemblance c. ailes relevées : règles pratiques d'inférence statistique

There has been discussion recently within the statistical community as to whether the rule of thumb "n > 30" really ensures the central limit theorem holds. Needless to say, the conclusion to date is that

Au cours d'échanges récents, on s'est demandé dans la communauté des statisticiens si la règle pratique « n > 30 » rendait réellement ce qu'invoque le théorème central limite. Inutile de dire qu'à ce jour, on conclut qu'il s'agit d'une sim-

Teaching Statistics: New Perspectives

Enseignement de la statistique : nouvelles perspectives

this is a gross oversimplification which holds for relatively “well behaved” data. How does this impact our teaching of introductory statistics courses? We want our students to be critical as to how and when basic techniques are applied, but we want them to feel confidence in the results when these techniques are used appropriately. I will discuss a simulation study for typical procedures taught in introductory statistics, examining how and when they can be judged to be reliable.

plification exagérée qui se tient en cas de données relativement « normales ». Quel en est l’impact sur l’enseignement des cours de base en statistique ? Soucieux que les étudiants exercent leur sens critique pour déterminer de quelle façon et à quel moment ils doivent appliquer les techniques de base, nous tenons aussi à ce qu’ils aient confiance qu’en les utilisant de façon appropriée, ils peuvent se fier aux résultats. Nous traitons d’une étude en simulation de procédures types enseignées dans les cours de base en statistique pour savoir comment et à quel moment elles peuvent être jugées fiables.

[Wednesday June 14/Mercredi 14 juin, 15:45-16:00]

Sohee Kang (University of Toronto Scarborough)

Effectiveness of Gamification Feature in Introductory Statistics Course

Efficacité de la ludification dans les cours d’introduction à la statistique

Two key components to successful learning are engagement with the material and practice. The introductory Statistics course attracts a large number of students with a wide range of abilities and interests. Twelve WeBWorK homework sets are assigned to students, and they receive instant feedback while they can make infinite attempts to achieve the correct answer until the deadline. We further added gamification feature in the exist WeBWorK, which is designed to increase the amount of time that students want to spend on homework. The gamification techniques we made use of are “leveling” and “achievement”. We created five levels from “Novice” to “Junior Statistician”, five challenge questions, and twelve various achievement badges. In general the achievement badges were constructed to reward students for practicing good homework habits and to encourage them to solve every problem. We will share the result and discuss the effectiveness of these gamification features.

L’engagement et la pratique sont deux éléments clés d’un apprentissage réussi. Les cours d’introduction à la statistique attirent un grand nombre d’étudiants avec des bagages et des intérêts variés. Nous donnons à des étudiants douze exercices WeBWorK : ils obtiennent une évaluation instantanée et peuvent répondre aux questions autant de fois que voulu jusqu’au temps limite pour obtenir la meilleure note. Nous ajoutons encore un élément de ludification au système WeBWorK existant, qui est conçu pour augmenter le temps que les étudiants ont envie d’y passer. Les techniques que nous utilisons sont l’évolution et la réalisation d’objectifs. Nous créons cinq niveaux, de « Novice » à « Statisticien junior », cinq questions-défis et douze badges de réalisations. En général, les badges servent à récompenser les étudiants pour leurs bonnes habitudes d’étude et de les encourager à résoudre chaque problème. Nous partageons les résultats de cette étude et discutons de l’efficacité de ces fonctions de ludification.

[Wednesday June 14/Mercredi 14 juin, 16:00-16:15]

Sotirios Damouras (University of Toronto Scarborough), **Sohee Kang** (University of Toronto Scarborough)

The Status of Statistics Curricula in Canada

L’état des programmes d’études en statistiques au Canada

This talk will review the current state of undergraduate Statistics curricula at universities across Canada. We will present both quantitative and qualitative information on the structure and composition of Major programs in Statistics. More specifically, we will look at the number and type of course requirements for each program, the learning outcomes they serve, the topics and skills they

Cette allocution porte sur l’état actuel des programmes d’études en statistiques de premier cycle dans les universités canadiennes. Nous présentons de l’information quantitative et qualitative sur la structure et la composition des programmes de spécialisation en statistiques. Nous nous attardons plus précisément sur le nombre et le type d’exigences pour les cours de chaque programme, les fins d’apprentissage de chacun, les matières et compétences acquises et d’autres

Teaching Statistics: New Perspectives

Enseignement de la statistique : nouvelles perspectives

develop, as well as other relevant information. The talk intends to give an overview of how we collectively educate Statisticians, with the ultimate goal of helping identify directions for future curricular development.

renseignements pertinents. Nous voulons donner un aperçu de l'éducation que nous tous prodiguons aux statisticiens, avec pour objectif final d'aider à définir les orientations pour l'élaboration de futurs programmes.

Author List • Liste des auteurs

Ng, Raymond T., 24, 82

Abarin, Taraneh, 37, 53, 155, 156, 241

Acar, Elif Fidan, 27, 43, 50, 100, 187, 221

Akakpo, Nathalie, 50, 222

Alam, Shomoita, 44, 193

Albert, Paul S., 25, 86

Albert-Green, Alisha, 19, 62

Almutiry, Waleed, 49, 219

Altman, Rachel MacKay, 135

Alvo, Mayer, 21, 71

Amu, Agnes Nessie, 27, 100

Andrews, Jeffrey L., 40, 173

Andrews, Jessica, 34, 137

Andronescu, Mirela, 49, 215

Arashi, Mohammad, 40, 173

Arshadi, Sahar, 37, 156

Arumugam, MohanaGowri, 37, 155

Asghari, Shabnam, 36, 153

Ashleik, Naeima, 23, 78

Awosoga, Oluwagbohunmi, 39, 167

Ayilara, Olawale Fatai, 40, 173

Bae, Taehan, 34, 134

Baerg, Krista, 40, 171

Bakal, Jeffrey A., 18, 58

Balamchi, Shabnam, 31, 120

Balbàs, Alejandro, 27

Balbás, Alejandro, 102

Bareinboim, Elias, 48, 211

Barry, Amadou Diogo, 21, 74

Beaumont, Jean-François, 25, 88

Bee, Marco, 50, 222

Belalia, Mohamed, 52, 237

Bell, Rhonda, 21, 76

Benchimol, Eric, 28, 105

Benedetti, Andrea, 20, 69

Bergeron, Pierre-Jerome, 21, 71

Bhatnagar, Sahir R., 35, 144

Bickel, David R., 43, 188

Bickis, Mikelis Guntars, 23, 78

Bierling, Vanessa, 26, 91

Bilayi-Biakana, Clemonell, 21, 71

Bilder, Christopher, 36, 152

Bin, Junchi, 22, 37, 151

Blais, Lucie, 28, 48, 105, 225

Boaz, Ray, 19, 61

Bosa, Kevin, 35, 139

Boteler, Claire, 36, 147

Bouezmarni, Taoufik, 21, 75

Braun, John, 31, 119

Brenner, Bluma, 50, 219

Brewster, John F., 52, 233

Brisebois, François, 35, 138

Brobbey, Anita, 45, 198

Brown, Matthew, 53, 240

Brown, Patrick E., 19, 62

Brown, Robert L., 52, 235

Bryan, Jenny, 24, 82

Bunge, Markus, 22

Burke, Kevin, 34, 134

Burkett, Kelly M., 18, 56

Bégin, Jean-François, 45, 199

Cadigan, Noel, 37, 155, 157

Cai, Kaida, 26, 96

Cai, Song, 53, 242

Campbell, Dave, 39, 169

Canji, David, 22

Canty, Angelo J., 49, 215

Cao, Jiguo, 21, 33, 75, 132

Capistran, Marcos A., 20, 66

Carenini, Giuseppe, 24, 82

Carey, Michelle, 20, 41, 64, 177

Chakraborty, Adrijo, 39, 164

Chakraborty, Bibhas, 20, 68

Chambers, Larry, 29, 109

Chang, Bo, 52, 236

- Chang, Hsing-Ming, 33, 130
 Chang, Yi Meng, 22
 Chaoubi, Ihsan, 46, 203
 Charest, Anne-Sophie, 19, 59
 Charpentier, Arthur, 21, 74
 Chatterjee, Snigdhasu, 39, 164
 Chaudhry, Mohan, 21, 71
 Che, Menglu, 21, 76
 Chekouo Tekougang, Thierry, 40, 169
 Chen, Li-Pang, 26, 96
 Chen, Lichen, 41, 180
 Chen, Yilin, 53, 242
 Cheng, Smiley, 52, 233
 Chkrebtii, Oksana, 20, 66
 Chowdhury, Monsur, 44, 195
 Chu, Haitao, 48, 212
 Chu, Jingjia, 37, 154
 Ciborowski, Jan J.H., 41, 176
 Cigsar, Candemir, 26, 30, 37, 94, 115, 157
 Clark, Robert, 44, 192
 Coache, Anthony, 38, 152
 Coia, Vincenzo, 24, 82
 Cook, Richard J., 24, 50, 84, 223
 Cossette, Hélène, 45, 46, 200, 203
 Cotton, Cecilia, 20, 68
 Coulombe, Janie, 23
 Cowen, Laura L.E., 30, 113
 Craiu, Radu, 47, 207
 Csorgo, Miklos, 29, 112
 Cui, Yifan, 25, 85
 Cunningham, Charles, 36, 146

 Dai, Ruixue, 22
 Daignault, Katherine, 44, 193
 Dalla Riva, Giulio V., 24, 82
 Daly, Caitlin, 36, 146
 Damouras, Sotirios, 54, 245
 Daoust, Pierre, 34, 137
 Dasgupta, Kaberi, 28, 105
 Datta, Gauri, 25, 39, 87, 164
 Davies, Katherine, 36, 148
 Day, Troy, 36, 147
 de Leon, Alexander R., 26, 33, 93, 131
 de Souza, Camila P.E., 49, 215
 Dean, Charmaine B., 30, 41, 42, 115, 176, 182
 Deardon, Rob, 36, 40, 49, 147, 170, 219
 Dehling, Herold, 24, 81
 Dey, Rajib, 37, 155

 Diedrichsen, Jörn, 23, 80
 Dillman, Don A., 18, 55
 Dinsdale, Daniel, 41, 177
 Dinu, Irina, 27, 31, 99, 121
 Do, Kim-Anh, 40, 169
 Dolovich, Lisa, 29, 109
 Dongmo Jiongo, Valéry, 45, 199
 Dowd, Mike, 20, 64
 Dubin, Joel, 49, 216
 Duchesne, Thierry, 47, 52, 206, 237
 Dupuis, Annie, 37, 150
 Dupuis, Debbie J., 50, 222
 Durocher, Stephane, 35, 141
 Dursun, Serdar, 53, 240
 Dyck, Justin, 22

 El Kasmi, Yassin, 45, 200
 Escobar, Michael, 37, 150

 Fan, Zhaozhi, 50, 221
 Fang, Yizhou, 35, 142
 Farewell, Vern, 52, 235
 Feng, Xiaolan, 45, 197
 Feng, Xin (Cindy), 19, 62
 Ferland, René, 38, 152
 Ferreira Guerra, Steve, 48, 225
 Flournoy, Nancy, 47, 209
 Forget, Amélie, 48, 225
 Foster, Gary, 36, 146
 Foster, Jared C., 25, 86
 Franczak, Brian, 37, 150
 Froda, Sorana, 38, 152
 Fu, Yunting, 22
 Fung, Karen, 41, 176

 Gagnon, François, 35, 139
 Gambino, Jack, 18
 Gao, Yu, 26, 91
 Garcia, Yury E., 20, 66
 Gardiner, Bryan, 37, 151
 Garrett, Rose, 22
 Garrido, Jose, 27, 102
 Gelbart, Michael A., 24, 82
 Ghahramani, Melody, 33, 131
 Gheiratmand, Mina, 53, 240
 Ghodsi, Ali, 32, 126
 Gibbs, Alison L., 33, 128
 Gong, Kaibo, 39, 164
 Gorun, Samantha, 22

- Graham, Jinko, 27, 31, 98, 119
 Greenlaw, Keelin, 49, 215
 Greenshaw, Andrew, 53, 240
 Greenwood, Celia M.T., 27, 31, 35, 49, 100, 120, 144, 214, 215
 Greiner, Russell, 53, 240
 Griffith, Skye, 22
 Grinek, Stepan, 27, 100
 Grover, Kanika, 23
 Guan, Zixiang, 22
 Guo, Cui, 53, 239
 Guo, Danqiao, 25, 89
 Gustafson, Paul, 24, 30, 82, 113
 Guttmann, Astrid, 20, 69
 Gweon, Hyukjun (Jay), 44, 196
- Halladay, Amanda, 35, 138
 Han, Peisong, 20, 29, 67, 111
 Hanley, James A., 43, 190
 Hariri, Ofir, 19, 62
 Hasan, Tariqul, 50, 222
 Hasler, Caren, 47, 207
 Hatefi, Armin, 44, 196
 Haziza, David, 25, 34, 88, 137
 He, Fang, 31, 121
 Henrey, Andrew, 34, 135
 Heyman, Megan, 39, 164
 Hidirolou, Mike, 25, 87
 Hiebert, Ian, 22
 Hill, Michael, 39, 167
 Himelein, Kristen, 43, 191
 Hong, Hanping, 25, 89
 Hoque, Md. Erfanul, 41, 179
 Horvath, Lajos, 41, 179
 Hossain, Md. Shakhawat, 40, 173
 Hsu, Grace Guan, 23
 Hu, Jianhua, 53, 243
 Hu, Pingzhao, 45, 197
 Hu, Rui, 44, 195
 Hu, Wei, 53, 239
 Hu, X. Joan, 24, 31, 40, 83, 119, 170
 Hu, Zheng Jing, 22
 Hua, Lei, 38, 160
 Huang, Jian, 53, 243
 Huang, Mei Ling, 40, 172
 Huo, Dezheng, 30, 113
 Hurley, Oliver, 36, 153
 Huynh, Kim P., 44, 191
- Islam, Md. Mohaiminul, 23, 45, 197
 Ivanoff, Gail, 21, 71
- Jafari Jozani, Mohammad, 40, 44, 173, 196
 Jiang, Min, 34, 138
 Jiang, Wenjun, 25, 89
 Jin, Xuejing, 36, 146
 Joe, Harry, 32, 52, 123, 236
 Johara, Fatema, 23
 Johnson, Timothy D., 53, 239
 Johnston, Steven, 44, 192
 Jolicoeur-Martineau, Alexia, 35, 144
 Juarez-Colunga, Elizabeth, 42, 183
- Kaczorowski, Janusz, 29, 109
 Kang, Sohee, 54, 245
 Karlis, Dimitris, 37, 156
 Karunarathna, Charith Bhagya, 27, 98
 Kellner, Ralf, 38, 162
 Khan, Shahedul A., 34, 40, 135, 172
 Khodayari Moez, Elham, 27, 31, 99, 121
 Khosa, Saima K., 34, 135
 Kim, Jae Kwang, 47, 207
 Kim, James, 21, 71
 Kokoszka, Piotr, 23, 80
 Kolkiewicz, Adam, 41, 180
 Kong, Dehan, 53, 239
 Kong, Linglong, 21, 34, 76, 133
 Kopciuk, Karen, 38, 52, 151, 235
 Koralage, Janaki Thakshila, 50, 221
 Kosorok, Michael, 25, 85
 Kroc, Edward, 53, 241
 Kulik, Rafal, 21, 71
 Kulperger, Reg, 31, 37, 121, 154
- Labbe, Aurélie, 27, 31, 49, 50, 100, 120, 214, 219
 Labuda, Aleks, 30, 117
 Lai, Yuanhao, 41, 180
 Landriault, David, 38, 161
 Larribe, Fabrice, 18, 57
 Lawless, Jerald F., 24, 51, 83, 233
 Lawson, Andrew, 19, 61
 Leblanc, Alexandre, 18, 35, 141
 Lee, Chel Hee, 40, 171
 Lee, J. Jack, 23, 79
 Lee, Jooyoung, 50, 223
 Lefebvre, Geneviève, 28, 105
 Lefebvre, Isabelle, 25, 88
 Lemieux, Christiane, 21, 72

- Lethebe, Brendan Cord, 28, 106
 Leung, Andy, 51, 229
 Li, Bin, 38, 161
 Li, Chun, 51, 228
 Li, Haocheng, 26, 45, 48, 93, 197, 224, 225
 Li, Jia, 22, 26, 93
 Li, Jiaxiu, 35, 141
 Li, Juan, 20, 64
 Li, Lennon, 19, 62
 Li, Longhai, 19, 62
 Li, Pengfei, 36, 146
 Li, Shuanming, 45, 201
 Li, Tian, 40, 170
 Li, William, 44, 193
 Li, Yunjing, 22
 Lian, Qinshu, 48, 212
 Liang, Kun, 36, 146
 Liaw, Chris, 42, 184
 Lim, Yongho, 30, 115
 Lin, Hongcan, 46, 203
 Lin, Lillian, 23
 Lin, Zhenhua, 33, 132
 Ling, Yun, 35, 142
 Liu, Aiyi, 25, 86
 Liu, Danping, 25, 86
 Liu, Dongmeng, 31, 119
 Liu, Hai Yan, 21, 71
 Liu, Juxin, 23, 30, 78, 113
 Liu, Lin, 44, 194
 Liu, Qi, 51, 228
 Liu, Zheng, 37, 151
 Lix, Lisa M., 28, 45, 105, 198
 Lkabous, Amine Mohamed, 27, 102
 Lockhart, Richard, 21, 40, 75, 171
 Loredó-Ostí, J. Concepcion, 18, 56
 Lou, Wendy, 30, 114
 Love, Matthew, 22
 Lu, Xuewen, 26, 96
 Lu, Yi, 45, 46, 200, 201, 203
 Lu, Zihang, 21, 74
 Lyon, Martha E., 40, 171
 Lysy, Martin, 30, 35, 117, 142
 Ma, Jinhui, 29, 109
 Ma, Renjun, 35, 50, 141, 222
 Ma, Yue, 23
 MacKay Altman, Rachel, 30, 34, 116
 Mackay, Jon, 32, 126
 MacKenzie, Gilbert, 34, 134
 Macpherson, Brian, 52, 233
 MacQueen, Sarah A., 36, 153
 Mahdavian, Masoud, 36, 153
 Mahsin, Md., 36, 147
 Maiti, Taps, 39, 164
 Manalo, Joshua, 22
 Mandal, Abhyuday, 39, 164
 Mandal, Saumen, 44, 195
 Marceau, Etienne, 45, 46, 200, 203
 Mariathas, Hensley Hubert, 36, 153
 Marshall, François Alastair, 49, 218
 Martsynyuk, Yuliya V., 29, 112
 Mason, Levi James, 22
 Matira, Kevin, 37, 155
 McCandless, Lawrence C., 39, 40, 166, 170
 McGrath, Sean, 20, 23, 69
 McGregor, Kevin D.J., 31, 120
 McIsaac, Michael A., 48, 224
 Mcleish, Don, 35, 142
 McLeod, A. Ian, 41, 49, 180, 218
 McNicholas, Paul D., 19, 26, 30, 37, 58, 91, 117, 150, 155, 156, 158
 Mehrabian, Abbas, 42, 184
 Menon, Bijoy, 39, 167
 Messer, Karen, 44, 194
 Metzler, Adam, 38, 162
 Millar, Anne Michele, 54, 244
 Mills, Shirley E., 50, 227
 Mireuta, Matei, 34, 137
 Mitric, Ilie Radu, 27, 102
 Mitsakakis, Nicholas, 22, 36, 76, 148
 Molinaro, Annette M., 25, 85
 Monette, Georges A., 33, 128
 Moodie, Erica E.M., 20, 29, 39, 43, 44, 47, 68, 110, 166, 186, 193, 209
 Moon, Nathalie, 24, 84
 Moradi, Fahimeh, 31, 121
 Murdoch, Duncan, 31, 121
 Murphy, Gail C., 24, 82
 Nadeem, Khurram, 41, 176
 Nakhla, Meranda, 20, 69
 Nasari, Masoud, 29, 112
 Nasri, Bouchra, 21, 75
 Natarajan, Loki, 44, 194
 Nathoo, Farouk, 19, 61
 Ndwapi, Nkumbuludzi, 26, 91

- Neslehova, Johanna, 52, 237
 Neusy, Elisabeth, 25, 87
 Nguyen, Paul, 19, 62
 Ni, Jiayi, 28, 105
 Nieuwoudt, Christina M., 27, 98
 Nirmalkanna, Ananthika, 53, 242
 Nirmalkanna, Kunasekaran, 26, 94
 Nkurunziza, Sévérin, 21, 72
 Norouzirad, Mina, 40, 173

 Okhrati, Ramin, 27, 102
 Oldford, Wayne, 52, 235
 Olshen, Adam, 25, 85
 Oualkacha, Karim, 21, 49, 74, 214, 216
 Ouhourane, Mohamed, 49, 216
 Ozturk, Omer, 44, 196

 Pathak, Dola, 36, 152
 Paul, Lauren, 48, 224
 Peng, Defen, 34, 134
 Peng, Yingwei, 42, 182
 Perreault, Andrea, 54, 244
 Perreault, Samuel, 52, 237
 Pesevski, Angelina, 37, 150
 Picka, Jeffrey D., 53, 244
 Pitselis, Georgios, 27, 102
 Pizzi, Nick, 17
 Plan, Yaniv, 42, 184
 Platt, Robert W., 20, 29, 68, 107
 Poirier, Gabrielle, 35, 139
 Porth, Lysa, 38, 160
 Prabuddha, Inesh, 23
 Pullenayegum, Eleanor M., 29, 107

 Qin, Yingli, 49, 218
 Qiu, Derek, 43, 189
 Qiu, Feng, 53, 243
 Qiu, Shi, 19, 62
 Quach, Harris, 26, 93
 Quan, Hude, 28, 106
 Quessy, Jean-François, 52, 236, 237

 Rahme, Elham, 20, 28, 69, 105
 Ramezan, Reza, 35, 144
 Ramsay, James O., 20, 41, 64, 177
 Ramsay, Kelly, 23, 35, 141
 Rao, J.N.K., 25, 87
 Rat, Ramona, 40, 172
 Ratnasekera, Pulindu, 43, 189

 Reid, Nancy, 39, 44, 166, 196
 Ren, Jiandong, 25, 89
 Reyna Vargas, Myrtha E., 22, 76
 Ricciuti, Sarah, 37, 158
 Rice, Gregory, 41, 49, 179, 216
 Rivest, Louis-Paul, 47, 52, 207, 236
 Robison, Scott, 46, 205
 Roger, Michel, 50, 219
 Roick, Tyler, 37, 156
 Ross, Nick, 32, 127
 Rostamiforooshani, Mehdi, 22
 Rémillard, Bruno, 21, 75

 Saarela, Olli, 29, 44, 110, 193
 Safari, Abdollah, 30, 116
 Saha, Sudipta, 22
 Sahakyan, Yeva, 22
 Sajobi, Tolulope T., 28, 39, 45, 106, 167, 198
 Salibian-Barrera, Matias, 41, 177
 Sanders, Barbara, 45, 200
 Sang, Peijun, 21, 33, 75, 132
 Saunders, David, 46, 203
 Scealy, Janice Lea, 51, 230
 Schachar, Russell, 37, 150
 Schiopu-Kratina, Ioana, 21, 71
 Schmid, Christopher, 48, 212
 Schnitzer, Mireille E., 43, 48, 186, 225
 Schnitzer, Mireille E. , 211
 Schnurr, Alexander, 24, 81
 Schonlau, Matthias, 44, 196
 Scory, Tayler Dawn, 38, 151
 Scott, Carlone, 22
 Senadeera, Prageeth Manujaya, 37, 157
 Sendova, Kristina, 46, 204
 Seng Tan, Ken, 38, 160
 Shah, Sohrab, 49, 215
 Sharma, Shubham, 22
 Shen, Hua, 26, 43, 48, 96, 186, 224
 Shen, Lingzhu, 22
 Shen, Weining, 53, 239
 Shen, Yi, 42, 184
 Shepherd, Bryan E., 51, 228
 Sheriff, John, 28, 104
 Shi, Fang, 22
 Shi, Yidan, 26, 94
 Shlomo, Natalie, 19, 59
 Shokoohi, Farhad, 49, 214
 Shrier, Ian, 34, 48, 134, 211

- Shu, Di, 20, 67
 Siddiqua, Joynob Ara, 23
 Simard, Clarence, 42, 184
 Simard, Marc, 20, 69
 Simms, Alvin, 36, 153
 Simoneau, Gabrielle, 20, 68
 Singh, Charanpal, 22
 Sinha, Sanjoy Kumar, 39, 47, 164, 210
 Sneddon, Gary, 50, 222
 Sobhan, Shamsia, 23
 Soleymani, Samira, 49, 218
 Soltanifar, Mohsen, 37, 150
 Somers, Julian M., 40, 170
 Song, Hui, 42, 182
 Song, Peter X., 23, 78
 Song, Yin, 19, 61
 Spinelli, John, 24, 83
 St-Pierre, Julien, 49, 214
 Stafford, Jamie, 19, 62
 Stallard, Jim, 46, 205
 Steele, Russell, 20, 34, 39, 48, 69, 134, 166, 211
 Steeves, Holly N., 51, 230
 Steiner, Stefan, 24, 44, 82, 196
 Stephens, David A., 29, 39, 44, 47, 49, 50, 110, 166, 193, 209, 214, 219
 Stevens, Nathaniel T., 43, 186
 Stewart, Connie, 51, 232
 Stingo, Francesco, 40, 169
 Stinner, Mark, 17, 19, 60
 Strawderman, Robert, 25, 85
 Su, Chien-Lin Mark, 34, 134
 Su, Jianxi, 38, 160
 Su, Wanhua, 45, 196
 Sun, Bingrui (Cindy), 37, 155
 Sun, Lei, 27, 46, 99, 202
 Sun, Shaun Zheng, 40, 171
 Sutton, Mitchell, 22

 Tait, Peter, 30, 117
 Talbot, Denis, 28, 105
 Taljaard, Monica, 29, 109
 Tang, Thai Son, 22
 Taniguchi, Yoshihiro, 21, 72
 Taylor, Steve, 30, 41, 115, 176
 Thabane, Lehana, 29, 36, 39, 109, 146, 167
 Thiessen, Michelle, 23
 Thompson, Mary E., 26, 52, 94, 235
 Thomson, David, 36, 147
 Thomson, Trevor, 31, 119
 Thornton, Jennifer, 54, 244
 Thorpe, Kevin E., 39, 167
 Tian, Ye, 23
 Tian, Zhaoyang, 36, 146
 Tichon, Jenna G., 47, 205
 Tieu, Jenny, 40, 172
 Tillé, Yves, 47, 53, 208, 242
 Timbers, Tiffany A., 24, 82
 Tintle, Nathan, 17, 28, 104
 Tomal, Javed H., 41, 176
 Torabi, Mahmoud, 25, 41, 87, 179
 Torkashvand, Elaheh, 49, 216
 Toupin, Marie-Hélène, 52, 236
 Trapin, Luca, 50, 222
 Trufin, Julien, 27, 102
 Tsagris, Michail, 51, 232
 Tsai, Cary Chi-Liang, 32, 124
 Tu, Dongsheng, 42, 182
 Turgeon, Maxime, 27, 100
 Tyas, Suzanne, 26, 94
 Tyson, Rebecca, 36, 153

 Vallée, Audrey-Anne, 47, 53, 208, 242
 Van Bussel, Melissa, 22
 van Buuren, Stef, 17
 Vandal, Nathalie, 20, 69
 Vanniyasingam, Thuva, 36, 146
 Vershynin, Roman, 42, 184
 Villandré, Luc, 50, 219
 Volterman, William, 36, 148

 Wagner, Brandie, 42, 183
 Wallace, Michael, 29, 47, 110, 209
 Wang, Huijing, 24, 83
 Wang, Jane, 35, 139
 Wang, Liangliang, 33, 132
 Wang, Ran, 23
 Wang, Suxin, 45, 200
 Wang, Xiao, 34, 133
 Wang, Yang, 45, 197
 Wang, Youdan, 48, 212
 Wang, Yueleng, 21, 72
 Wang, Zhongya, 23
 Warriyar, Vineetha K. V., 40
 Warriyar, Vineetha K.V., 170
 Watier, Francois, 45, 200
 Watt, Sarah, 22

- Wazana, Ashley, 35, 144
 Weaver, Colin, 28, 106
 Wei, Xiaoyu, 23
 Welsh, Alan, 51, 230
 Weng, Chengguo, 25, 46, 89, 203
 Weygang, Mark, 22
 White, Scott S., 22, 33, 131
 Wiebe, Samuel, 45, 198
 Wiens, Doug, 44, 195
 Williamson, Tyler, 28, 45, 106, 198
 Wilson, James David, 32, 126
 Wirjanto, Tony, 25, 41, 89, 180
 Woolford, Douglas, 41, 176
 Wu, Adelaide Di, 32, 124
 Wu, Changbao, 20, 34, 67, 137
 Wu, Jingjing, 43, 45, 186, 197
 Wu, Sin-Hong, 33, 130
 Wu, Siyu, 22
 Wu, Steven, 30, 116
 Wu, Yilei, 49, 218

 Xi, Da Zhong (Dexen), 30, 115
 Xie, Yuying, 20, 68
 Xu, Di (Cindy), 38, 161
 Xu, Wei, 43, 187
 Xu, Xiaojian, 47, 210
 Xu, Yue, 22

 Yan, Guohua, 35, 141
 Yan, Ying, 48, 224
 Yang, Chung-Ming, 33, 130
 Yang, Jian, 45, 197
 Yang, Lusi, 22
 Yang, Ming Jie, 35, 138
 Yang, Po, 44, 193
 Yang, Shu, 47, 207
 Yang, Su-Fen, 33, 130
 Yang, Yi, 35, 49, 144, 216
 Yao, Fang, 33, 132
 Yates, Bryan, 30, 117
 Yi, Grace, 20, 48, 67, 225
 Yilmaz, Yildiz, 53, 242
 Yin, Yue, 22
 Yohai, Victor J., 51, 229
 You, Jiaying, 23
 You, Jinhong, 40, 174
 Yu, Hao, 37, 154
 Yuan, Yan, 21, 76

 Yue, Jiajia, 37, 157
 Yung, Wesley, 25, 87

 Zamar, Ruben H., 51, 229
 Zemanick, Edith, 42, 183
 Zeng, Leilei, 24, 26, 84, 94
 Zhang, Lidan (Lucy), 46, 203
 Zhang, Lin, 27, 99
 Zhang, Ruixi, 46, 204
 Zhang, Shixiao, 20, 67
 Zhang, Ting, 22
 Zhang, Yuan, 36, 146
 Zhang, Zhengwu, 34, 133
 Zhao, Joanna, 43, 189
 Zhao, Kaiqiong, 23, 50, 221
 Zhao, Puying, 34, 137
 Zheng, Xin, 22
 Zhong, Lehang, 22
 Zhou, Haoxuan, 39, 169
 Zhou, Hua, 53, 239
 Zhou, Ling, 23, 78
 Zhou, Mai, 25, 85
 Zhou, Rui, 32, 124
 Zhu, Hongtu, 34, 133
 Zhu, Mu, 49, 218
 Zhu, Ruqing, 25, 85
 Zhu, Wenjun, 38, 160
 Zhu, Yayuan, 24, 83
 Zhu, Yeying, 20, 68
 Zumbo, Bruno D., 53, 241

