



43rd Annual Meeting
of the
Statistical Society of Canada

43^e Congrès annuel
de la
Société statistique du Canada

June 14 - 17, 2015

14 au 17 juin 2015

Dalhousie University

Table of Contents • Table des matières

Table of Contents • Table des matières	1
Welcome to • Bienvenue à Halifax	2
Sponsors • Commanditaires	3
Exhibitors • Exposants	3
Organizers • Organismes	4
General Information • Informations générales	5
The Conference • Le congrès	7
Social Events • Activités sociales	9
Committees and Meetings • Comités et réunions	11
Social and Information Events • Événements sociaux et informatifs	14
List of Sessions • Liste des sessions	16
Program • Programme	19
Abstracts • Résumés	64

Welcome to • Bienvenue à Halifax

Dalhousie University welcomes you to the 43rd Annual Meeting of the Statistical Society of Canada. Founded in 1818, Dalhousie consists of three Halifax campuses. Studley, the location of most events for the meeting, is the largest of the three and houses the Faculties of Arts, Science, Computer Science and administrative buildings. Carleton, which is along the way from hotels to the conference, is near two hospitals and home to the Faculties of Medicine, Dentistry and Health Professions. Farther down and nearer to the harbour is Sexton, home of the Faculty of Engineering. Dalhousie has 18,500 students, 1,000 faculty and is a research-intensive university with approximately \$140 million in annual research funding.

With a population of over 400,000 people, Halifax is the capital city of Nova Scotia and the educational, cultural and economic hub of Atlantic Canada. Halifax boasts a rich history and scenic waterfront. The site of the conference and main hotels are all within walking distance to the downtown core where you'll find art galleries, museums, theatres, festivals, sport events, lively pubs, quaint cafés and, of course, friendly Haligonians. Check out shops, restaurants and historic buildings as you stroll along the world's longest harbourfront boardwalk.

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

L'Université Dalhousie vous souhaite la bienvenue au 43^{ième} congrès annuel de la Société statistique du Canada. Fondée en 1818, Dalhousie est constituée de trois campus à Halifax. Studley, l'endroit de la plupart des événements du congrès, est le plus grand des trois et comprend les facultés des arts, des sciences, de l'informatique et des édifices administratifs. Carleton, qui est sur le chemin entre les hôtels et le congrès, est proche de deux hôpitaux et comprend les facultés de médecine, de médecine dentaire et des professions de la santé. Plus proche du port on trouve Sexton, qui comprend la faculté de génie. Dalhousie a 18 500 étudiants, 1000 professeurs et constitue un institut de recherche avec environ 140 millions de dollars en subventions de recherche annuelles.

Avec une population de 400 000 habitants, Halifax est la capitale de la Nouvelle-Écosse et le centre éducatif, culturel et économique des provinces de l'Atlantique. Halifax se vante d'une histoire riche et d'un bord de mer panoramique. Le centre-ville est accessible à pied du site du congrès et des hôtels principaux, où on peut trouver des galeries d'arts, des musées, des théâtres, des festivals, des événements sportifs, des pubs animés, de charmants cafés, et bien sûr, des Haligoniens amicaux. Visitez des boutiques, des restaurants et des édifices historiques en marchant le long de la plus longue promenade portuaire au monde.

Pour plus d'information pour planifier votre voyage, visitez le site web <http://www.novascotia.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 la généreuse contribution a rendu possible la tenue de ce congrès :

- President's Office, Dalhousie University
- Faculty of Science, Dalhousie University
- Department of Mathematics and Statistics, Dalhousie University
- Atlantic Association for Research in the Mathematical Sciences
- The Fields Institute
- Pacific Institute for the Mathematical Sciences
- Centre de Recherches Mathématiques
- Canadian Statistical Sciences Institute /
Institut canadien des sciences statistiques

Exhibitors • Exposants

The exhibitors, providing displays for examination and purchase, can be found in the McInnes Room of the Student Union Building (SUB). The exhibitors will be available Monday, June 15 and Tuesday, June 16 from 9:00 am to 4:00pm.

Les kiosques des exposants, pour bouquiner ou acheter, sont situés dans la salle McInnes du Student Union Building (SUB). Les exposants seront présents lundi 15 juin et mardi 16 juin de 9 h à 16 h.

- CRC Press <http://www.crcpress.com>
- MacMillan Education <http://www.macmillanhighered.com>
- Tech Observer <http://www.tech-observer.com/>
- Wiley <http://ca.wiley.com/WileyCDA>

Organizers • Organisateur

Local Arrangements Committee • Comité des arrangements locaux

- Edward Susko
Dalhousie University
- Joanna Mills-Flemming
Dalhousie University
- David Hamilton
Dalhousie University
- Chris Field
Dalhousie University
- Queena Crooker-Smith
Dalhousie University

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

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

Program Committee • Comité du programme

- Paul Gustafson (Chair • Président) *University of British Columbia*
- Lehana Thabane *McMaster University*, Biostatistics - Biostatistique
- Reg Kulperger *Western University*, Probability - Probabilité
- Paul McNicholas *McMaster University*, Business and Industrial Statistics- Industrielle et Gestion
- Wesley Yung *Statistics Canada*, Survey Methods - Enquête
- Bethany White *Western University*, Statistical Education - Éducation en statistique
- Jose Garrido *Concordia University*, Actuarial Science - Science Actuarielle

General Information • Informations générales

Directions • Emplacement

Campus maps are available on the conference website and in this program. For those staying at any of the conference hotels, walking directions to the McCain building are available at <http://www.ssc.ca/en/meetings/2015/accommodations>

Des cartes du campus sont disponibles sur le site web du congrès et dans ce programme. Pour ceux qui s'hébergent dans n'importe quel des hôtels du congrès, les itinéraires pour se rendre à l'édifice McCain sont disponibles sur <http://www.ssc.ca/fr/congrès/2015/hébergement>

Registration • Inscription

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

- Sunday, June 14: 8:00 am to 5:30 pm
McCain Building
- Monday, June 15: 7:30 am to 5:00 pm
McCain Building
- Tuesday, June 16: 8:15 am to 5:00 pm
McCain Building
- Wednesday, June 17: 8:15 am to 4:00 pm
McCain Building

Les participants déjà inscrits peuvent venir chercher leur matériel de conférence 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 14 juin : 08 h à 17 h 30
McCain Building
- Lundi 15 juin : 7 h 30 à 17 h
McCain Building
- Mardi 16 juin : 8 h 15 à 17 h
McCain Building
- Mercredi 17 juin : 8 h 15 à 17 h
McCain Building

Parking on Campus • Stationnement sur le campus

Guests staying on campus can request free parking passes when they arrive. Parking passes for off-campus guests can be purchased from Security Services at (902) 893-4190 for \$14.50. A visitor parking location map is available at <http://www.ssc.ca/en/meetings/2015/transportation>. Please note that the Dunn Parking lot is restricted to on-campus guests.

Les visiteurs hébergés sur le campus peuvent demander des passes de stationnement gratuites lorsqu'ils arrivent. Des passes de stationnement pour les visiteurs hors campus peuvent être achetées au service de sécurité au (902) 893-4190 pour 14,50 \$. Une carte des endroits de stationnement sur le campus est disponible sur <http://www.ssc.ca/fr/congrès/2015/transport>. Veuillez noter que le stationnement Dunn est réservé uniquement aux visiteurs hébergés sur le campus.

Campus Security • Sécurité sur le campus

- Fire/Police/Ambulance: 911
- Campus Police:
902-494-4109 (emergencies) or 4109 from any campus pay phone
902-494-6400 (general enquiries and parking questions)
- Halifax Police Service: 902-490-5020
- Pompiers/Police/Ambulance : 911
- Force policière du campus :
902-494-4109 (urgences) ou 4109 à partir de n'importe quel téléphone payant sur le campus
902-494-6400 (informations générales et stationnement)
- Force policière de Halifax : 902-490-5020

Internet Access • Accès internet

Wireless access is available for all attendees. A unique login and password for Wi-Fi access can be found at the back of your badge. The wireless network name is Dal-WPA2.

L'accès sans fil est disponible pour tous les participants. Un nom d'utilisateur et un mot de passe pour l'accès sans fil sont indiqués derrière votre porte-nom. Le nom du réseau sans fil est Dal-WPA2.

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 two locations: **Howe Hall** and **Risley Hall**. Each participant has been assigned to a particular location and must obtain their lunch at the location indicated on their badge. Individuals attending committee meetings should obtain their lunch from the McCain registration desk.

The **Student Union Building** is the location of the coffee breaks and has separate coffee and cafeteria locations. The **Killam Library** has an indoor atrium with tables, chairs, a Second Cup and Subway.

Most restaurants are located a short walking distance from the conference location. The Saeger Bistro is a short distance away and there are a number of Pub-style restaurants along nearby Spring Garden Road.

For dinner, there are a number of good restaurants in the downtown core, The Press Gang and The Wooden Monkey providing examples. For people headed to the waterfront, The Bicycle Thief, Ristorante a Mano and the Cut Steakhouse are good choices.

Les repas du midi de lundi à mercredi sont inclus avec l'inscription et servis à deux endroits : **Howe Hall** et **Risley Hall**. Chaque participant est assigné à une cafétéria indiquée sur son porte-nom. Les membres de comités participant à des réunions doivent aller chercher leur repas au kiosque d'inscription McCain.

Le **Student Union Building** est l'endroit pour des pauses-café et a des locaux séparés pour le café et les cafétérias. La **Killam Library** a un atrium avec des tables, des chaises, un café Second Cup et un restaurant Subway.

La plupart des restaurants sont à une courte distance à pied du congrès. Le Saeger Bistro est tout près et il y a de nombreux restaurants style pub sur le chemin Spring Garden avoisinant.

Pour le souper, il y a de nombreux bons restaurants au centre-ville, comme The Press Gang et The Wooden Monkey par exemple. Pour ceux allant vers le bord de l'eau, The Bicycle Thief, Ristorante a Mano et le Cut Steakhouse sont de bons choix.

Athletics Facilities • Installations sportives

Dalplex (<http://athletics.dal.ca/>) is the main fitness facility at Dalhousie. It has a gym, cardio room, pool, indoor running track and squash courts. Conference attendees staying on campus have regular membership privileges included as part of their accommodation. Day passes for other attendees are available for \$13.





Dalplex (<http://athletics.dal.ca/>) est le centre d'entraînement principal de Dalhousie. Il comprend un gymnase, une salle de cardio, une piscine, une piste d'athlétisme intérieure et des courts de squash. Les participants au congrès hébergés sur le campus reçoivent les privilèges de membres réguliers avec leur hébergement. Des passes de jour sont disponibles pour les autres congressistes au prix de 13 \$.





The Conference • Le congrès

Language • Langue

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

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

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

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

Rooms • Salles

Coffee breaks, exhibits and posters - McInnes Room, **Student Union Building**

Plenary sessions - Ondaatje Hall, **McCain**

Scientific sessions - **McCain, Rowe** or **Mona Campbell**

Pauses-café, exposants et affiches - salle McInnes, **Student Union Building**

Séances plénières - hall Ondaatje, **McCain**

Séances scientifiques - **McCain, Rowe** ou **Mona Campbell**

Poster Sessions • Séances d'affichage

Contributed posters and Case Study posters will be in the McInnes Room, with Case study posters being displayed Monday and contributed research posters appearing on Tuesday. Posters will be displayed from 12:00 to 5:30 pm. The authors will be with their posters from 1:30 pm until the end of coffee break at 3:30 pm.

Les séances d'affiches libres et d'études de cas seront dans la salle McInnes, avec les affiches des études de cas exposées le lundi et les affiches libres exposées le mardi. Les affiches seront exposées de midi à 17 h 30. Les auteurs seront avec leur affiche de 13 h 30 jusqu'à la fin de la pause-café à 15 h 30.

Workshops • Ateliers

Sunday, June 14, 9:00 - 4:00 pm • Dimanche 14 juin, 9 h - 16 h

Rowe and McCain • Rowe et McCain

Workshops organized by the sections will be held Sunday 9:00 am to 4:00 pm in the McCain (Survey Methods) and Rowe (all other workshops) buildings. Coffee breaks will be held in the McCain lobby at 10:20 am - 10:40 am and 2:50 pm - 3:10pm. Lunch break is 12 noon - 1:30 pm.

Les ateliers organisés par les groupes auront lieu le dimanche de 9 h à 16 h dans les édifices McCain (Méthodes d'enquête) et Rowe (tous les autres ateliers). Les pauses-café auront lieu dans le hall de McCain de 10 h 20 à 10 h 40 et de 14 h 50 à 15 h 10. Le lunch est de midi à 13 h 30.

Other Meetings • Autres réunions

SSC Annual General Meeting: Monday 5-6pm
McCain Scotia Bank Auditorium

Section AGM's: Tuesday 5-6pm
various rooms in McCain and Rowe

Committee meetings: Most noontime business meetings will occur in the McCain building, with lunch for business meeting participants available at either Risley or Howe Hall. Indicated on your badge is the location, Howe Hall or Risley Hall, where you are scheduled to pick up your lunch.

NSERC 2014 Results: Sunday 5-6
Rowe 1011

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

Assemblée générale annuelle : lundi 17 h-18 h, McCain Scotia Bank Auditorium

AGM des groupes : mardi 17 h-18 h, plusieurs salles à McCain et Rowe

Réunions de comités : La plupart des réunions du midi ont lieu dans l'édifice McCain, avec les repas pour les participants dans Risley Hall et Howe Hall. L'endroit désigné pour aller chercher votre repas, Howe Hall ou Risley Hall, est indiqué sur votre porte-nom.

Information concours 2014 CRSNG : dimanche 17 h-18 h Rowe 1011

Atelier subventions du CRSNG : mardi 17 h-18 h, McCain Scotia Bank Auditorium.
Cet atelier, présenté par le personnel des subventions à la découverte du CRSNG, couvrira le processus du formulaire 180, le processus d'évaluation des subventions à la découverte (principes, critères et cotes), le modèle de conférence et des conseils pour la préparation d'une demande de subvention à la découverte. À la fin de l'atelier, les participants pourront poser des questions.

Social Events • Activités sociales

Welcome Reception • Réception de bienvenue

Sunday, June 14, 6:00 - 8:30 pm • Dimanche 14 juin, 18 h 00 - 22 h 30McCain Lobby

The welcome reception will be held in the McCain lobby. La réception de bienvenue aura lieu dans le hall du McCain Building.

Barbeque • Barbecue

Monday, June 15, 6:30 - 8:30 pm • Lundi 15 juin, 18h30 - 20h30 Cameron Dining Room, Howe Hall

The student BBQ is free for students. The BBQ will take place in the Cameron Dining Room of Howe Hall. Non-student participants can join the event by purchasing a ticket (\$15).

Le barbecue étudiant est gratuit pour les étudiants. Il aura lieu au Cameron Dining Room de Howe Hall. Les participants qui ne sont pas des étudiants peuvent se joindre à l'événement en achetant un billet (15 \$)

Banquet

Tuesday, June 16, 6:15 pm • Mardi 16 juin, 18h15 Cunard Centre

The banquet will be held at the **Cunard Centre**, with cocktails and a cash bar at 6:15 pm and dinner service beginning at 7:05 pm. All conference participants who have not declined the banquet will find a banquet ticket in their registration envelope. You are asked to place this on the table in front of you. If you have a ticket but do not wish to attend the banquet, please return your ticket to the registration desk by noon on Tuesday, June 16. Buses are arranged to transport participants from Dalhousie University (in front of the McCain Building) and main hotels from 6:00 pm to 6:30 pm to the banquet location (last bus at 6:30 pm) and return participants after the banquet (10:00 pm to 10:30 pm). We recommend that all participants use the bus service, and we ask that you show your banquet ticket prior to boarding the bus. Parking is also available at the banquet site.

Le banquet aura lieu au **Centre Cunard**, avec cocktails et bar payant à 18 h 15 et souper servi à partir de 19 h 05. Les participants au congrès qui n'ont pas refusé le banquet trouveront un ticket dans leur enveloppe d'inscription. Vous devez placer ce ticket devant vous sur la table. Si vous avez un ticket mais que vous ne souhaitez pas participer au banquet, veuillez le rendre au bureau des inscriptions avant mardi le 16 juin à midi. Des autobus transporteront les participants de Dalhousie (devant l'édifice McCain) et des hôtels principaux entre 18 h et 18 h 30 vers le site du banquet (dernier bus à 18 h 30) et les ramèneront en fin de soirée (22 h à 22 h 30). Nous recommandons à tous les participants de se prévaloir de ces bus et vous demandons de montrer votre ticket de banquet en montant dans le bus. Un stationnement est également disponible sur site.

Other Social Events • Autres activités sociales

New investigators event Sunday following the welcome reception
Split Crow Pub Halifax, 1855 Granville Street

Événement des nouveaux chercheurs Dimanche après la réception de bienvenue
Split Crow Pub Halifax, 1855 Granville Street

COWIS dinner Sunday at 7:30pm
The Wooden Monkey, 1707 Grafton Street
reservations under the name `COWIS'

Souper du Comité de promotion de la femme en statistique
dimanche 19 h 30
The Wooden Monkey, 1707 Grafton Street
réservations sous le nom de `COWIS'

Committees and Meetings • Comités et réunions

Saturday June 13

samedi 13 juin

18:00 - 23:00 Executive Committee 1 / Comité exécutif 1
Lord Nelson Hotel - Vanguard 1

Sunday June 14

dimanche 14 juin

9:00 - 14:00 Board of Directors 1 / Conseil d'administration 1
McCain 2021

14:30 - 17:00 Fundraising Committee / Comité de collecte de fonds
McCain 2198

14:30 - 17:00 Website Redesign Committee / Comité de la restructuration du site Web
McCain 1198

14:30 - 17:00 Department Heads Meeting / Rencontre des directeurs de départements
McCain 2190

Monday June 15

lundi 15 juin

12:00 - 13:30 Business and Industrial Statistics Section Executive Committee 1 / Comité exécutif du groupe de statistique industrielle et de gestion 1
McCain 2184

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

12:00 - 13:30 Biostatistics Section Executive Committee 1 / Comité exécutif du groupe de biostatistique 1
McCain 2170

12:00 - 13:30 Finance Committee / Comité des finances
Rowe 1028

12:00 - 13:30 Probability Section Executive Committee 1 / Comité exécutif du groupe de probabilité 1
McCain 2198

12:00 - 13:30 Publications Committee / Comité des publications
McCain - Ondaatje

12:00 - 13:30 Research Committee / Comité de la recherche
McCain 2102

12:00 - 13:30 SSC Bylaws Committee / Comité de statuts de la SSC
Rowe 1011

12:00 - 13:30 Statistical Education Section Executive Committee 1 / Comité exécutif du groupe d'éducation en statistique 1
McCain 2132

12:00 - 18:00 Award for Case Studies in Data Analysis Committee 1 / Comité du prix pour les études de cas et l'analyse de données 1
McCain 2116

17:00 - 18:00 SSC Annual General Meeting / Assemblée générale annuelle de la SSC
McCain - Scotiabank

Tuesday June 16

mardi 16 juin

12:00 - 13:00 MacMillan Education Demo: Using LaunchPad in Introductory Statistics Courses for Online Homework and Assessment / À l'aide de LaunchPad pour un cours d'introduction aux statistiques aux fins d'évaluation et de travaux en ligne
Rowe 1020

12:00 - 13:30	CJS Editorial Board / Comité de rédaction de la RCS <i>McCain 2102</i>
12:00 - 13:30	Census at School Committee / Comité de recensement à l'école <i>McCain 2198</i>
12:00 - 13:30	Office Committee / Comité du Bureau <i>Rowe 1011</i>
12:00 - 13:30	Public Relations Committee / Comité des relations publiques <i>McCain - Ondaatje</i>
12:00 - 13:30	Statistical Education Committee / Comité d'éducation en statistique <i>McCain - Scotiabank</i>
12:00 - 13:30	Committee on New Investigators / Comité des nouveaux chercheurs <i>McCain 2132</i>
12:00 - 13:30	Accreditation General Meeting / Assemblée générale des membres accrédités <i>McCain 2170</i>
12:00 - 18:00	Award for Case Studies in Data Analysis Committee 2 / Comité du prix pour les études de cas et l'analyse de données 2 <i>McCain 2184</i>
17:00 - 18:00	Student Research Presentation Award Committee / Comité du prix pour la présentation de recherche étudiante <i>McCain 2118</i>
17:00 - 18:00	Actuarial Science Section General Meeting / Groupe d'actuariat assemblée générale <i>McCain 2170</i>
17:00 - 18:00	Biostatistics Section General Meeting / Groupe de biostatistique assemblée générale <i>Rowe 1011</i>
17:00 - 18:00	Probability Section General Meeting / Groupe de probabilité assemblée générale <i>McCain 2184</i>
17:00 - 18:00	Statistical Education Section General Meeting / Groupe d'éducation en statistique assemblée générale <i>McCain 2198</i>
17:00 - 18:00	Survey Methods Section General Meeting / Groupe de méthodes d'enquête assemblée générale <i>McCain 2132</i>

Wednesday June 17
mercredi 17 juin

12:00 - 13:30	Business and Industrial Statistics Section General Meeting / Groupe de statistique industrielle et de gestion assemblée générale <i>McCain 2198</i>
12:00 - 13:30	Accreditation Committee / Comité d'accréditation <i>McCain 2021</i>
12:00 - 13:30	Actuarial Science Section Executive Committee 2 / Comité exécutif du groupe d'actuariat 2 <i>McCain 2162</i>
12:00 - 13:30	Biostatistics Section Executive Committee 2 / Comité exécutif du groupe de biostatistique 2 <i>McCain 2176</i>
12:00 - 13:30	Probability Section Executive Committee 2 / Comité exécutif du groupe de probabilité 2 <i>McCain 2116</i>
12:00 - 13:30	Program Committee / Comité du programme <i>McCain - Scotiabank</i>
12:00 - 13:30	Statistical Education Section Executive Committee 2 / Comité exécutif du groupe d'éducation en statistique 2 <i>McCain 2102</i>

12:00 - 13:30	Student and Recent Graduate Committee / Comité des étudiants et diplômés récents <i>Rowe 1011</i>
12:00 - 13:30	Treasurers / Trésoriers <i>Rowe 1028</i>
16:45 - 18:00	Board of Directors 2 / Conseil d'administration 2 <i>McCain 2021</i>
18:00 - 19:00	Executive Committee 2 / Comité exécutif 2 <i>Risley Hall, Rooms 1 & 2</i>

Social and Information Events • Événements sociaux et informatifs

Sunday June 14

dimanche 14 juin

8:00 - 17:30	Registration / Inscription <i>McCain Lobby</i>
17:00 - 18:00	Information on 2015 NSERC Discovery Grant Competition / Information sur le concours 2015 des subventions à la découverte du CRSNG <i>Rowe 1011</i>
18:00 - 20:00	Welcome Reception / Réception de bienvenue <i>McCain Lobby</i>
19:30	COWIS dinner (open to all) / Souper du Comité sur les femmes en statistique (ouvert à tous) <i>The Wooden Monkey, 1707 Grafton Street, Halifax</i>
20:00	Informal Gathering of New Investigators / Rassemblement informel des nouveaux chercheurs <i>Split Crow Pub, 1855 Granville Street, Halifax</i>

Monday June 15

lundi 15 juin

7:30 - 17:00	Registration / Inscription <i>McCain Lobby</i>
9:50 - 10:30	Coffee Break / Pause-café <i>Student Union Building - McInness</i>
14:45 - 15:30	Coffee Break / Pause-café <i>Student Union Building - McInness</i>
17:00 - 19:00	Accredited Members Mixer / Réception pour les membres accrédités <i>Howe Hall - Cameron</i>
18:30 - 20:30	Student BBQ / Barbecue étudiant <i>Howe Hall - Cameron Dining Room</i>

Tuesday June 16

mardi 16 juin

8:15 - 17:00	Registration / Inscription <i>McCain Lobby</i>
9:50 - 10:30	Coffee Break / Pause-café <i>Student Union Building - McInness</i>
14:45 - 15:30	Coffee Break / Pause-café <i>Student Union Building - McInness</i>
17:00 - 18:00	NSERC Discovery Grant Application Workshop / Atelier sur les demandes de subventions à la découverte du CRSNG <i>McCain - Scotiabank</i>
18:15 - 19:05	Banquet- Cocktails and Cash Bar / Banquet- Bar payant <i>Cunard Centre</i>
19:05 - 22:00	Banquet- Dinner Service / Banquet- Service du repas <i>Cunard Centre</i>

Wednesday June 17

mercredi 17 juin

8:15 - 16:00	Registration / Inscription <i>McCain Lobby</i>
9:50 - 10:30	Coffee Break / Pause-café <i>Student Union Building - McInness</i>

14:30 - 15:00 Coffee Break / Pause-café
Student Union Building - McInness

List of Sessions • Liste des sessions

	Title	Page	Titre
1A-A1	Inaugural Session	22	Séance inaugurale
1A-A2	SSC Presidential Invited Address	22	Allocution de l'invité du président de la SSC
1B-I1	Combinatorial Probability	22	Probabilité combinatoire
1B-I2	Getting to Know our Students in the Electronic Age: Did the Dog Really Eat the Homework?	23	Apprendre à connaître nos étudiants à l'ère électronique : le chien a-t-il vraiment mangé leurs devoirs?
1B-I3	Recent Advances in Predictive Methods in Actuarial Science	23	Avancées récentes dans les modèles prédictifs en science actuarielle
1B-I4	Recursive Partitioning and Regression Trees	23	Partitionnement récursif et arbres de régression
1B-I5	Statistics in Biosciences: Statistical Methods for Big Data from Health Science	24	Statistique en biosciences : méthodes statistique pour les données volumineuses en sciences de la santé
1B-C1	Biostatistics: Genetics and Imaging Topics	24	Biostatistique : génétique et imagerie
1B-C2	Environmetrics	25	Environnemétrie
1B-C3	Overdispersion, Mixtures, and Related Topics	25	Surdispersion, mélanges et sujets connexes
1B-C4	Time Series and Probability	26	Séries chronologiques et probabilité
1C-D1	Case Study 1: Youth Employment and the 2008 Crisis	27	Étude de cas 1: Emploi des jeunes et la crise de 2008
1C-D2	Case Study 2: Baseball Strategies	28	Étude de cas 2: Stratégies au baseball
1D-I1	Advances in High-Dimensional Statistical Inference	28	Avancées en inférence statistique à dimensions élevées
1D-I2	Business and Industrial Applications	28	Applications industrielles et de gestion
1D-I3	Novel Statistical Methods for Patient-Reported Outcome Measures	29	Nouveaux modèles statistiques pour les mesures de résultats rapportées par les patients
1D-I4	Recent Developments in Small Area Estimation	29	Développements récents en estimation pour petites régions
1D-I5	BFF Inference and Beyond -- Building a Unified Foundation for Statistical Inference	29	Inférence BFF et au-delà -- les fondements d'une inférence statistique unifiée
1D-C1	Actuarial Science and Finance 1	30	Science actuarielle et finance 1
1D-C2	Biostatistics in Action 1	30	Biostatistique en action 1
1D-C3	Biostatistics: Dealing with Incomplete Data	31	Biostatistique : comment traiter les données incomplètes
1D-C4	Foundations and Nonparametrics	32	Fondements et statistique non paramétrique
1E-I1	Big Data Analysis: Methodology and Applications	32	Analyse des données volumineuses : méthodologie et applications
1E-I2	Methodological developments that could impact the practice of statistics	33	Avancées méthodologiques pouvant influencer la pratique de la statistique
1E-I3	Novel Assessment Methods in Undergraduate Education	33	Nouvelles méthodes d'évaluation dans le premier cycle d'études
1E-I4	Statistical Methods in Molecular Evolution	33	Méthodes statistiques en évolution moléculaire
1E-I5	Stochastic Processes and Applications	34	Processus stochastiques et applications
1E-C1	Complex Correlation Structures and Zero-Heavy Data	34	Structures de corrélation complexes et données à surreprésentation de zéros
1E-C2	Experimental Design, Simulation, and Reliability	35	Plan expérimental, simulation et fiabilité
1E-C3	Statistical Science for Ocean Ecosystems	35	Science statistique pour les écosystèmes océaniques
1E-C4	Statistics in Action: A Smörgåsbord	36	Statistique en action : un «pot-pourri »de présentations
2A-A	Gold Medal Address	38	Allocution du récipiendaire de la Médaille d'or

	Title	Page	Titre
2B-I1	Modelling of Temporally Correlated Multivariate Counts	38	Modélisation de décomptes multivariés corrélés temporellement
2B-I2	Recent Advances in Copula Dependence Modelling	38	Avancées récentes en modélisation de la dépendance au moyen de copules
2B-I3	Recent Advances in Infectious Disease Epidemiology	39	Avancées récentes et épidémiologie des maladies infectieuses
2B-I4	Statistical Inference in the Presence of Missing Survey Data	39	Inférence statistique en l'absence de certaines données de sondage
2B-I5	Uncertainty Quantification in Ocean and Environmental Models	39	Quantification de l'incertitude dans les modèles océanographiques et environnementaux
2B-C1	Biostatistics: Methodological Innovation 1	40	Biostatistique : innovation méthodologique 1
2B-C2	Methodological Advances: A Smörgåsbord	40	Progrès méthodologiques : un «pot-pourri »de présentations
2B-C3	Statistical Science for Genetics and Genomics	41	Science statistique pour la génétique et la génomique
2B-C4	Statistics Education	42	Éducation en statistique
2C-P	Poster Session	43	Séance d'affichage
2D-I1	Actuarial Graduate Students Presentations	45	Présentations des étudiants gradués en science actuarielle
2D-I2	Big Data and Neuroimaging	46	Données volumineuses et neuroimagerie
2D-I3	Recent Developments in Nonparametric Statistics	46	Développements récents en statistique non paramétrique
2D-I4	Showcasing New Investigating Women in Statistics	47	Nouvelles chercheuses en statistique
2D-I5	Survival Modelling with Environmental Applications	47	Modélisation de la survie et applications environnementales
2D-I6	Teaching Effective Communication of Statistical Ideas to Non-Statisticians	47	Comment enseigner à bien communiquer les idées statistiques aux non-statisticiens
2D-C1	Bayes and Likelihood: Models and Algorithms	48	Bayes et la vraisemblance : modèles et algorithmes
2D-C2	Robust Methods	48	Méthodes robustes
2D-C3	Theoretical Developments	49	Développements théoriques
2E-I1	Extreme Value Theory and Applications	49	Théorie des valeurs extrêmes et applications
2E-I3	Recent Developments in Statistical Genomic Modelling and Analysis	50	Développements récents en modélisation et analyse génomiques statistiques
2E-I4	Statistical Approaches for Big Data	50	Approches statistiques des données volumineuses
2E-I5	eLearning in Statistics -- Successes and Opportunities	51	Apprentissage électronique en statistique -- succès et opportunités
2E-C1	Actuarial Science and Finance 2	51	Science actuarielle et finance 2
2E-C2	Biostatistics: Methodological Innovation 2	52	Biostatistique : innovation méthodologique 2
2E-C3	Data Visualization and Multivariate Methods	52	Visualisation de données et méthodes multivariées
2E-C4	Survey Methodology	53	Méthodes d'enquêtes
3A-A1	CRM-SSC Prize in Statistics Address	55	Allocution du récipiendaire du Prix CRM-SSC en statistique
3A-A2	SSC Impact Award Address	55	Allocution du récipiendaire du Prix pour impact de la SSC
3B-I1	Actuarial, Financial and Quantitative Risk Management	55	Gestion du risque actuariel, financier et quantitatif
3B-I2	DASF 90 - A Celebration of the Contributions of Don Fraser, OC	56	DASF 90 -- Une célébration des contributions de Don Fraser, OC

	Title	Page	Titre
3B-I3	Statistical Methods for Genetic Dissection of Complex Traits	56	Méthodes statistiques pour la dissection génétique de caractères complexes
3B-I4	Total Survey Error	56	Erreur d'enquête totale
3B-I5	Using TMB to Quickly and Robustly Solve Problems from Marine Ecology	57	Comment utiliser TMB pour résoudre des problèmes en écologie marine de manière rapide et robuste
3B-C1	Big Data and Networks	57	Données volumineuses et réseaux
3B-C2	Biostatistics in Action 2	58	Biostatistique en action 2
3B-C3	Biostatistics: Causal Inference and Measurement Error	58	Biostatistique : inférence causale et erreur de mesure
3B-C4	Time Series and Longitudinal Analysis	59	Séries chronologiques et analyse longitudinale
3C-A1	Isobel Loutit Invited Address	61	Allocution invitée Isobel Loutit
3C-A2	Survey Methods Section Presidential Invited Address	61	Allocution de l'invité du président du Groupe des méthodes d'enquête
3C-A3	CJS Award Address	61	Allocution du récipiendaire du Prix de la RCS
3C-A4	Pierre Robillard Award Address	61	Allocution du récipiendaire du Prix Pierre-Robillard
3D-I1	Applications of Mixed-Effect Models in the Health Sciences	62	Applications des modèles à effets mixtes en sciences de la santé
3D-I2	Limit Theorems With Applications	62	Théorèmes limites et applications
3D-I3	Maritime Ecology	62	Écologie marine
3D-I4	R Packages for Teaching	63	Utiliser R pour l'enseignement
3D-I5	Statistical Application and Modeling for Functional Data Analysis	63	Application et modélisation statistiques pour l'analyse de données fonctionnelles

Program • Programme

Saturday June 13

samedi 13 juin

18:00 - 23:00

Lord Nelson Hotel - Vanguard 1

Executive Committee 1

Comité exécutif 1

Sunday June 14**dimanche 14 juin****8:00 - 17:30****McCain Lobby****Registration
Inscription****9:00 - 14:00****McCain 2021****Board of Directors 1
Conseil d'administration 1****9:00 - 16:00****Rowe 1014****GARCH Modeling in Financial Engineering
Modélisation GARCH en génie financier**Organizer and Chair / Responsable et président: Reg Kulperger (Western University)
Probability Section Workshop
Atelier du Groupe de probabilité

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

9:00 - 16:00**Rowe 1007****Identifying and Addressing Difficult Concepts for Students in an Introductory Statistics Course
Identification et explication de concepts difficiles pour les étudiants des cours d'introduction à la statistique**Organizer and Chair / Responsable et présidente: Bethany White (Western University)
Statistical Education Workshop
Atelier de formation statistique

9:00 - 16:00 Marjorie Bond (Monmouth College)

9:00 - 16:00**Rowe 1011****P propensity Scores: Methods, Models, and Adjustments
Scores de propension : méthodes, modèles et ajustements**Organizer and Chair / Responsable et président: Lehana Thabane (McMaster University)
Biostatistics Section Workshop
Atelier du Groupe de biostatistique

9:00 - 16:00 David Stephens (McGill University)

9:00 - 16:00**McCain 2184****Record Linkage Methods: Theory and Application
Méthodes de couplage d'enregistrements : théorie et application**Organizer and Chair / Responsable et président: Wesley Yung (Statistics Canada / Statistique Canada)
Survey Methods Section Workshop
Atelier du Groupe sur les méthodes d'enquête

9:00 - 16:00 Abdelnasser Saïdi (Statistics Canada / Statistique Canada)

14:30 - 17:00**McCain 2190****Department Heads Meeting
Rencontre des directeurs de départements**

14:30 - 17:00 **McCain 2198**
Fundraising Committee
Comité de collecte de fonds

14:30 - 17:00 **McCain 1198**
Website Redesign Committee
Comité de la restructuration du site Web

17:00 - 18:00 **Rowe 1011**
Information on 2015 NSERC Discovery Grant Competition
Information sur le concours 2015 des subventions à la découverte du CRSNG

18:00 - 20:00 **McCain Lobby**
Welcome Reception
Réception de bienvenue

19:30 **The Wooden Monkey, 1707 Grafton Street, Halifax**
COWIS dinner (open to all)
Souper du Comité sur les femmes en statistique (ouvert à tous)

20:00 **Split Crow Pub, 1855 Granville Street, Halifax**
Informal Gathering of New Investigators
Rassemblement informel des nouveaux chercheurs

Monday June 15**lundi 15 juin****7:30 - 17:00****McCain Lobby****Registration
Inscription****8:30 - 8:45 Session 1A-A1****McCain - Ondaatje****Inaugural Session
Séance inaugurale**

Special Session / Conférence spéciale

8:30 - 8:35 Ed Susko (Local Arrangements Chair)



8:35 - 8:45 John Petkau (SSC President)

8:45 - 9:50 Session 1A-A2 (abstract/résumé - page 65)**McCain - Ondaatje****SSC Presidential Invited Address
Allocution de l'invité du président de la SSC**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: John Petkau (University of British Columbia)

8:45 - 9:50 Xiao-Li Meng (Harvard University)

Lifelines for Statistics (and Statisticians): Generalists, Specialists, or Your-lists? / Bouées de sauvetage pour les statistiques (et les statisticiens) : généralistes, spécialistes ou vos-listes ?  **9:50 - 10:30****Student Union Building - McInness****Coffee Break
Pause-café****10:20 - 11:50 Session 1B-I1** (abstracts/résumés - page 66)**Rowe 1011****Combinatorial Probability
Probabilité combinatoire**



Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Neal Madras (York University)


Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité



10:20 - 10:50 Jane Gao (University of Waterloo), Mike Molloy (University of Toronto)

From Computer Science to Statistical Physics: Solution Clustering in Random CSPs / De l'informatique à la physique statistique : solution de regroupement en cas de problèmes de satisfaction des contraintes aléatoires  

10:50 - 11:20 Jeannette Janssen & Hoda Chuangpishit (Dalhousie University), Mahya Ghandehari (University of Delaware), Nauzer Kalyaniwalla (Dalhousie University)

Linear Embeddings of Large Graphs / Plongements linéaires de grands graphiques  

11:20 - 11:50 Lerna Pehlivan (University of Washington), Neal Madras (York University)

Random 312 Avoiding Permutations / Permutations évitant la séquence aléatoire 312  

10:20 - 11:50 **Session 1B-I2** (abstracts/résumés - page 68) **Rowe 1028**

Getting to Know our Students in the Electronic Age: Did the Dog Really Eat the Homework?

Apprendre à connaître nos étudiants à l'ère électronique : le chien a-t-il vraiment mangé leurs devoirs ?







Invited Paper Session / Présentations sur invitation

Chair/Président: Gary Sneddon (Mount Saint Vincent University)

Organizer/Responsable: Michele Millar (Mount Saint Vincent University)

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

- 10:20 - 10:45 Marjorie Bond (Monmouth College)
 A Qualitative Survey's Journey: From Conception to Publication / Le parcours d'un sondage
 qualitatif : de la conception à la publication  
- 10:45 - 11:10 Danielle Cox (Acadia University)
 The Inside Scoop on Students - Experiences in the Classroom / Ce qu'il faut savoir sur les
 étudiants - expériences dans la salle de classe  
- 11:10 - 11:35 Michelle Millar (Mount Saint Vincent University)
 Designing and Analyzing Student Surveys: Pitfalls and Rewards / Concevoir et analyser des
 enquêtes menées auprès d'étudiants : embûches et récompenses  
- 11:35 - 11:50 Discussion

10:20 - 11:50 **Session 1B-I3** (abstracts/résumés - page 70) **Mona Campbell 1108**

Recent Advances in Predictive Methods in Actuarial Science







Avancées récentes dans les modèles prédictifs en science actuarielle

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Manuel Morales (Université de Montréal)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe de science actuarielle

- 10:20 - 10:50 Cary Tsai (Simon Fraser University)
 A Linear Relational Approach to Modelling Mortality Rates / Approche relationnelle linéaire de
 la modélisation des taux de mortalité  
- 10:50 - 11:20 Maciej Augustyniak (Université de Montréal), Mathieu Boudreault (Université du Québec à Mon-
 tréal), Manuel Morales (Université de Montréal)
 Estimating the Markov-Switching GARCH Model with a Deterministic Particle Filter / Estima-
 tion du modèle GARCH à changement de régimes avec un filtre particulaire déterministe  
- 11:20 - 11:50 Anne Mackay (ETH Zurich), Mario Wüthrich (ETH Zürich)
 Predicting Best-Estimate Yield Curves in Incomplete Bond Markets / Meilleure évaluation des
 courbes de rendement sur les marchés obligataires incomplets  



10:20 - 11:50 **Session 1B-I4** (abstracts/résumés - page 72) **McCain - Scotiabank**





Recursive Partitioning and Regression Trees

Partitionnement récursif et arbres de régression

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Liqun Diao (University of Waterloo)

- 10:20 - 10:50 Michael LeBlanc (Fred Hutchinson Cancer Research Center)
 Refinement Rules for Prognosis and Enhanced Treatment Effect Subgroups / Règles de
 raffinement pour les pronostics et pour les effets de traitements améliorés dans des sous-
 groupes  

- 10:50 - 11:20 Robert Strawderman (University of Rochester), Jon Steingrimsson (Cornell University), Liqun Diao (University of Waterloo)
Doubly Robust Survival Trees / Arbres de survie doublement robustes  
- 11:20 - 11:50 Thomas Loughin & Andrew Henrey (Simon Fraser University)
Heteroscedastic Regression Trees for Joint Modeling of Means and Variances / Arbres de régression hétéroscédastique pour une modélisation conjointe des moyennes et des variances  







10:20 - 11:50 Session 1B-I5 (abstracts/résumés - page 74) **McCain - Ondaatje**

Statistics in Biosciences: Statistical Methods for Big Data from Health Science

Statistique en biosciences : méthodes statistique pour les données volumineuses en sciences de la santé

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Grace Yi (University of Waterloo)

- 10:20 - 10:50 Xihong Lin, Ian Barnett & Rajarshi Murkerjee (Harvard University)
The Generalized Higher Criticism for Testing SNP-set Effects in Genetic Association Studies / Haute critique généralisée pour tester des effets d'ensemble de SNP dans les études d'association génétique  
- 10:50 - 11:20 Hongzhe Li (University of Pennsylvania)
Sparse Simultaneous Signal Detection and Its Applications in Genomics / Détection de signaux rares simultanés et ses applications en génomique  
- 11:20 - 11:50 Charmaine Dean & Mark Wolters (Western University & Fudan University)
Parameter Estimation in Autologistic Regression Models for Detection of Smoke in Satellite Images / Estimation des paramètres dans les modèles autologistiques de régression pour la détection de fumée dans les images satellites  

10:20 - 11:50 Session 1B-C1 (abstracts/résumés - page 76) **McCain 2102**

Biostatistics: Genetics and Imaging Topics

Biostatistique : génétique et imagerie

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Zeny Feng (University of Guelph)

- 10:20 - 10:35 Yunan Zhu & Ivor Cribben (University of Alberta)
Estimating Sparse Graphical Models / Estimation des modèles graphiques épars  
- 10:35 - 10:50 Yongliang Zhai & Alexandre Bouchard-Côté (University of British Columbia)
A Poissonian Model of Indel Rate Variation for Phylogenetic Tree Inference / Modèle poissonien de la variation du taux d'indel pour l'inférence d'arbres phylogénétiques  
- 10:50 - 11:05 Sean Jewell & Alexandre Bouchard-Côté (University of British Columbia)
Phylogenetic Inference by Divide and Conquer Sequential Monte Carlo (D&C SMC) / Inférence phylogénétique par division et conquête séquentielle de Monte-Carlo (D&C SMC)  
- 11:05 - 11:20 Yun Cai, Hong Gu & Toby Kenney (Dalhousie University)
Supervised Non-Negative Matrix Factorization for Analysis of Microbiome Data / Factorisation de matrice non-négative supervisée pour l'analyse de données du microbiome  
- 11:20 - 11:35 Bo Chen (University of Toronto)
Statistical Analysis of Sex Chromosome in Genome-wide Association Studies / L'analyse statistique du chromosome sexuel dans les études d'association pangénomiques  
- 11:35 - 11:50 Keelin Greenlaw, Farouk Nathoo & Mary Lesperance (University of Victoria), Elena Szefer & Jinko Graham (Simon Fraser University)
Empirical Bayes Multivariate Group-Sparse Regression for Brain Imaging Genomics / Régression multivariée bayésienne empirique à groupes clairsemés en génomique de l'imagerie cérébrale  

10:20 - 11:50 Session 1B-C2 (abstracts/résumés - page 79) **McCain 2170**

**Environmetrics
Environnementrie**

Contributed Paper Session / Séance de communications libres

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

- 10:20 - 10:35 David J Thomson, David L Riegert & Aaron Springford (Queen's University)
Experiments ``Fixing" Ground-Level Ozone / Expériences sur le rétablissement de l'ozone au sol E E
- 10:35 - 10:50 Herbert Nkwimi Tchahou, Claude Girard & Martin Hamel (Statistique Canada)
Indicators of Database Quality for Wetlands in Canada: Example of Analysis of Environmental Data / Indicateurs de qualité des bases de données des terres humides au Canada : un exemple d'analyse de données environnementales F E
- 10:50 - 11:05 Ofir Harari (Simon Fraser University), Matthew T. Pratola (Ohio State University), Derek Bingham & Gwenn E. Flowers (Simon Fraser University)
Design on Non-Convex Regions: Optimal Experiments for Spatial Process Prediction / Plan d'expérience sur des régions non convexes : expériences optimales pour la prévision de processus spatiaux E E
- 11:05 - 11:20 Audrey-Anne Vallée (Université de Neuchâtel), Bastien Ferland-Raymond (Ministère des forêts, de la faune et des parcs du Québec), Louis-Paul Rivest (Université Laval), Yves Tillé (Université de Neuchâtel)
Incorporating Spatial and Operational Constraints in the Sampling Designs for Forest Inventories / Incorporation de contraintes spatiales et opérationnelles dans des plans d'échantillonnage pour des inventaires forestiers E E
- 11:20 - 11:35 Camila Casquilho-Resende (University of British Columbia), Nhu D. Le (BC Cancer Agency), James V. Zidek (University of British Columbia)
Spatio-temporal Modelling of Temperature Fields in the Pacific Northwest / Modélisation spatio-temporelle des champs de température dans le Nord-Ouest Pacifique E E
- 11:35 - 11:50 Peter Starszyk & Martin Lysy (University of Waterloo)
Inferring Chemical Kinetics from Direct and Indirect Concentration Data / Inférence en cinétique chimique à partir de données de concentration directes et indirectes E E





10:20 - 11:35 Session 1B-C3 (abstracts/résumés - page 82) **McCain 2184**

**Overdispersion, Mixtures, and Related Topics
Surdispersion, mélanges et sujets connexes**

Contributed Paper Session / Séance de communications libres

Chair/Président: Alexandre Leblanc (University of Manitoba)

- 10:20 - 10:35 Xingde Duan & Renjun Ma (University of New Brunswick)
``Multivariate Poisson mixed models with unstructured covariance structure" / Modèles de Poisson mixtes multivariés avec une structure de covariance non structurée E E
- 10:35 - 10:50 Mohammad Rajibul Islam Mian & Sudhir Paul (University of Windsor)
Estimation for Zero Inflated Overdispersed Generalized Linear Model with Missing Covariates as well as Response / Estimation pour les modèles linéaires généralisés surdispersés à surreprésentation de zéros avec covariables et réponses manquantes E E
- 10:50 - 11:05 William H. Aeberhard (Dalhousie University), Eva Cantoni (University of Geneva), Stephane Heritier (Monash University)
Nonparametric Estimation of Mixing Distributions for Mixed Poisson Models / Estimation non paramétrique des lois mélanges pour les mélanges de Poisson E E

- 11:05 - 11:20 Christine Nguyen, Justyne Mottola & Mei Ling Huang (Brock University)
On Inference of a Mixture Pareto Distribution / Sur l'inférence d'une distribution Pareto de mélanges  
- 11:20 - 11:35 Mohamed Amezziane (Central Michigan University)
The Exact Distribution of a Certain Class of Nonparametric Multivariate Two-Sample Test Statistics / Distribution exacte d'une certaine classe de statistiques de test multivariées non paramétriques à deux échantillons  

10:20 - 11:50 Session 1B-C4 (abstracts/résumés - page 85) McCain 2198

Time Series and Probability
Séries chronologiques et probabilité

Contributed Paper Session / Séance de communications libres
Chair/Président: Alexandre Bouchard-Côté (University of British Columbia)

- 10:20 - 10:35 Abdollah Safari & Rachel MacKay Altman (Simon Fraser University), Brian Leroux (University of Washington)
Estimators of Regression Coefficients in Models for Time Series of Counts / Estimateurs des coefficients de régression dans les modèles de séries chronologiques de dénombrement  
- 10:35 - 10:50 Jinjun Tong & J Concepción Loredó-Osti (Memorial University of Newfoundland)
Estimation and Filtering of State Space Model with GEV marginals and Exponential Levy-type Stable Errors / Estimation et filtrage d'un modèle d'espace d'états au moyen de marges de valeur extrême généralisée (GEV) et d'erreurs exponentielles stables de type Lévy  
- 10:50 - 11:05 Trevor Thomson, Melody Ghahramani & Shakhawat Hossain (University of Winnipeg)
Efficient Estimation for Time Series Following Generalized Linear Models / Estimation efficace des séries chronologiques qui suivent des modèles linéaires généralisés  
- 11:05 - 11:20 Haosui Duanmu, William Weiss & Jeffrey Rosenthal (University of Toronto)
Nonstandard Analysis and its Application to Markov Chain Convergence Problem / Analyse non standard appliquée au problème de convergence des chaînes de Markov  
- 11:20 - 11:35 Jinyoung Yang & Jeffrey Rosenthal (University of Toronto)
Automatically Tuned General-Purpose MCMC via New Adaptive Diagnostics / Méthodes MCMC d'usage général auto-réglées via de nouveaux diagnostics adaptatifs  
- 11:35 - 11:50 Sudhakar Achath & Amrita Vishwa Vidyapeetham (University of Saskatchewan)
Imprecise Probabilities for Measuring Tracking Error Risk / Probabilités imprécises pour mesurer les risques d'erreur de suivi  

12:00 - 13:30 McCain - Scotiabank

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

12:00 - 18:00 McCain 2116

Award for Case Studies in Data Analysis Committee 1
Comité du prix pour les études de cas et l'analyse de données 1

12:00 - 13:30 McCain 2170

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

12:00 - 13:30 **McCain 2184**
Business and Industrial Statistics Section Executive Committee 1
Comité exécutif du groupe de statistique industrielle et de gestion 1

12:00 - 13:30 **Rowe 1028**
Finance Committee
Comité des finances

12:00 - 13:30 **McCain 2198**
Probability Section Executive Committee 1
Comité exécutif du groupe de probabilité 1

12:00 - 13:30 **McCain - Ondaatje**
Publications Committee
Comité des publications

12:00 - 13:30 **McCain 2102**
Research Committee
Comité de la recherche

12:00 - 13:30 **Rowe 1011**
SSC Bylaws Committee
Comité de statuts de la SSC

12:00 - 13:30 **McCain 2132**
Statistical Education Section Executive Committee 1
Comité exécutif du groupe d'éducation en statistique 1

12:00 - 17:30 **Session 1C-D1** **Student Union Building - McInness**
Case Study 1: Youth Employment and the 2008 Crisis
Étude de cas 1 : Emploi des jeunes et la crise de 2008

Case Study / Étude de cas

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

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)

(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

12:00 - 17:30 Asma Bahamyirou, Alexei Nordell-Markovits, Eric Tsala & Teng Ke Xiong (Université de Sherbrooke)

12:00 - 17:30 Menglu Che, Sile Tao & Wei Tu (University of Alberta)

12:00 - 17:30 Bo Chen, Reihaneh Entezari & Tadeu Ferreira (University of Toronto)

12:00 - 17:30 Philippe Gagnon, Antoine Baudouin, Félix Carrier & Rafael Sobrino (HEC Montréal)

12:00 - 17:30 Bowen Gao, Chendi Liao, Genming He, Jia Jing, Megan Kan & Qian Yang (University of Toronto)

12:00 - 17:30 Kamal Rai, Laura Lin, Joanne Xu, Danny Lu & Zixin Nie (University of Waterloo)

12:00 - 17:30 Gabrielle Simoneau & Pablo Gonzalez Ginestet (McGill University)

12:00 - 17:30 Abhishek Varu, Pratheep Kunwar, Yuvraj Pannu, Sudipta Saha, Bowei Xiao, Hiroaki Iijima, Janarthan Baral, Mohsen Soltanifar & Maheshan Indralingam (University of Toronto)

12:00 - 17:30 Session 1C-D2 Student Union Building - McInness
Case Study 2: Baseball Strategies**Étude de cas 2 : Stratégies au baseball**

Case Study / Étude de cas

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

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)







(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

- 12:00 - 17:30 Junior Chuang, David Dai, Katherine Daignault, Bingqing Shen, Kelly Sunderland, Mai Xu & Jie Zhu (University of Toronto)
- 12:00 - 17:30 Chendi Liao, Genming He, Jia Jing & Wenqi Fan (University of Toronto)
- 12:00 - 17:30 Sergio Vicente, Isabelle Lefebvre & Steve Ferreira (Université de Montréal)

13:30 - 15:00 Session 1D-I1 (abstracts/résumés - page 88) Rowe 1028
Advances in High-Dimensional Statistical Inference**Avancées en inférence statistique à dimensions élevées**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Yingli Qin (University of Waterloo)

- 13:30 - 14:00 Gourab Mukherjee (University of Southern California), Lawrence D. Brown (University of Pennsylvania), Paat Rusmevichientong (University of Southern California)
Adaptive Inventory Management Strategies for a Large Scale Newsvendor / Stratégies adaptatives de gestion des stocks pour une grande échelle de marchands de journaux  
- 14:00 - 14:30 Zhigang Yao (National University of Singapore), Shiqiong Huang & Jiashun Jin (Carnegie Mellon University)
Partial Correlation Screening for Estimating Large Precision Matrices, with Applications to Classification / Filtrage par corrélation partielle afin d'estimer des matrices de grande précision, avec applications à la classification  
- 14:30 - 15:00 Abbas Khalili (McGill University), Anand Vidyashankar (George Mason University)
Post-Model Selection Inference for Finite Mixture of Regression (FMR) Models / Inférence post-sélection de modèles pour les modèles de mélange fini de régression  





13:30 - 15:00 Session 1D-I2 (abstracts/résumés - page 90) McCain - Scotiabank
Business and Industrial Applications**Applications industrielles et de gestion**

Invited Paper Session / Présentations sur invitation

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

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 13:30 - 14:15 Shirley Mills & Dan Wu (Carleton University)
Privacy Preserving Classification for Partitioned Data / Protéger la confidentialité de la classification des données partitionnées  
- 14:15 - 15:00 Stefan Steiner, Tricia Cooper-Barfoot & Jock MacKay (University of Waterloo)
Regulating the Bias-Variance Tradeoff in Analysis of Survey Data over Time with Application to a Customer Loyalty Survey / Réguler le compromis biais-variance dans l'analyse des données d'enquête au fil du temps avec application à une enquête de fidélisation de la clientèle  
-

13:30 - 15:00 **Session 1D-I3** (abstracts/résumés - page 92) **McCain - Ondaatje**

Novel Statistical Methods for Patient-Reported Outcome Measures







Nouveaux modèles statistiques pour les mesures de résultats rapportées par les patients

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Lisa Lix (University of Manitoba)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 13:30 - 14:00 Myriam Blanchin, Alice Guilleux, Jean-Benoit Hardouin & Véronique Sébille (University of Nantes)
Assessing Change in Longitudinal PROMs Studies: Using IRT Polytomous Models to Detect and Adjust for Response Shift / Mesurer l'évolution dans les études longitudinales utilisant des PROMs : une méthode basée sur les modèles IRT polytomiques pour la détection et la prise en compte du décalage de la réponse  
- 14:00 - 14:30 Eleanor Pullenayegum (University of Toronto)
Handling Health Utility Data: the Example of the EQ-5D / Traitement des données sur l'état de santé : l'exemple de l'EQ-5D  
- 14:30 - 15:00 Tolulope Sajobi & Yukun Zhang (University of Calgary), Lisa Lix (University of Manitoba)
On Discrimination in Multivariate Non-Normal Repeated Measures Data / Discrimination dans des données multivariées non normales de mesures répétées  

13:30 - 15:00 **Session 1D-I4** (abstracts/résumés - page 94) **McCain 2170**

Recent Developments in Small Area Estimation

Développements récents en estimation pour petites régions

Invited Paper Session / Présentations sur invitation

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

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 13:30 - 14:00 Malay Ghosh (University of Florida), Tatsuya Kubokawa & Shonosuke Sugawara (University of Tokyo), Sanjay Chaudhuri (National University of Singapore)
Prediction in Heteroscedastic Nested Error Regression Models with Random Dispersions / Prédiction dans les modèles de régression à erreurs emboîtées hétéroscédastiques avec dispersions aléatoires  
- 14:00 - 14:30 Jiming Jiang & Rohosen Bandyopadhyay (University of California, Davis)
Benchmarking the Observed Best Predictor / Étalonnage du meilleur prédicteur observé  
- 14:30 - 15:00 Gauri Datta (University of Georgia), Adrijo Chakraborty (National Opinion Research Center), Abhyuday Mandal (University of Georgia)
Small Area Estimation for Normal Mixture Models / Estimation pour les petits domaines dans les modèles de mélanges normaux  

13:30 - 15:00 **Session 1D-I5** (abstracts/résumés - page 96) **Rowe 1011**

BFF Inference and Beyond – Building a Unified Foundation for Statistical Inference



Inférence BFF et au-delà – les fondements d'une inférence statistique unifiée





Chair/Président: Xiao-Li Meng (Harvard University)

Organizers/Responsables: Xiao-Li Meng (Harvard University) and/et Min-ge Xie (Rutgers University)

Session sponsored by the ICOSA – Canada Chapter

Séance parrainée par le chapitre canadien de l'ICOSA

- 13:30 - 14:00 Nancy Reid (University of Toronto), David Cox (University of Oxford)
On some Principles of Statistical Inference / Sur certains principes de l'inférence statistique  

- 14:00 - 14:30 Jenny Shi & Jan Hannig (University of North Carolina at Chapel Hill), Thomas Lee & Chu Shing Lai (University of California at Davis)
A fiducial approach to the estimation of covariance matrix / Une approche fiduciaire à l'estimation de la matrice de covariance  
- 14:30 - 15:00 Ming-ge Xie (Rutgers University)
Confidence Distribution (CD) and a Unifying Framework for Statistical Inferences / Distribution de confiance (DC) et un cadre unifié pour les inférences statistiques  

13:30 - 14:45 Session 1D-C1 (abstracts/résumés - page 98) **McCain 2184**

Actuarial Science and Finance 1

Science actuarielle et finance 1

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Michelle Xia (Northern Illinois University)

- 13:30 - 13:45 Toby Kenney (Dalhousie University)
Credibility Classification with Missing Data / Classification basée sur la crédibilité avec des données manquantes  
- 13:45 - 14:00 Etienne Marceau, Hélène Cossette, Itre Mtalai & Déry Veilleux (Université Laval)
Discrete-Time Risk Models, with Exchangeable Archimedean Copulas / Modèles de risque en temps discret avec copules archimédiennes échangeables  
- 14:00 - 14:15 Alexander Melnikov (University of Alberta)
Financial Markets with Long-range Dependence: Partial Hedging and Statistical Study / Marchés financiers avec dépendance à long terme : couverture de risque partielle et étude statistique  
- 14:15 - 14:30 Lealand Morin (Capital One)
Keeping Diffusion Processes within Bounds: Using Information between Observations / Maintenir des processus de diffusion à l'intérieur des limites : utiliser l'information entre les observations  
- 14:30 - 14:45 David Saunders (University of Waterloo), Lung Kwan Tsui (HSBC), Satish Iyengar (University of Pittsburgh)
Lower Tail Independence of Hitting Times of Two-Dimensional Diffusion Processes / Indépendance de l'aile inférieure des temps de frappe de processus de diffusion a deux dimensions  





13:30 - 15:00 Session 1D-C2 (abstracts/résumés - page 101) **McCain 2198**









Biostatistics in Action 1

Biostatistique en action 1

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Celia Greenwood (McGill University)

- 13:30 - 13:45 Junior Chuang, Wei Hao (David) Dai & Wendy Lou (University of Toronto), Kelly Sabaliauskas (Southern Ontario Centre for Atmospheric Aerosol Research, University of Toronto)
Impact of Ultra-Fine Particles on Childhood Development of Lung Functions, Results from Canadian Healthy Infant Longitudinal Development (CHILD) Study / Impact des particules ultrafines sur le développement infantile des fonctions pulmonaires, résultats de l'Étude longitudinale nationale sur le développement des enfants en santé (CHILD)  
- 13:45 - 14:00 Mohammad Ehsanul Karim, Paul Gustafson, John Petkau, Helen Tremlett & The BeAMS Study Group (University of British Columbia)
Comparison of Statistical Approaches Dealing with Immortal Time Bias in Drug Effectiveness Studies / Comparaison d'approches statistiques qui traitent de biais d'immortalité dans les études d'efficacité de médicaments  

- 14:00 - 14:15 Alisha Albert-Green (University of Western Ontario), John Braun (University of British Columbia Okanagan), Charmaine Dean (University of Western Ontario)
A General Framework for the Joint Modelling of Multivariate Zero-Inflated Spatial Processes / Un cadre général pour la modélisation conjointe des processus spatiaux multivariés à surreprésentation de zéros  
- 14:15 - 14:30 Wei Hao (David) Dai (University of Toronto), Zihang Lu (The Hospital For Sick Children), Wendy Lou (University of Toronto)
Composite Indices for High-Dimensional Questionnaire Data / Indices composites pour des données de questionnaire de haute dimension  
- 14:30 - 14:45 Xuan Li, Hanna Jankowski, Xiaogang Wang & Jane Heffernan (York University)
A Method for Clustering Hemagglutinin Influenza Protein Sequences / Une méthode pour regrouper les séquences de protéines hémagglutinines de la grippe  
- 14:45 - 15:00 Ahalya Sivathayalan & Patrick Farrell (Carleton University), Daniel Krewski (University of Ottawa)
Empirical Comparisons of the Signal-to-Noise Crossover Dose and the Benchmark Dose / Comparaisons empiriques de dose de croisement signal-bruit et la dose de référence  













13:30 - 15:00 **Session 1D-C3** (abstracts/résumés - page 104) **McCain 2116**

Biostatistics: Dealing with Incomplete Data

Biostatistique : comment traiter les données incomplètes

Contributed Paper Session / Séance de communications libres

Chair/Président: Joel Dubin (University of Waterloo)

- 13:30 - 13:45 Liqun Diao, Richard Cook & Leilei Zeng (University of Waterloo)
Vine Copula Models for Recurrent Exacerbations in Chronic Disease / Modèles de copules en vignes pour les exacerbations récurrentes dans les maladies chroniques  
- 13:45 - 14:00 Yayuan Zhu, Cecilia Cotton & Jerry Lawless (University of Waterloo)
Analysis of Complex Event Outcome Data from Disease Clinics / Analyse des données de résultat d'un événement complexe pour des maladies cliniques  
- 14:00 - 14:15 Michael McIsaac (Queen's University)
Designing Studies with Incomplete Data: An Adaptive Approach to Optimal Response-Dependent Mean Score Sampling / Concevoir des études avec des données incomplètes : une approche adaptative au score moyen d'échantillonnage dépendant de la réponse optimale  
- 14:15 - 14:30 Chithran Vasudevan (Memorial University), Asokan Mulayath Variyath & Zhao Zhi Fan (Memorial University)
Censored Quantile Regression with the Effective Usage of Auxiliary Information / Régression quantile censurée avec l'utilisation efficace de l'information auxiliaire  
- 14:30 - 14:45 Ying Wu & Richard Cook (University of Waterloo)
Assessing the Accuracy of Predictive Models with Interval-censored Data / Évaluation de la précision des modèles prédictifs sur des données censurées par intervalle  
- 14:45 - 15:00 Christopher A. Gravel & Patrick Farrell (Carleton University), Daniel Krewski (University of Ottawa), Anup Dewanji (Indian Statistical Institute)
A Validation Sampling Approach for Unbiased Estimation in Misclassified Right Censored Continuous Time Survival Data / Méthode de validation par échantillonnage pour estimation non biaisée de données de survie continues dans le temps censurées à droite mal classifiées  

13:30 - 15:00 **Session 1D-C4** (abstracts/résumés - page 107) **McCain 2102**

Foundations and Nonparametrics

Fondements et statistique non paramétrique

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Mireille Schnitzer (Université de Montréal)

- 13:30 - 13:45 Asma Bahamyirou & Éric Marchand (Université de Sherbrooke)
On the Discrepancy between Bayes Credibility and Frequentist Probability of Coverage / Sur le décalage entre la crédibilité bayésienne et la probabilité de couverture fréquentiste.  
- 13:45 - 14:00 Weili Li (Dalla Lana School of Public Health, University of Toronto), Lisa Strug (The Hospital for Sick Children)
Extending the Evidential Paradigm to Accommodate Composite Hypotheses / L'extension du paradigme de la preuve pour tenir compte des hypothèses composites  
- 14:00 - 14:15 Félix Camirand Lemyre & Taoufik Bouezmarni (Université de Sherbrooke), Jean-François Quesy (Université du Québec à Trois-Rivières)
Conditional Distributions from Top to Bottom. / Les distributions conditionnelles de long en large.  
- 14:15 - 14:30 Mikelis Bickis (University of Saskatchewan)
Overthrowing the Ω Delusion and Making Probability Pointless / Bouleverser le délire d' Ω par la probabilité sans points  
- 14:30 - 14:45 Alexandre Leblanc (University of Manitoba)
Local Polynomial Regression: A Bernstein Polynomial Perspective / Régression polynomiale locale : une approche basée sur les polynômes de Bernstein  
- 14:45 - 15:00 Mohamed Belalia & Taoufik Bouezmarni (Université de Sherbrooke), Alexandre Leblanc (University of Manitoba)
Bernstein Conditional Density Estimation with Application to Conditional Distribution and Regression Functions / Estimation de la densité conditionnelle à l'aide des polynômes de Bernstein avec application aux lois conditionnelles et aux fonctions de régression  

14:45 - 15:30

Coffee Break

Pause-café

Student Union Building - McInness







15:30 - 17:00 **Session 1E-I1** (abstracts/résumés - page 110) **McCain - Scotiabank**

Big Data Analysis: Methodology and Applications

Analyse des données volumineuses : méthodologie et applications

Invited Paper Session / Présentations sur invitation





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

- 15:30 - 16:00 Hao Chai & Ma Shuangge (Yale University)
Identifying Gene-Environment Interactions using a Robust Penalization Approach / Identification des interactions gène-environnement à l'aide d'une approche de pénalisation robuste  
- 16:00 - 16:30 Farouk Nathoo (University of Victoria)
A High-Dimensional State-Space Model for the Joint Analysis of EEG and MEG Data / Modèle spatiotemporel de grande dimension pour l'analyse commune de données EEG et MEG  
- 16:30 - 17:00 Anand Vidyashankar (George Mason University), Ahmed Ejaz (Brock University)
Robust Inference for Conditionally Specified Models via Divergences / Inférence robuste pour les modèles à spécifications conditionnelles via la divergence  

15:30 - 17:00 **Session 1E-I2** (abstracts/résumés - page 112) **McCain - Ondaatje**

Methodological developments that could impact the practice of statistics
Avancées méthodologiques pouvant influencer la pratique de la statistique







Invited Paper Session / Présentations sur invitation
 Organizer and Chair / Responsable et président: Fernando Camacho (Damos Inc.)
 Session sponsored by the Accreditation Committee
 Séance parrainée par le Comité d'accréditation

- 15:30 - 16:15 Richard Cook (University of Waterloo)
 Adaptive Clinical Trials: Statistical Methods and Practical Considerations. / Essais cliniques adaptatifs : méthodes statistiques et considérations pratiques.  
- 16:15 - 17:00 Lehana Thabane (McMaster University)
 Issues in Vaccine Trials: Recent Debate, Methodological Challenges and the Ethics / Enjeux dans les essais de vaccin : débat récent, défis méthodologiques et éthique  

15:30 - 17:00 **Session 1E-I3** (abstracts/résumés - page 113) **Rowe 1011**

Novel Assessment Methods in Undergraduate Education
Nouvelles méthodes d'évaluation dans le premier cycle d'études




Invited Paper Session / Présentations sur invitation
 Chair/Présidente: Alison Gibbs (University of Toronto)
 Organizer/Responsable: Bruce Dunham (University of British Columbia)
 Session sponsored by the Statistics Education Section
 Séance parrainée par le Groupe d'éducation en statistique



- 15:30 - 16:00 Susan Kinniburgh, Bree Wilton & Susan Chen (Camosun College)
 Outcomes of an Introductory Statistics Course Using WeBWork Online Assessment / Résultats d'un cours d'introduction à la statistique utilisant l'évaluation en ligne WeBWork  
- 16:00 - 16:30 Bruce Dunham (University of British Columbia)
 Using Two-Stage Examinations in Undergraduate Courses / Utilisation d'examens à deux degrés dans les cours de premier cycle  
- 16:30 - 17:00 Douglas Whitaker & Tim Jacobbe (University of Florida)
 LOCUS: Assessing Conceptual Understanding of Statistics / LOCUS : Évaluation de la compréhension conceptuelle de la statistique  

15:30 - 17:00 **Session 1E-I4** (abstracts/résumés - page 115) **Rowe 1028**

Statistical Methods in Molecular Evolution
Méthodes statistiques en évolution moléculaire

Invited Paper Session / Présentations sur invitation
 Organizer and Chair / Responsable et présidente: Liangliang Wang (Simon Fraser University)

- 15:30 - 16:00 Cécile Ané (University of Wisconsin-Madison), Claudia Solís-Lemus (University of Wisconsin - Madison)
 Statistical Methods to Detect Reticulate Evolution / Méthodes statistiques pour réseaux phylogénétiques avec hybridations  
- 16:00 - 16:30 Alexandre Bouchard-Côté (University of British Columbia)
 Inferring the Phylogeny of Cancer Cells / La reconstruction de la phylogénèse de cellules cancéreuses  

- 16:30 - 17:00 Forrest Crawford (Yale University), Marc Suchard (University of California, Los Angeles), Vladimir Minin (University of Washington)
Estimation for General Birth-Death Processes in Molecular Evolution / Estimation des processus généraux de naissance et de mort en évolution moléculaire  

15:30 - 17:00 Session 1E-I5 (abstracts/résumés - page 117) **McCain 2170**







Stochastic Processes and Applications
Processus stochastiques et applications

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Gail Ivanoff (University of Ottawa)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité









- 15:30 - 16:00 Gennady Shaikhet & Evangelos Kranakis (Carleton University)
Asymptotically Tight Bounds for Sensor Allocation Problems on the Real Line / Bornes asymptotiquement étroites pour les problèmes de répartition de capteurs sur la droite des réels  
- 16:00 - 16:30 Aaron Smith (University of Ottawa), Natesh Pillai (Harvard University)
Mixing Time of the Constrained Ising Process on the Lattice at Low Density / Temps de mélange du processus d'Ising sur le treillis à faible densité  
- 16:30 - 17:00 Louis-Pierre Arguin (Université de Montréal), David Belius (NYU), Adam Harper (Cambridge)
Probabilistic Approach for the Maxima of the Riemann Zeta Function on the Critical Line / Approche probabiliste pour le maximum de la fonction zeta de Riemann sur la droite critique  





15:30 - 17:00 Session 1E-C1 (abstracts/résumés - page 119) **McCain 2198**

Complex Correlation Structures and Zero-Heavy Data
Structures de corrélation complexes et données à surreprésentation de zéros

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Ying Zhang (Acadia University)

- 15:30 - 15:45 Tiago Pellegrini, Tariqul Hasan & Renjun Ma (University of New Brunswick)
Tweedie Mixed Models for Spatial Areal Data / Modèles mixtes de Tweedie pour des données géospatiales  
- 15:45 - 16:00 Shijia Wang & Noel Cadigan (Memorial University of Newfoundland), Hugues Benoit (Fisheries and Oceans Canada)
Inference About Regression Parameters Using Highly Stratified Survey Count Data with Overdispersion and Repeated Measurements / Inférence sur les paramètres de régression par l'utilisation de données de dénombrement de sondages hautement stratifiées présentant une surdispersion et des mesures répétées  
- 16:00 - 16:15 Tariqul Hasan (University of New Brunswick), Gary Sneddon (Mount St. Vincent University), Renjun Ma (University of New Brunswick)
Simultaneous Modelling of Clustered Marginal Counts and Multinomial Proportions with Zero-Inflation with Application to Analysis of Osteoporotic Fractures Data / Modélisation simultanée de données de dénombrement marginal en grappes et de proportions multinomiales avec surreprésentation de zéros appliquée à l'analyse de données de fractures ostéoporotiques  
- 16:15 - 16:30 Chunlin Wang, Pengfei Li & Paul Marriott (University of Waterloo)
Semiparametric Inferences on the Means of Groups with Excess Zero Observations / Les inférences semi-paramétriques sur les moyennes de groupes avec des observations à surreprésentation de zéros  

- 16:30 - 16:45 Erin Lundy & Charmaine Dean (Western University), Elizabeth Juarez-Colunga (University of Colorado Denver)
Joint Models for Multivariate Longitudinal Zero-Heavy Count Data Analysis / Modèles conjoints pour l'analyse de données de dénombrement multivariées longitudinales à surreprésentation de zéros  
- 16:45 - 17:00 Haocheng Li (University of Calgary), Raymond Carroll (Texas A&M University)
Joint Modeling Methods for Functional Data Mixed with Zero-and-One Inflated Continuous Proportions and Skewed Continuous Measurements / Méthodes de modélisation conjointe pour des données fonctionnelles mélangées avec des proportions continues à surreprésentation de zéros et de uns et des mesures continues asymétriques  













15:30 - 17:00 **Session 1E-C2** (abstracts/résumés - page 123) **McCain 2116**

Experimental Design, Simulation, and Reliability

Plan expérimental, simulation et fiabilité

Contributed Paper Session / Séance de communications libres

Chair/Président: Stefan Steiner (University of Waterloo)

- 15:30 - 15:45 Hugh Chipman & Pritam Ranjan (Acadia University)
Simulation Studies for Statistical Procedures: Why Can't We Practice What We Preach? / Études de simulation pour des procédures statistiques : pourquoi ne pouvons-nous pas faire ce que nous prêchons ?  
- 15:45 - 16:00 Fatimah Al Ahmad & Hugh Chipman (Acadia University)
Assessing the Accuracy of BART Credible Intervals for Prediction / Évaluation de la précision des intervalles de crédibilité à des fins prédictives  
- 16:00 - 16:15 Nathaniel Stevens, Stefan Steiner & Jock MacKay (University of Waterloo)
Quantifying Agreement Between Two Methods of Measurement / Quantification de l'adéquation entre deux méthodes de mesure  
- 16:15 - 16:30 Candemir Cigsar (Memorial University of Newfoundland)
Carryover Effects in Repairable Systems / Effets de report dans les systèmes réparables  
- 16:30 - 16:45 Hensley Hubert Mariathas & Candemir Cigsar (Memorial University of Newfoundland)
The Effect of Heterogeneity on the Assessment of Imperfect Repairs / Effet de l'hétérogénéité sur l'évaluation des réparations imparfaites  
- 16:45 - 17:00 Po Yang (University of Manitoba), Chang-Yun Lin (National Chung Hsing University)
Response Surface Methodology using Split-Plot Definitive Screening Designs / Méthodologie de surface de réponse avec plans de criblage définitifs à parcelles divisées  





15:30 - 17:00 **Session 1E-C3** (abstracts/résumés - page 126) **McCain 2184**







Statistical Science for Ocean Ecosystems

Science statistique pour les écosystèmes océaniques

Contributed Paper Session / Séance de communications libres

Chair/Président: Duncan Murdoch (Western University)

- 15:30 - 15:45 Joanna Mills Flemming (Dalhousie University)
Statistical Methods for Assessing Impacts of Overfishing / Méthodes statistiques pour l'évaluation des répercussions de la surpêche  
- 15:45 - 16:00 Stuart Carson & Joanna Mills Flemming (Dalhousie University)
Spatio-Temporal Statistical Modelling for Atlantic Cod (Gadus morhua) Abundance on the Scotian Shelf using Sample Survey Data / Modélisation statistique spatio-temporelle de l'abondance de la morue de l'Atlantique (Gadus morhua) sur le plateau néo-écossais à l'aide de données d'enquête  

- 16:00 - 16:15 Noel Cadigan & Jon Fisher (Marine Institute of Memorial University of Newfoundland), Darrell Mullowney & Earl Dawe (Fisheries and Oceans Canada)
A Spatio-Temporal Stock Assessment Model for Snow Crab in Newfoundland and Labrador / Un modèle spatio-temporel d'évaluation des stocks de crabe des neiges à Terre-Neuve-et-Labrador  
- 16:15 - 16:30 Lin Qin & Martin Lysy (University of Waterloo)
Detecting Sea Water Migration in Canadian Arctic Fish / Détection de la migration des poissons d'eau de mer en Arctique canadien
- 16:30 - 16:45 Daniel Dinsdale & Matías Salibián-Barrera (University of British Columbia)
Stochastic Modelling of Marine Mammal Tag Data / Modélisation stochastique des données d'étiquetage des mammifères marins  
- 16:45 - 17:00 Connie Stewart (University of New Brunswick Saint John)
Measuring Distance between Compositional Data Containing Zeros / Mesure de la distance entre des données compositionnelles comportant des zéros  

15:30 - 17:00 **Session 1E-C4** (abstracts/résumés - page 129)

McCain 2102

Statistics in Action: A Smörgåsbord

Statistique en action : un « pot-pourri » de présentations

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Wendy Lou (University of Toronto)

- 15:30 - 15:45 Yizhou Fang & Martin Lysy (University of Waterloo)
Fast and Flexible Non-linear Modeling of Financial Assets / Modélisation non linéaire rapide et souple des actifs financiers  
- 15:45 - 16:00 Philippa Swartz, Derek Bingham, Michael Grosskopf & Tim Swartz (Simon Fraser University)
The Quality of Pitches in Major League Baseball / La qualité des lancers dans la Ligue majeure de baseball  
- 16:00 - 16:15 Michael Guerzhoy & Nathan Taback (University of Toronto)
Hierarchical Bayesian Models for Uncertainty-Quantified Ranking of Restaurant Chains by Food Safety Compliance / Modèles bayésiens hiérarchiques pour classement avec incertitude quantifiée des chaînes de restaurants selon la conformité aux normes sanitaires alimentaires  
- 16:15 - 16:30 Anita Brobbey & Asokan Mulayath Variyath (Memorial University of Newfoundland)
Analysis of Ordered Categorical Data from Designed Experiments / Analyse de données catégorielles ordonnées de plans d'expériences  
- 16:30 - 16:45 Thuva Vanniyasingam, Gary Foster, Lehana Thabane & Alfonso Iorio (McMaster University)
Determining Individual-Specific Terminal Half-Life Estimates and Optimal Sampling Time Points for Patients with Hemophilia B Administered rFIXfc / Détermination des estimations individuelles spécifiques de demi-vie terminale et échantillonnage optimal des points dans le temps pour des patients atteints d'hémophilie B à qui on a administré rFIXfc  
- 16:45 - 17:00 Charles Bordet & Louis-Paul Rivest (Université Laval)
A Stochastic Pella Tomlinson Model and its Maximum Sustainable Yield / Le modèle stochastique de Pella Tomlinson et son rendement équilibré maximal  

17:00 - 19:00

Howe Hall - Cameron

Accredited Members Mixer

Réception pour les membres accrédités

17:00 - 18:00

**SSC Annual General Meeting
Assemblée générale annuelle de la SSC**

McCain - Scotiabank

18:30 - 20:30

**Student BBQ
Barbecue étudiant**

Howe Hall - Cameron Dining Room

Tuesday June 16**mardi 16 juin****8:15 - 17:00****McCain Lobby****Registration
Inscription****8:45 - 9:50 Session 2A-A (abstract/résumé - page 132)****McCain - Ondaatje****Gold Medal Address****Allocution du récipiendaire de la Médaille d'or**

Special Session / Conférence spéciale

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

8:45 - 9:50 Jiahua Chen (University of British Columbia)

EM-Test for Finite Mixture Models / Test EM pour les modèles de mélanges finis

**9:50 - 10:30****Student Union Building - McInness****Coffee Break****Pause-café****10:20 - 11:50 Session 2B-I1 (abstracts/résumés - page 133)****Mona Campbell 1108****Modelling of Temporally Correlated Multivariate Counts****Modélisation de décomptes multivariés corrélés temporellement**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Tariqul Hasan (University of New Brunswick)

10:20 - 10:50 Dimitris Karlis (Athens University of Economics and Business)

On Multivariate Time Series for Counts Based on Multivariate Integer Autoregressive models / Séries chronologiques multivariées pour les dénombrements basées sur des modèles autorégressifs d'entiers multivariés



10:50 - 11:20 Renjun Ma (University of New Brunswick), Edward Hughes (Edward Hughes Consulting)

Analysis of Multivariate Longitudinal Count Data / Analyse des données de dénombrement longitudinales multivariées



11:20 - 11:50 Gary Sneddon (Mount Saint Vincent University), Tariqul Hasan & Renjun Ma (University of New Brunswick)

A State-Space Model for Multivariate Count Time Series with Extra Zeros / Modèle d'espace d'états pour séries chronologiques de dénombrement multivariées avec surreprésentation de zéros

**10:20 - 11:50 Session 2B-I2 (abstracts/résumés - page 135)****McCain - Ondaatje****Recent Advances in Copula Dependence Modelling****Avancées récentes en modélisation de la dépendance au moyen de copules**

Invited Paper Session / Présentations sur invitation

Chair/Présidente: Debbie Dupuis (HEC Montréal)

Organizer/Responsable: Elif Acar (University of Manitoba)

10:20 - 10:50 Louis-Paul Rivest (Université Laval), François Verret (Statistique Canada), Sophie Baillargeon (Université Laval)

Copulas and Small Area Estimation / Copules et estimation dans des petits domaines



10:50 - 11:20 Bruno Rémillard (HEC Montréal), Christian Genest & Johanna Nešlehová (McGill University)

Copulas for Discrete or Mixed Data and Applications / Copules pour données discrètes ou mixtes et applications





11:20 - 11:50 Elif Acar (University of Manitoba)
Non-simplified Vine Copula Models / Modèles de copules en vignes non simplifiées  



10:20 - 11:50 Session 2B-I3 (abstracts/résumés - page 137) **Rowe 1020**



Recent Advances in Infectious Disease Epidemiology
Avancées récentes et épidémiologie des maladies infectieuses

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Rob Deardon (University of Calgary)

10:20 - 10:50 Elaine Nsoesie (Harvard University)
Modeling the Chikungunya Epidemic in the Caribbean / Modélisation de l'épidémie de chikungunya dans les Caraïbes  

10:50 - 11:20 Patrick Brown (University of Toronto)
Having Fun with the Spatio-Temporal Log-Gaussian Cox Process / Amusons-nous avec le processus Cox log-gaussien spatio-temporel  

11:20 - 11:50 Razvan Romanescu (University of Guelph), Rob Deardon (University of Calgary)
Explaining Patterns in Influenza Count Data via Epidemic Network Models / Explications des tendances dans les données de dénombrement de la grippe à l'aide de modèles épidémiologiques en réseau  

10:20 - 11:50 Session 2B-I4 (abstracts/résumés - page 139) **Rowe 1028**

Statistical Inference in the Presence of Missing Survey Data
Inférence statistique en l'absence de certaines données de sondage



Invited Paper Session / Présentations sur invitation



Chair/Présidente: Jenny Thompson (US Census Bureau)



Organizer/Responsable: David Haziza (Université de Montréal)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

10:20 - 10:50 Jae-Kwang Kim (Iowa State University), Shu Yang (Harvard University)
A Note on Multiple Imputation for General-Purpose Estimation / Note sur l'imputation multiple pour l'estimation ordinaire  

10:50 - 11:20 Changbao Wu, Min Chen & Mary Thompson (University of Waterloo)
Empirical Likelihood Methods for Two Survey Samples with Data Missing-by-Design / Méthodes de vraisemblance empirique pour deux échantillons de sondage avec données délibérément omises  

11:20 - 11:50 Jongho Im, Jae-Kwang Kim & Wayne A. Fuller (Iowa State University)
Two-Per-Stratum Sampling for Fractional Hot Deck Imputation / Échantillonnage à deux échantillons par strate pour imputation hot-deck fractionnaire  

10:20 - 11:50 Session 2B-I5 (abstracts/résumés - page 141) **McCain - Scotiabank**







Uncertainty Quantification in Ocean and Environmental Models
Quantification de l'incertitude dans les modèles océanographiques et environnementaux

Invited Paper Session / Présentations sur invitation

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

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

- 10:20 - 10:50 Oxana Chkrebti (Ohio State University), David Campbell (Simon Fraser University), Mark Girolami (University of Warwick), Ben Calderhead (Imperial College London)
Quantifying Numerical Uncertainty in Models of Fluid Dynamics / Quantification de l'incertitude numérique dans les modèles de dynamique des fluides  
- 10:50 - 11:20 Michael Dowd (Dalhousie University), Paul Mattern (University of California), Jon Briggs (University of Auckland), Katja Fennel (Dalhousie University), Renate Meyer (University of Auckland)
Uncertainty Quantification for Biological Ocean Models / Quantification de l'incertitude dans les modèles océaniques biologiques  
- 11:20 - 11:50 William Welch, Tianji Shi & Douw G. Steyn (University of British Columbia)
Dynamic Space-Time Modelling of Ozone Processes / Modélisation spatiotemporelle dynamique des processus d'ozone  

10:20 - 11:35 Session 2B-C1 (abstracts/résumés - page 143) **McCain 2198**

Biostatistics: Methodological Innovation 1

Biostatistique : innovation méthodologique 1

Contributed Paper Session / Séance de communications libres

Chair/Président: Jason Nielsen (Carleton University)

- 10:20 - 10:35 Hai Yan Liu, Pierre-Jérôme Bergeron & Ioana Schiopu-Kratina (University of Ottawa)
Conditional GEE for Recurrent Gap Times / Équation d'estimation généralisée conditionnelle pour événement récurrents  
- 10:35 - 10:50 Nathalie Moon, Richard Cook & Leilei Zeng (University of Waterloo)
Estimating Disease Incidence with Truncated and Misclassified Response Data / Estimer l'incidence de maladie avec des réponses mal classées et tronquées  
- 10:50 - 11:05 Omidali Aghababaei Jazi, Masoud Asgharian & Abbas Khalili (McGill University)
Simultaneous Parameter Estimation and Variable Selection in Cox Proportional Hazards Model for Length-Biased Data / Estimation de paramètres et sélection de variables simultanées dans les modèles de risques proportionnels de Cox pour des données à biais de longueur  
- 11:05 - 11:20 Bo Deng & Keumhee Chough Carriere (University of Alberta)
Testing Simultaneous Marginal Homogeneity in Clustered Matched-Pair Multinomial Data / Test d'homogénéité marginale simultanée avec données multinomiales groupées par paires  
- 11:20 - 11:35 Parisa Gavanji (Queen's university), Bingshu Chen & Wenyu Jiang (Queen's University)
Penalized Likelihood Ratio Test for Biomarker Threshold Model with an Unknown Cut-Point in Clinical Trials / Test du rapport des vraisemblances pénalisé pour modèle de seuil de biomarqueurs avec seuil inconnu lors d'essais cliniques  





10:20 - 11:50 Session 2B-C2 (abstracts/résumés - page 146) **McCain 2102**









Methodological Advances: A Smörgåsbord

Progrès méthodologiques : un « pot-pourri » de présentations

Contributed Paper Session / Séance de communications libres

Chair/Président: Linglong Kong (University of Alberta)

- 10:20 - 10:35 Farnoosh Abbas Aghababazadeh, Mayer Alvo & David R. Bickel (University of Ottawa)
Local False Discovery Rate Estimate Based on Reference Set / Estimation du taux local de fausses découvertes à partir d'un ensemble de référence  
- 10:35 - 10:50 Iraj Yadegari & Éric Marchand (Université de Sherbrooke)
Post-Selection Point Estimation and Predictive Inference / Estimation et densité prédictive après sélection  

- 10:50 - 11:05 Aziz Imoudden & Éric Marchand (Université de Sherbrooke), William Strawderman (Rutgers University), Othmane Kortbi (United Arab Emirates University)
Estimation of Predictive Density for a Constrained Gamma Model / Estimation de densité prédictive pour un modèle Gamma sous contrainte  
- 11:05 - 11:20 Nasser Sadeghkhan & Éric Marchand (Université de Sherbrooke)
Predictive Density Estimation with Additional Information / Estimation d'une densité prédictive avec information additionnelle  
- 11:20 - 11:35 Adam Rahman & Wayne Oldford (University of Waterloo)
Minimal Spanning Tree Based Data Configurations -- Construction and Distributions / Configurations de données basées sur des arbres couvrants minimaux -- Construction et distributions  
- 11:35 - 11:50 Matthew Pencer (McGill University)
Post-Selection Inference in Multinomial Logistic Regression / Inférence post-sélection en régression logistique multinomiale  













10:20 - 11:50 **Session 2B-C3** (abstracts/résumés - page 149) **McCain 2170**

Statistical Science for Genetics and Genomics

Science statistique pour la génétique et la génomique

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Karen Kopciuk (Cancer Epidemiology and Prevention Research, AHS/University of Calgary)









- 10:20 - 10:35 Shofiqul Islam, Sonia Anand, Jemila Hamid, Lehana Thabane & Joseph Beyene (McMaster University)
Comparing the Performance of Linear and Nonlinear Principal Components in the Context of High Dimensional Genomic Data Integration / Comparer la performance des composantes principales linéaires et non-linéaires dans le contexte de l'intégration des données génomiques de grande dimension  
- 10:35 - 10:50 Gun Ho Jang, Lincoln Stein & John McPherson (Ontario Institute for Cancer Research)
A Copy-Number Recalibration Method using Object Integration for Next Generation Sequencing Data / Une méthode de recalage du nombre de copies utilisant l'intégration objet pour séquencer les données de prochaine génération  
- 10:50 - 11:05 Li Li, Hong Gu & Toby Kenney (Dalhousie University)
A New Approach to Detect Genetic Recombination in Phylogeny / Nouvelle approche pour détecter la recombinaison génétique dans la phylogénèse  
- 11:05 - 11:20 Kun Liang (University of Waterloo)
A Hidden Markov Tree Model for Multiple Hypotheses Testing of Gene Ontology Gene Sets / Modèle d'arbre de Markov caché pour tester les nombreuses hypothèses de l'ontologie des gènes sur des ensembles de gènes  
- 11:20 - 11:35 Farhad Shokoohi & Aurélie Labbe (McGill University), David Stephens (McGill University)
DNA Methylation Analysis; A Thorough Comparison of Available Analytic Tools / Analyse de données de méthylation de l'ADN : une comparaison des outils d'analyse disponibles  
- 11:35 - 11:50 Thierry Chekouo Tekougang, Francesco Stingo, Michele Guindani & Kim-Anh Do (UT MD Anderson Cancer Center)
A Bayesian Predictive Model for Imaging Genetics with Application to Schizophrenia / Un modèle prédictif bayésien pour la génétique d'imagerie avec application à la schizophrénie  
-

10:20 - 11:40 Session 2B-C4 (abstracts/résumés - page 152) **McCain 2184**

Statistics Education**Éducation en statistique**

Contributed Paper Session / Séance de communications libres

Chair/Président: Jeffrey Picka (University of New Brunswick)

- 10:20 - 10:40 Sohee Kang (University of Toronto Scarborough)
Do Collaborative Learning Strategies Improve Students' Learning in Statistics Course? / Est-ce que les stratégies d'apprentissage coopératif améliorent l'apprentissage des étudiants dans un cours de statistique ?  
- 10:40 - 11:00 Jennifer Thornton (Mount Saint Vincent University)
Online Teaching and Assessing Student Learning / Enseignement en ligne et évaluation de l'apprentissage des étudiants  
- 11:00 - 11:20 Asokan Mulayath Variyath (Memorial University of Newfoundland)
Teaching Statistics to First Year Undergraduate Students - Challenges / Enseigner les statistiques aux étudiants de première année du premier cycle - Défis  
- 11:20 - 11:40 Swarna Weerasinghe, Kevin Wilson & Adam Muir (Dalhousie University)
Use of Social Media as a Secure and Supportive Learning Environment for a Graduate Level Biostatistics Course in Medicine / Utilisation des médias sociaux comme un milieu d'apprentissage sécuritaire et favorable pour un cours de biostatistique de deuxième cycle en médecine  

12:00 - 13:30 **McCain 2170**

Accreditation General Meeting**Assemblée générale des membres accrédités**

12:00 - 18:00 **McCain 2184**

Award for Case Studies in Data Analysis Committee 2**Comité du prix pour les études de cas et l'analyse de données 2**

12:00 - 13:30 **McCain 2102**

CJS Editorial Board**Comité de rédaction de la RCS**

12:00 - 13:30 **McCain 2198**

Census at School Committee**Comité de recensement à l'école**

12:00 - 13:30 **McCain 2132**

Committee on New Investigators**Comité des nouveaux chercheurs**

12:00 - 13:00 **Rowe 1020**

MacMillan Education Demo: Using LaunchPad in Introductory Statistics Courses for Online Homework and Assessment**À l'aide de LaunchPad pour un cours d'introduction aux statistiques aux fins d'évaluation et de travaux en ligne**

12:00 - 13:30**Rowe 1011****Office Committee
Comité du Bureau**

12:00 - 13:30**McCain - Ondaatje****Public Relations Committee
Comité des relations publiques**

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

12:00 - 17:30**Session 2C-P**

(abstracts/résumés - page 154)

Student Union Building - McInness**Poster Session****Séance d'affichage**

Contributed Poster Session / Séance d'affichage libre

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)

(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

12:00 - 17:30

Olu Awosoga & Bonnie Lee (University of Lethbridge)

Multi-site Pilot Randomized Control Trial of Congruence Couple Therapy for Problem Gamblers / Étude pilote randomisée contrôlée multi-sites pour la thérapie de couple basée sur la congruence pour les problèmes de jeu

12:00 - 17:30

Entao Chen, Ying Zhang, Holger Teismann & Khurram Nadeem (Acadia University)

A Beta Distribution Based Hierarchical Multinomial Model for Age Structure Analysis of Wildlife Animal / Modèle multinomial hiérarchique basé sur la loi bêta pour l'analyse de la structure d'âges d'animaux sauvages

12:00 - 17:30

Hao (Nelson) Chen (University of British Columbia)

Tree-based Methods for Emulation of a Complex Computer Model / Méthodes basées sur les arbres pour l'émulation d'un modèle informatique complexe

12:00 - 17:30

Jingjia Chu, Reg Kulperger & Hao Yu (University of Western Ontario)

Modelling the Common Risk among Equities: A Multivariate Time Series Model with an Additive GARCH Structure / Modélisation du risque commun entre les actions : un modèle multivarié de séries chronologiques avec une structure GARCH additive

12:00 - 17:30

Hyukjun Gweon (University of Waterloo), Lars Kaczmarek (GESIS–Leibniz Institute for the Social Sciences), Matthias Schonlau (University of Waterloo)

Automated Occupational Coding Using Machine Learning / Codification automatisée des professions par l'apprentissage automatique

12:00 - 17:30

Fang He & Duncan Murdoch (University of Western Ontario)

Subscene design in R package rgl / Conception de sous-scène dans le paquet rgl du progiciel R

12:00 - 17:30

James Kiberd, Adam Muir & Kevin Wilson (Dalhousie University)

The Good, the Bad, and the Cite-Able: Study Quality and Statistical Significance as Predictors of Study Citation Rates / Le bon, le mauvais et l'invocable : qualité des études et signification statistique, comme les prédicteurs des taux de citation d'études

12:00 - 17:30

Karen A Kopciuk (Cancer Epidemiology and Prevention Research, AHS/University of Calgary), Ji Ruan, Farshad Farshidfar & Oliver F. Bathe (University of Calgary)

Distribution-Based Imputation for Left-Censored Metabolomics Data / Imputation basée sur la distribution pour des données métabolomiques censurées à gauche

- 12:00 - 17:30 Takuma Kurosawa, Asanao Shimokawa & Etsuo Miyaoka (Tokyo University of Science)
On Asymptotic Normality of Maximum Partial Likelihood Estimators for Binary Time Series with Transition Models / Normalité asymptotique des estimateurs du maximum de vraisemblance partielle pour des séries chronologiques binaires sur des modèles de transition
- 12:00 - 17:30 Wan-Chen Lee (Health Canada)
Statistical Strategies for Evaluating the Exposure of Chemical Mixtures during Pregnancy / Stratégies statistiques pour évaluer l'exposition aux à des mélanges chimiques pendant la grossesse
- 12:00 - 17:30 Michael Li & Benjamin Bolker (McMaster University)
Modeling High-Resolution Animal Telemetry Data: Hidden Markov Models and Extensions / Modélisation en haute résolution des données animales par télémétrie : modèles à chaîne de Markov cachée et extensions
- 12:00 - 17:30 Jason Loeppky (University of British Columbia)
Parameterization of the Gaussian Process for Modelling a Blackbox Function / Paramétrage du processus gaussien pour la modélisation d'une fonction de boîte noire
- 12:00 - 17:30 Bin Luo, Alanna Edge, Charmaine Dean, Kathleen Hill & Reg Kulperger (University of Western Ontario)
Spatial Statistical Tools for Genome-Wide Mutation Thundershower Detection under a Microarray Probe Sampling System / Outils statistiques spatiaux pour la détection d'orage de mutations pangénomique sous un système d'échantillonnage de sondes de microréseaux
- 12:00 - 17:30 Adam Muir, James Kiberd & Kevin Wilson (Dalhousie University)
Did a Systematic Review Just Eat Your Citation? The Impact of Reviews on the Citation Frequency of Primary Studies / Est-ce qu'un article de revue vient de dévorer votre citation ? L'impact des revues sur la fréquence de citation des études primaires
- 12:00 - 17:30 Drew Neish, Sanjeena Subedi & Zeny Feng (University of Guelph)
Clustering Using Mixtures of Dirichlet-multinomial Regression Models / Partitionnement des données à l'aide de modèles de régression de mélanges Dirichlet multinomiaux
- 12:00 - 17:30 Aurélien Nicosia, Thierry Duchesne & Louis-Paul Rivest (Laval University)
A General Directional Random Walk Model: Application to Animal Movement / Un modèle général de marche aléatoire directionnelle : application au mouvement animalier
- 12:00 - 17:30 Andrew Porter, O. Brian Allen & Gary Umphrey (University of Guelph)
Estimating an Experimental Error Variance for Fractional Factorial Designs / Estimation de la variance de l'erreur expérimentale pour des plans factoriels fractionnaires
- 12:00 - 17:30 Shi Qiu, Longhai Li & Cindy Feng (University of Saskatchewan)
Cross-Validatory Model Comparison and Divergent Region Detection using iIS and iWAIC for Disease Mapping / Modèle de comparaison validé croisé et détection de région divergente en employant iIS et iWAIC pour la cartographie d'une maladie
- 12:00 - 17:30 Rhonda Rosychuk & Iram Usman (University of Alberta)
An Examination of Spatial Scan Statistics Based on Time to Event Data / Étude des statistiques de scanographie spatiale basées sur des données de durée avant événement
- 12:00 - 17:30 Asanao Shimokawa, Takuma Kurosawa & Etsuo Miyaoka (Tokyo University of Science)
Construction of Tree-Structured Prediction Model Based on Interval-Valued Covariates / Création de modèles arborescents de prédiction basés sur des covariables par intervalle
- 12:00 - 17:30 Anjali Silva, Sanjeena Dang, Steven Rothstein, Gregory Downs & Yong-Mei Bi (University of Guelph)
Comparative Analysis of Clustering Techniques for RNA-seq Data / Analyse comparative des techniques de classification des données de séquençage ARN
- 12:00 - 17:30 Gabrielle Simoneau (McGill University)
An Empirical Comparison of Methods to Meta-Analyze Individual Patient Data of Diagnostic Accuracy / Comparaison empirique de méthodes de méta-analyse des données individuelles sur la précision du diagnostic









- 12:00 - 17:30 Yubin Sung (University of Guelph)
A Two-Step Method for Genetic Association Analysis with Multiple Longitudinal Traits of Samples of Related Subjects / Méthode en deux étapes pour l'analyse d'associations génétiques avec des caractères longitudinaux multiples d'échantillons de sujets apparentés
- 12:00 - 17:30 Fode Tounkara & Louis-Paul Rivest (Université Laval)
Archimedean Copulas for Clustered Binary Data / Copule archimédienne pour des données de Bernoulli corrélées
- 12:00 - 17:30 Douglas Whitaker (University of Florida)
Students' Understanding of Histograms and Bar Graphs: Results from the LOCUS Assessments / Compréhension estudiantine des histogrammes et diagrammes à bâtons : résultats des évaluations LOCUS
- 12:00 - 17:30 Douglas Whitaker (University of Florida)
Transition and Collaboration: The Identity of an Advanced Placement Statistics Teacher / Transition et collaboration : l'identité d'un enseignant d'un cours de perfectionnement de placement en statistique
- 12:00 - 17:30 Kevin Wilson, James Kiberd & Adam Muir (Dalhousie University)
Out of Cite, Out of Mind? The Effects of Citation by Review Articles on Frequency of Citation in the Epidemiological Literature / Hors de vue, oublié ? Effets de la citation par des articles de synthèse sur la fréquence de citation dans la littérature épidémiologique
- 12:00 - 17:30 Lingyun Ye (Canadian Centre for Vaccinology), Bruce Smith (Dalhousie University), Shelly McNeil (Canadian Centre for Vaccinology)
The Validity of Test-Negative Case-Control Design / Validité de la méthode cas-témoins avec tests négatifs
- 12:00 - 17:30 Boyko Zlatev (University of Alberta)
Studying Selections of Poems by Statistical Methods / Étude de sélections de poèmes par des méthodes statistiques





13:30 - 15:00 **Session 2D-I1** (abstracts/résumés - page 168) **Rowe 1011**

Actuarial Graduate Students Presentations

Présentations des étudiants gradués en science actuarielle

Invited Paper Session / Présentations sur invitation
Chair/Président: Étienne Marceau (Université Laval)
Organizer/Responsable: Hélène Cossette (Université Laval)
Session sponsored by the Actuarial Science Section
Séance parrainée par le Groupe de science actuarielle

- 13:30 - 13:45 Sabrina Zhang & Cary Tsai (Simon Fraser University)
The Optimal Payment Reduction Ratios For a Catastrophe Bond. / Les rapports de réduction de paiements optimaux pour une obligation catastrophe  
- 13:45 - 14:00 Ji Ruan & Alex de Leon (University of Calgary)
Cluster Analysis of Gene Expression Profiles via Flexible Count Models for RNA-seq Data / Analyse de regroupement de profils d'expression génique via des modèles de dénombrement flexibles pour des données seq-ARN  
- 14:00 - 14:15 Jean-François Bégin (HEC Montréal), Mathieu Boudreault (Université du Québec à Montréal), Geneviève Gauthier (HEC Montréal)
Credit Risk in Corporate Spreads during the Financial Crisis of 2008 / Risque de crédit dans les rendements des obligations de sociétés durant la crise financière de 2008  
- 14:15 - 14:30 Itre Mtalai, Hélène Cossette, Etienne Marceau & Déry Veilleux (Université Laval)
Aggregation Methods for Portfolios of Dependent Risks with Archimedean Copulas / Méthodes d'agrégation des portefeuilles de risques dépendants en utilisant les copules archimédiennes  







- 14:30 - 14:45 Mirabelle Huynh & David Landriault (University of Waterloo), Tianxiang Shi (University of Nebraska-Lincoln), Gordon E Willmot (University of Waterloo)
On a Risk Model with Claim Investigation / Un modèle de risque avec enquête de réclamation  
- 14:45 - 15:00 Yisub Kye & Edward Furman (York University), Raluca Vernic (Institute for Mathematical Statistics and Applied Mathematics, Bucharest)
On the probability of ruin for non-identically distributed and dependent claims / À propos de la probabilité de ruine pour des réclamations non-identiquement distribuées et dépendantes  

13:30 - 15:00 Session 2D-I2 (abstracts/résumés - page 172) **Rowe 1020**

Big Data and Neuroimaging
Données volumineuses et neuroimagerie

Invited Paper Session / Présentations sur invitation

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

- 13:30 - 14:00 Sebastian Kurtek (Ohio State University), Ian H. Jermyn (Durham University), Qian Xie & Anuj Srivastava (Florida State University)
Elastic Statistical Shape Models of Anatomical Surfaces using Square-Root Normal Fields / Modèles de forme statistiques élastiques de surfaces anatomiques utilisant des champs normaux racine carrée  
- 14:00 - 14:30 Linglong Kong (University of Alberta), Hongtu Zhu (University of North Carolina at Chapel Hill), Jianqing Fan (Princeton University), Xingcai Zhou (Tongling University), Partha Sarathi Mukherjee (Boise State University), Baiguo An & Chao Huang (University of North Carolina at Chapel Hill)
Robust Spatial Varying Coefficient Model in Neuroimaging Data Analysis / Modèle de coefficient spatial variable en analyse de données en neuro-imagerie  
- 14:30 - 15:00 Timothy Johnson & Ming Feng (University of Michigan), Farouk Nathoo (University of Victoria)
Bayesian Computation for Log Gaussian Cox Processes with Application to Neuroimaging / Calcul bayésien pour des processus de Cox log-gaussiens avec applications en neuro-imagerie  







13:30 - 15:00 Session 2D-I3 (abstracts/résumés - page 174) **McCain 2102**

Recent Developments in Nonparametric Statistics
Développements récents en statistique non paramétrique

Invited Paper Session / Présentations sur invitation

Chair/Président: Paul Cabilio (Acadia University)

Organizer/Responsable: Ying Zhang (Acadia University)







- 13:30 - 14:00 Mayer Alvo (University of Ottawa)
Bridging the Gap: A Nonparametric Likelihood Function for Ranking Data / Combler le fossé : une fonction de vraisemblance non paramétrique pour des données de classement  
- 14:00 - 14:30 Herold Dehling (Ruhr-Universitaet Bochum), Olivier Durieu (Universite de Tours), Marco Tusche (Ruhr-University Bochum), Dalibor Volny (Universite de Rouen)
New Techniques for Empirical Processes of Dependent Data / Nouvelles techniques pour les processus empiriques de données dépendantes  
- 14:30 - 15:00 Ying Zhang, Paul Cabilio & Amber Corkum (Acadia University)
Robust Confidence Interval Construction with Applications to Time Series Intervention Analysis / Construction d'intervalles de confiance robustes avec applications à l'analyse d'intervention en séries chronologiques  

13:30 - 15:00 **Session 2D-I4** (abstracts/résumés - page 176) **McCain - Scotiabank**

Showcasing New Investigating Women in Statistics

Nouvelles chercheuses en statistique

Invited Paper Session / Présentations sur invitation
 Chair/Présidente: Juli Atherton (McGill University)
 Organizer/Responsable: Laura Cowen (University of Victoria)
 Session sponsored by the Committee on Women in Statistics
 Séance parrainée par le Comité des femmes en statistique







- 13:30 - 14:00 Mireille Schnitzer (Université de Montréal), Russell Steele, Michèle Bally & Ian Shrier (McGill University)
 The Causal Inference Paradigm for Network Meta-analysis with Implications for Feasibility and Practice / Paradigme de l'inférence causale pour la méta-analyse en réseau et ses conséquences sur la faisabilité et la pratique  
- 14:00 - 14:30 Yildiz Yilmaz (Memorial University)
 Genetic Association Studies Based on Joint Models of Multiple Phenotypes / Études d'association génétique basées sur des modèles conjoints de phénotypes multiples  
- 14:30 - 15:00 Liangliang Wang (Simon Fraser University), Alexandre Bouchard-Côté (University of British Columbia), Arnaud Doucet (Oxford University)
 Scalable Monte Carlo Methods for Bayesian Phylogenetic Inference / Méthodes de Monte-Carlo extensibles pour l'inférence bayésienne phylogénétique  

13:30 - 15:00 **Session 2D-I5** (abstracts/résumés - page 178) **Rowe 1028**

Survival Modelling with Environmental Applications

Modélisation de la survie et applications environnementales

Invited Paper Session / Présentations sur invitation
 Chair/Président: John Braun (University of British Columbia)
 Organizers/Responsables: John Braun (University of British Columbia) and/et Douglas Woolford (Wilfred Laurier University)
 Session sponsored by the Biostatistics Section
 Séance parrainée par le Groupe de biostatistique





- 13:30 - 14:00 James Stafford (University of Toronto)
 Accommodating Spatial and Temporal Uncertainty in Wildfire Ignitions. / Prise en compte de l'incertitude spatiale et temporelle dans le déclenchement des feux de forêts  
- 14:00 - 14:30 Yi Xiong & Joan Hu (Simon Fraser University), John Braun (University of British Columbia)
 Exploring Forest Fire Duration Times / Exploration de la durée des feux de forêts  
- 14:30 - 15:00 Thierry Duchesne (Université Laval), John Braun (University of British Columbia), Peter Hall (University of Melbourne)
 Kernel and Deconvolution Methods of Density Estimation for Interval-Censored Data / Méthodes du noyau et de déconvolution pour l'estimation d'une densité avec données censurées par intervalles  

13:30 - 15:00 **Session 2D-I6** (abstracts/résumés - page 180) **McCain - Ondaatje**

Teaching Effective Communication of Statistical Ideas to Non-Statisticians

Comment enseigner à bien communiquer les idées statistiques aux non-statisticiens

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

- 13:30 - 14:00 Nathan Taback (University of Toronto)
Teaching Effective Communication Skills to Statistics Students / Enseignement de techniques de communication efficaces pour des étudiants en statistique  
- 14:00 - 14:30 Lehana Thabane (McMaster University)
Effective Strategies to Communicate Statistical Issues to Clinicians in Health Research / Stratégies efficaces pour communiquer les enjeux statistiques aux cliniciens en recherche sur la santé  
- 14:30 - 15:00 Don McLeish, Discussant (University of Waterloo)

13:30 - 14:45 Session 2D-C1 (abstracts/résumés - page 182) **McCain 2170**

Bayes and Likelihood: Models and Algorithms
Bayes et la vraisemblance : modèles et algorithmes







Contributed Paper Session / Séance de communications libres
Chair/Président: Renjun Ma (University of New Brunswick)





- 13:30 - 13:45 Michael Newton (University of Wisconsin, Madison)
Graph Partition Modeling / Modélisation de partitionnement des graphes  
- 13:45 - 14:00 Biljana Jonoska Stojkova & David Campbell (Simon Fraser University)
Simulated Tempering without Normalizing Constants / Atténuation simulée sans constantes normalisantes  
- 14:00 - 14:15 Daniel Evan Hockey, Shirley Mills & Song Cai (Carleton University)
Predicting the Attributes of Nodes in a Dynamic Network / Prévion des attributs des noeuds dans un réseau dynamique  
- 14:15 - 14:30 Neil Spencer & Alexandre Bouchard-Côté (University of British Columbia)
A Sequential Monte Carlo Sampler for Deterministically Related Continuous Random Variables / Échantillonneur de Monte-Carlo séquentiel pour des variables aléatoires continues liées de manière déterministe  
- 14:30 - 14:45 Michael Regier (West Virginia University), Erica Moodie (McGill University)
The Orthogonally Partitioned EM Algorithm: Extending the EM Algorithm for Algorithmic Stability for Bias Correction due to Imperfect Data / L'algorithmes EM orthogonalement partitionné : étendre l'algorithmes EM pour la stabilité algorithmique dans la correction du biais attribué à des données imparfaites  

13:30 - 14:45 Session 2D-C2 (abstracts/résumés - page 185) **McCain 2198**

Robust Methods
Méthodes robustes

Contributed Paper Session / Séance de communications libres
Chair/Présidente: Debbie Dupuis (HEC Montréal)

- 13:30 - 13:45 Philippe Gagnon (Université de Montréal), Alain Desgagné (Université du Québec à Montréal), Mylène Bédard (Université de Montréal)
Robustness to Outliers in a Bayesian Simple Linear Regression Model / Inférence robuste sur les paramètres d'une régression linéaire simple bayésienne  
- 13:45 - 14:00 Maryam Sohrabi & Mahmoud Zarepour (University of Ottawa)
Bootstrapping the Robust Estimates of the Mean Vector for Multivariate Heavy-Tailed Distributions with Different Indices of Stability / Bootstrap pour les estimations robustes du vecteur moyen pour des distributions de probabilité à ailes lourdes aléatoires avec différents indices de stabilité  
- 14:00 - 14:15 Wei Tu & Rohana Karunamuni (University of Alberta)
Robust Efficient Generalized M-Estimation in Regression Models / M-Estimation généralisée robuste et efficace dans des modèles de régression  











- 14:15 - 14:30 Olu Awosoga (University of Lethbridge), Joseph McKean & Bradley Huitema (Western Michigan University)
Meta-type Analysis of Multiple Baseline Time Series Design Intervention Models for Dependent and Independent Series / Méta-analyse de modèles d'interventions en séries chronologiques avec plusieurs situations initiales pour des séries dépendantes et indépendantes  
- 14:30 - 14:45 Yang Zhao (University of Regina)
Robust Imputation to Unified Approach for Regression Models with Data Missing by Design / Imputation robuste à l'approche unifiée pour les modèles de régression avec données manquantes selon le modèle  

13:30 - 14:45 Session 2D-C3 (abstracts/résumés - page 188) **McCain 2184**

Theoretical Developments

Développements théoriques

Contributed Paper Session / Séance de communications libres
Chair/Président: Deli Li (Lakehead University)

- 13:30 - 13:45 Sara DeVries & Mohamed Amezziane (Central Michigan University), S. Ejaz Ahmed (Brock University)
On Shrinking Nonparametric Regression Models to Linear Regression Models / À propos du rétrécissement de modèles de régression non paramétriques à des modèles de régression linéaire  
- 13:45 - 14:00 Fuqi Chen (The Fields Institute for Research in Mathematical Sciences), Sévérien Nkurunziza (Université de Windsor / Université de Sherbrooke)
On Stein-Rules in Multivariate Regression with Change-Points / Sur les règles de Stein en régression multivariée avec des points de rupture  
- 14:00 - 14:15 Miklos Csorgo (Carleton University), Yuliya Martsynyuk (University of Manitoba)
Testing for Change in the Mean via Weighted Partial Sums Processes / Vérification du changement dans les processus de sommes partielles moyens et pondérés  
- 14:15 - 14:30 Chi Dong & Michael Kouritzin (University of Alberta)
Replication and Approximation of Stochastic Processes in Generalized Spaces / Réplication et approximation de processus stochastiques dans des espaces généralisés  
- 14:30 - 14:45 Ammar Sarhan (Dalhousie University), Joseph Apaloo (St. Francis Xavier University)
The Bivariate Two Parameter Bathtub Hazard Rate Distribution / La loi bivariée avec taux de défaillance en baignoire à deux paramètres  



14:45 - 15:30 **Student Union Building - McInness**
Coffee Break
Pause-café


15:30 - 17:00 Session 2E-I1 (abstracts/résumés - page 191) **McCain - Ondaatje**

Extreme Value Theory and Applications

Théorie des valeurs extrêmes et applications

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

- 15:30 - 16:00 Francis Zwiers (University of Victoria)
Detection, Attribution and Extreme Climate and Weather Events / Détection, attribution et événements climatiques et météorologiques extrêmes  

- 16:00 - 16:30 Debbie Dupuis (HEC Montréal)
The Dynamics of Extreme Daily Temperatures / La dynamique des températures journalières extrêmes  
- 16:30 - 17:00 Edit Gombay (University of Alberta), Fuxiao Li (Northwest Polytechnical University, Xi'an, P.R. China), Hao Yu (University of Western Ontario)
Change Detection in Categorical Time Series / Détection des changements dans les séries temporelles catégorielles  

15:30 - 17:00 Session 2E-I2 (description - page 193) **McCain - Scotiabank**

Interviews, Grants, and Family Balance for New Investigators: A Panel Discussion
Entrevues, subventions et équilibre familial pour les nouveaux chercheurs (une table ronde)







Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et présidente: Mélina Mailhot (Concordia University)
Session sponsored by the New Investigator Committee
Séance parrainée par le Comité des nouveaux chercheurs

- 15:30 - 17:00 Jose Garrido (Concordia University)
15:30 - 17:00 Geneviève Lefebvre (Université du Québec à Montréal)
15:30 - 17:00 Stefan Steiner (University of Waterloo)

15:30 - 17:00 Session 2E-I3 (abstracts/résumés - page 194) **Rowe 1020**

Recent Developments in Statistical Genomic Modelling and Analysis
Développements récents en modélisation et analyse génomiques statistiques



Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et présidente: Jinko Graham (Simon Fraser University)
Session sponsored by the Biostatistics Section
Séance parrainée par le Groupe de biostatistique





- 15:30 - 16:00 Celia Greenwood, Aurélie Labbe, Antonio Ciampi & David Stephens (McGill University)
Methods for Analysis of DNA Methylation / Méthodes d'analyse pour la méthylation de l'ADN  
- 16:00 - 16:30 Gabriela Cohen Freue (University of British Columbia)
Borrowing Information from Genomics Data to Boost Proteomics Discoveries / Augmentation des découvertes en protéomique grâce à des données de génomique  
- 16:30 - 17:00 Laurent Briollais (University of Toronto)
Bayesian Graphical Models for Gene Network Analysis in Large-Scale Problems / Modèles graphiques bayésiens pour l'analyse de réseaux de gènes dans des problèmes à grande échelle  

15:30 - 17:00 Session 2E-I4 (abstracts/résumés - page 196) **Rowe 1028**

Statistical Approaches for Big Data
Approches statistiques des données volumineuses

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

- 15:30 - 16:00 Jeffrey Andrews (MacEwan University), Paul McNicholas (McMaster University), Mathieu Chali-four (MacEwan University)
Robust Model-Based Clustering of High-Dimensional Data / Regroupement à l'aide de modèles robustes de données à grandes dimensions  

- 16:00 - 16:30 Hong Gu, Michael Butler, Shen Ling, Toby Kenney & Chris Field (Dalhousie University)
Machine Learning Diagnoses on Patients Presenting Abdominal Pain / Diagnostics d'apprentissage automatique sur des patients présentant des douleurs abdominales  
- 16:30 - 17:00 Sunny Wang & Zhouqin He (St. Francis Xavier University)
Business Failure Prediction for Canadian Charitable Organizations Using Data Mining Tools / Préviation de faillite pour les organismes de bienfaisance canadiens par l'utilisation d'outils d'exploration de données  

15:30 - 17:00 Session 2E-I5 (abstracts/résumés - page 198) **Rowe 1011**

eLearning in Statistics – Successes and Opportunities

Apprentissage électronique en statistique – succès et opportunités

Invited Paper Session / Présentations sur invitation

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

Organizer/Responsable: Bethany White (Western University)

- 15:30 - 16:00 Ivo Dinov, Alexandr Kalinin & Syed Husain (University of Michigan), Nicolas Christou (UCLA)
Statistics Online Computational Resource (SOCR) - An Infrastructure for Technology-enhanced Trans-disciplinary Education / Statistics Online Computational Resource (SOCR) - Une infrastructure pour l'éducation transdisciplinaire améliorée par la technologie  
- 16:00 - 16:30 Alison Gibbs (University of Toronto)
A New Blend: An Introductory Statistics Course Integrating a MOOC and Active Learning Sessions, with Seasoning from the Students' Areas of Study / Un nouveau mélange : un cours d'introduction à la statistique intégrant un MOOC et des sessions d'apprentissage actif pimentées selon les domaines d'étude des étudiants  
- 16:30 - 17:00 Bethany White (Western University)
Going Online in a Big Way: Developing a First-year Statistics Course for the Online Ontario Initiative / Passer en ligne de façon importante : développer un cours de statistique de première année pour l'Initiative en ligne de l'Ontario  









15:30 - 16:45 Session 2E-C1 (abstracts/résumés - page 200) **McCain 2198**



Actuarial Science and Finance 2

Science actuarielle et finance 2

Contributed Paper Session / Séance de communications libres

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

- 15:30 - 15:45 René Ferland & François Watier (Université du Québec à Montréal)
Optimal Weighted Mean-Variance Portfolio Under a No-Bankruptcy Constraint / Portefeuille optimal à variance moyenne pondérée dans un contexte de contrainte de non-faillite  
- 15:45 - 16:00 François Watier (Université du Québec à Montréal), Moussa Kounta (The College of the Bahamas)
Regime-Switching Multi-Period Mean-Variance Optimal Portfolio Without Short-Selling / Portefeuille optimal multi-période de type moyenne-variance avec changement de régime sans vente à découvert  
- 16:00 - 16:15 Michelle Carey, James Ramsay & Christian Genest (McGill University)
A Generalized Smoothing Approach to the Calibration, Pricing and Hedging of Financial Options / Une approche de lissage généralisée à l'étalonnage, la tarification et la couverture des options financières  
- 16:15 - 16:30 Joel Sango & Pierre Duchesne (Université de Montréal)
On the Diagnostic of Vector Multiplicative Error Models / Sur le diagnostic des modèles vectoriels multiplicatifs à erreurs non négatives  

16:30 - 16:45 Martin Lysy (University of Waterloo), Natesh Pillai (Harvard University)
Statistical Inference for Stochastic Differential Equations with Memory / Inférence stochastique pour les équations différentielles stochastiques avec mémoire  



15:30 - 17:00 Session 2E-C2 (abstracts/résumés - page 203) **McCain 2102**



Biostatistics: Methodological Innovation 2



Biostatistique : innovation méthodologique 2



Contributed Paper Session / Séance de communications libres



Chair/Présidente: Cindy Feng (University of Saskatchewan)



15:30 - 15:45 Scarlett Kazimer (McGill University), Wenyu Jiang (Queen's University)
Sample Size Planning for the Cox Regression Model with Treatment-Biomarker Interaction / Planification de la taille échantillonnale pour le modèle de régression de Cox en présence d'interaction biomarqueur-traitement  

15:45 - 16:00 Matthew Stephenson (University of Guelph), Neil Klar (Western University), Gerarda Darlington (University of Guelph), Stephanie Dixon (Western University)
Effect of Accrual on Sample Size Calculation for Cluster Randomized Studies / Effet d'un accroissement sur le calcul de la taille échantillonnale pour des essais randomisés par grappes  

16:00 - 16:15 Osvaldo Espin-Garcia (University of Toronto / Lunenfeld-Tanenbaum Research Institute), Linda Hiraki (The Hospital for Sick Children), Angelo J. Canty (McMaster University), Andrew Paterson (The Hospital for Sick Children / University of Toronto), Shelley B. Bull (Lunenfeld-Tanenbaum Research Institute / University of Toronto)
Understanding Joint Models for Longitudinal and Survival Data Analysis of Genetic Association: A Simulation Study / Comprendre les modèles conjoints pour l'analyse des données longitudinales et de survie dans les associations génétiques : une étude de simulation  

16:15 - 16:30 Wanhua Su (MacEwan University), Yan Yuan (University of Alberta), Mu Zhu (University of Waterloo)
Two Threshold-Free Performance Metrics for Medical Tests: Average Positive Predictive Value and Area Under the ROC Curve / Deux mesures de performance sans seuil pour les tests médicaux : valeur prédictive positive moyenne et aire sous la courbe ROC  

16:30 - 16:45 Siu Hung Cheung (The Chinese University of Hong Kong), TongYu Lu (China Jiliang University), Wai Yin Poon (The Chinese University of Hong Kong)
Multiple Comparisons with a Control for Treatments with Skewed Ordinal Responses / Comparaisons multiples au moyen d'un contrôle des traitements avec des réponses ordinales asymétriques  

16:45 - 17:00 Abdelaziz Alsubie & Jianan Peng (Acadia University)
Simultaneous Statistical Inference for Low Dose Risk Estimation with Quantal Data in Benchmark Analysis / Inférence simultanée pour l'estimation du risque à faible dose avec des données quantiques pour l'analyse comparative  



15:30 - 16:45 Session 2E-C3 (abstracts/résumés - page 206) **McCain 2170**




Data Visualization and Multivariate Methods

Visualisation de données et méthodes multivariées

Contributed Paper Session / Séance de communications libres

Chair/Président: Richard Lockhart (Simon Fraser University)

15:30 - 15:45 Wayne Oldford & Adrian Waddell (University of Waterloo)
Loon: An Interactive and Extendable Statistical Data Visualization Toolkit / Loon : une boîte à outils interactive et extensible pour la visualisation de données statistiques  

- 15:45 - 16:00 Jeffrey Picka (University of New Brunswick)
The Need for Replication in Exploratory Computer Experiments / Le besoin de réplication dans les expériences exploratoires sur ordinateur  
- 16:00 - 16:15 Liang Jiayi, Shojaeddin Chenouri & Christopher Small (University of Waterloo)
Performance Analysis for Dimensionality Reduction / Analyse de performance pour la réduction de dimension  
- 16:15 - 16:30 Shojaeddin Chenouri (University of Waterloo), Behdad Mostafaiy & MohammadReza FaridRohani (Shahid Beheshti University)
Sparse Functional Linear Regression and Singular Function Estimation by Reproducing Kernel Hilbert Space Methods / Régression linéaire fonctionnelle simultanément éparsée et estimation singulière des fonctions par la reproduction des méthodes d'espace de Kernel Hilbert  
- 16:30 - 16:45 Jules Joseph S. de Tibeiro (Université de Moncton, Campus de Shippagan)
Statistical Validity Tests Based on Usual Probabilistic Models: An Alternative to Correspondence Analysis Based on Hellinger Metric. / Tests de validité statistique basés sur des modèles classiques de probabilité : une alternative à l'analyse des correspondances fondée sur la métrique de Hellinger.  













15:30 - 17:00 **Session 2E-C4** (abstracts/résumés - page 209) **McCain 2184**

Survey Methodology

Méthodes d'enquêtes

Contributed Paper Session / Séance de communications libres

Chair/Président: Noel Cadigan (Memorial University)

- 15:30 - 15:45 J. N. K. Rao (Carleton University), Yongsong Qin (Guangxi Normal University), Malgorzata Winiszewska (Carleton University)
Confidence Intervals for Mean and Distribution Function Under Imputation for Missing Data / Intervalles de confiance pour la moyenne et la fonction de répartition sous imputation pour les données manquantes  
- 15:45 - 16:00 Caren Hasler (Université de Neuchâtel), Radu V. Craiu (University of Toronto)
Nonparametric Imputation Method for Nonresponse in Surveys / Imputation non paramétrique pour la non-réponse dans les enquêtes  
- 16:00 - 16:15 Kyle Vincent (Bank of Canada), Steve Thompson (Simon Fraser University)
Estimating the Size and Distribution of Networked Populations with Snowball Sampling / Estimer la taille et la distribution de populations en réseau avec l'échantillonnage en boule de neige  
- 16:15 - 16:30 Valéry Dongmo Jiongo (Statistique Canada), Pierre Ngumkeu (Georgia State University)
Bootstrapping Mean Squared Errors of Robust Small Area Estimators / Bootstrap de l'erreur quadratique moyenne d'estimateurs robustes de petits domaines  
- 16:30 - 16:45 Xichen She & Changbao Wu (University of Waterloo)
Analysis of Ordinal Survey Responses with "Don't Know" / Analyse de réponses de sondage ordinales « Je ne sais pas »  
- 16:45 - 17:00 Christiane Laperriere (Statistics Canada)
Data Collection Using a Diary: The Experience of the Survey of Household Spending / Collecte de données à l'aide d'un carnet : l'expérience de l'Enquête sur les dépenses des ménages  

17:00 - 18:00

McCain - Scotiabank

NSERC Discovery Grant Application Workshop



Atelier sur les demandes de subventions à la découverte du CRSNG

17:00 - 18:00 Actuarial Science Section General Meeting Groupe d'actuariat assemblée générale	McCain 2170
17:00 - 18:00 Biostatistics Section General Meeting Groupe de biostatistique assemblée générale	Rowe 1011
17:00 - 18:00 Probability Section General Meeting Groupe de probabilité assemblée générale	McCain 2184
17:00 - 18:00 Statistical Education Section General Meeting Groupe d'éducation en statistique assemblée générale	McCain 2198
17:00 - 18:00 Student Research Presentation Award Committee Comité du prix pour la présentation de recherche étudiante	McCain 2118
17:00 - 18:00 Survey Methods Section General Meeting Groupe de méthodes d'enquête assemblée générale	McCain 2132
18:15 - 19:05 Banquet- Cocktails and Cash Bar Banquet- Bar payant	Cunard Centre
19:05 - 22:00 Banquet- Dinner Service Banquet- Service du repas	Cunard Centre

Wednesday June 17**mercredi 17 juin****8:15 - 16:00****McCain Lobby****Registration
Inscription****8:45 - 9:50 Session 3A-A1 (abstract/résumé - page 212) McCain - Scotiabank****CRM-SSC Prize in Statistics Address****Allocution du récipiendaire du Prix CRM-SSC en statistique**

Special Session / Conférence spéciale

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



8:45 - 9:50 Matías Salibián-Barrera (University of British Columbia)
 Robust Statistics, from Pessimism to Realism / Statistique robuste : du pessimisme au réalisme  

8:45 - 9:50 Session 3A-A2 (abstract/résumé - page 213) McCain - Ondaatje**SSC Impact Award Address****Allocution du récipiendaire du Prix pour impact de la SSC**

Special Session / Conférence spéciale

Chair/Président: Richard Routledge (Simon Fraser University)

Organizer/Responsable: Thérèse Stukel (Institute for Clinical Evaluative Sciences)

8:45 - 9:50 John Petkau (University of British Columbia)
 Collaborative Research Impacting on Clinical Trials in Multiple Sclerosis / Recherche concertée avec des répercussions sur les essais cliniques portant sur la sclérose en plaques  

9:50 - 10:30**Student Union Building - McInness****Coffee Break
Pause-café****10:20 - 11:50 Session 3B-I1 (abstracts/résumés - page 214) McCain - Ondaatje****Actuarial, Financial and Quantitative Risk Management****Gestion du risque actuariel, financier et quantitatif**



Invited Paper Session / Présentations sur invitation



Chair/Président: Jose Garrido (Concordia University)



Organizer/Responsable: Ruodo Wang (University of Waterloo)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe de science actuarielle

10:20 - 10:50 Adam Metzler (Wilfrid Laurier University)
 State Dependent Correlations in Linear Factor Models / Corrélations dépendant de l'état dans les modèles à facteurs linéaires  

10:50 - 11:20 Jiandong Ren (Western University), Qiming He (University of Waterloo)
 Parameter Estimation of Discrete Multivariate Phase-Type Distributions / Estimation des paramètres des distributions de type phase multivariées en temps discret  

11:20 - 11:50 Chengguo Weng (University of Waterloo)
 Optimal Reinsurance Design / Optimisation de la réassurance  









10:20 - 11:50 Session 3B-I2 (abstracts/résumés - page 216) **McCain - Scotiabank**

DASF 90 - A Celebration of the Contributions of Don Fraser, OC

DASF 90 – Une célébration des contributions de Don Fraser, OC

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Nancy Reid (University of Toronto)

- 10:20 - 10:40 Stephen Fienberg (Carnegie Mellon University)
Fraserian Statistical Inference / Inférence statistique fraserienne  
- 10:40 - 11:00 Ana-Maria Staicu, Yichi Zhang & Arnab Maity (North Carolina State University)
Testing for Additivity in Nonparametric Regression / Tests d'additivité en régression non paramétrique  
- 11:00 - 11:20 Grace Yi (University of Waterloo), Yanyuan Ma (University of South Carolina), Donna Spiegelman (Harvard School of Public Health), Raymond Carroll (Texas A&M University)
Functional and Structural Methods with Mixed Measurement Error and Misclassification in Covariates / Méthodes fonctionnelles et structurelles tenant compte de l'erreur de mesure et de classification des covariables  
- 11:20 - 11:40 David Bellhouse (Western University)
Pre-Don in Toronto / Époque pré-Don à Toronto  
- 11:40 - 11:50 Don Fraser, Discussant (University of Toronto)







10:20 - 11:50 Session 3B-I3 (abstracts/résumés - page 218) **Rowe 1011**

Statistical Methods for Genetic Dissection of Complex Traits

Méthodes statistiques pour la dissection génétique de caractères complexes

Invited Paper Session / Présentations sur invitation

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

- 10:20 - 10:50 Jerry Lawless (University of Waterloo), Andriy Derkach & Lei Sun (University of Toronto)
Outcome-Dependent Sampling in Two-Phase Genetic Association Studies / Échantillonnage dépendant des résultats dans des études d'association génétique biphasée  
- 10:50 - 11:20 Lisa Strug & Zeynep Baskurt (University of Toronto)
Evidential Analysis of Genetic Association in General Pedigrees Using Robust Composite Profile Likelihoods / Analyse probante d'association génétique dans des pedigrees quelconques utilisant le profil robuste des vraisemblances composites  
- 11:20 - 11:50 Zeny Feng (University of Guelph)
A Genetic Association Test to Identify Pleiotropic Effect with Multivariate Longitudinal Traits / Un test d'association génétique pour identifier l'effet pléiotropique sur des caractères longitudinaux multivariés  

10:20 - 11:50 Session 3B-I4 (abstracts/résumés - page 220) **Mona Campbell 1108**

Total Survey Error







Erreur d'enquête totale

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Karla Fox (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

- 10:20 - 10:50 Melanie Abeysondera, Marie-Hélène Miville, Karla Fox, Jeannine Claveau, Wisner Jocelyn, Félix Labrecque-Synnott & Remaining members of the mode effect working group (Statistics Canada / Statistique Canada)
Using Total Survey Error Frameworks to Study Mode Effect / L'utilisation des cadres de l'erreur d'enquête totale pour étudier l'effet du mode  
- 10:50 - 11:20 Jenny Thompson, Broderick Oliver & Jennifer Beck (U.S. Census Bureau)
Using Response Rates and Other Quality Metrics to Assess the Effects of the Mixed Collection Modes for Business Surveys: A Case Study / Utilisation des taux de réponse et d'autres mesures de qualité pour évaluer les effets des modes de collecte mixtes pour les enquêtes auprès des entreprises : une étude de cas  
- 11:20 - 11:50 Brady West (University of Michigan), Michael Elliott (University of Michigan-Ann Arbor)
New Methodologies for the Study and Decomposition of Interviewer Effects in Surveys / Nouvelles méthodologies pour l'étude et la décomposition des effets des intervieweurs dans les enquêtes  







10:20 - 11:50 Session 3B-I5 (abstracts/résumés - page 222) **Rowe 1028**

Using TMB to Quickly and Robustly Solve Problems from Marine Ecology

Comment utiliser TMB pour résoudre des problèmes en écologie marine de manière rapide et robuste

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Joanna Mills Flemming (Dalhousie University)

- 10:20 - 10:50 Christoffer Albertsen & Kasper Kristensen (Technical University of Denmark)
Maximum Likelihood Estimation Using Template Model Builder / Estimation du maximum de vraisemblance au moyen du modélisateur Template Model Builder  
- 10:50 - 11:20 Marie Auger-Methe (Dalhousie University), Ian D. Jonsen (Macquarie University), Christoffer Albertsen (Technical University of Denmark), Glenn T. Crossin (Dalhousie University), Andrew E. Derocher (University of Alberta), Katharine Studholme & Joanna Mills Flemming (Dalhousie University)
Tackling the Challenges of Fitting Movement Models to Marine Data / Relever les défis de l'ajustement des modèles de déplacement des données maritimes  
- 11:20 - 11:50 Kim Whoriskey (Dalhousie University), Christoffer Albertsen (Technical University of Denmark), David Yurkowski (University of Windsor), Anders Nielsen (Technical University of Denmark), Joanna Mills Flemming (Dalhousie University)
Fast Fitting of Robust State-Space Models to Animal Movement Data via Template Model Builder / Ajustement rapide des modèles à espace d'états robustes pour des données de déplacement d'animaux par le modélisateur Template Model Builder  





10:20 - 11:35 Session 3B-C1 (abstracts/résumés - page 224) **McCain 2198**





Big Data and Networks

Données volumineuses et réseaux

Contributed Paper Session / Séance de communications libres

Chair/Président: Thierry Duchesne (Université Laval)

- 10:20 - 10:35 Yilei Wu, Yingli Qin, Yilei Wu & Mu Zhu (University of Waterloo)
Linear and Quadratic Discriminant Analyses for High-Dimensional Data / Analyses discriminantes linéaire et quadratique pour des données de grande dimension  
- 10:35 - 10:50 Yingli Qin (University of Waterloo), Weiming Li (Beijing University of Posts and Telecommunications)
Testing the Order of a Population Spectral Distribution for High-Dimensional Data / Vérification de l'ordre d'une distribution spectrale de la population pour des données de grande dimension  

- 11:05 - 11:20 Javed Tomal (University of Toronto Scarborough), William Welch & Ruben H. Zamar (University of British Columbia)
Ranking Homologous Proteins using a Fast Ensemble of Phalanxes Based on Logistic Regression Model / Classement de protéines homologues à l'aide d'un ensemble rapide de phalanges fondé sur un modèle de régression logistique  
- 11:20 - 11:35 Avik Halder (Queen's University and b3Intelligence), Glen Takahara (Queen's University)
Social Media Analytics using Multistate Intensity Models / Analytique des médias sociaux à l'aide de modèles d'intensité multiétats  





10:20 - 11:50 **Session 3B-C2** (abstracts/résumés - page 226) **McCain 2102**

Biostatistics in Action 2

Biostatistique en action 2

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Geneviève Lefebvre (Université du Québec à Montréal)

- 10:20 - 10:35 Joel Dubin & Joon Lee (University of Waterloo)
Challenges in Predicting Outcomes for Patients in Intensive Care Units (ICU's) / Défis lors de la prédiction de résultats pour les patients dans les unités de soins intensifs (USI)  
- 10:35 - 10:50 Brian Younho Hong (University of Ottawa), Ellen McDonald, Gary Foster, Maxime Rhéaume, Nader Khalidi, Maggie Larche & Lehana Thabane (McMaster University)
Exploring the Potential Preventative Effect of Immunomodulators on Digital Ulcer Formation for Scleroderma Patients: A Retrospective Cohort Study / Analyse de l'effet préventif potentiel des immunomodulateurs sur la formation d'ulcères digitaux chez les patients atteints de sclérodémie : étude de cohorte rétrospective  
- 10:50 - 11:05 Bei Jiang (Columbia University and New York University), Eva Petkova (New York University), Thaddeus Tarpey (Wright State University), Todd Ogden (Columbia University)
Modeling Placebo Response using EEG Data / Modélisation de la réponse placebo à l'aide de données EEG  
- 11:05 - 11:20 Narges Nazeri Rad (Lunenfeld-Tanenbaum Research Institute), Yun-Hee Choi (University of Western Ontario), Karen A Kopciuk (University of Calgary), Jerry Lawless (University of Waterloo), Laurent Briollais (Lunenfeld-Tanenbaum Research Institute)
Multistate Models for Screening Evaluation in Family-based Designs / Modèles multi-états pour l'évaluation du dépistage dans des plans de type familiaux  
- 11:20 - 11:35 Yingying Xu & Joel Dubin (University of Waterloo), John Danziger (Harvard Medical School), Joon Lee (University of Waterloo)
Predicting Recovery from Acute Kidney Injury Among Critically Ill Patients / Prédiction du rétablissement d'une insuffisance rénale aiguë chez les patients gravement malades  
- 11:35 - 11:50 Chelsea Ugenti (Wilfrid Laurier University), Colin McFayden (Ontario Ministry of Natural Resources), Douglas Woolford (Wilfrid Laurier University)
Exploring the Impact of Restricted Fire Zones on the Risk of People-caused Forest Fires in Ontario / Explorer l'impact de zones de feu restreintes sur le risque de feux de forêts causés par les gens en Ontario  

10:20 - 11:50 **Session 3B-C3** (abstracts/résumés - page 229) **McCain 2116**

Biostatistics: Causal Inference and Measurement Error

Biostatistique : inférence causale et erreur de mesure

Contributed Paper Session / Séance de communications libres

Chair/Président: Mohammad Ehsanul Karim (University of British Columbia)

- 10:20 - 10:35 Steve Ferreira, Mireille Schnitzer, Amélie Forget & Lucie Blais (Université de Montréal)
Causal Inference Methodology to Evaluate the Time-Varying Effect of Asthma Medication on Pregnancy Duration / Méthodes en inférence causale afin d'évaluer l'effet variant dans le temps de certains médicaments contre l'asthme sur le temps de gestation  
- 10:35 - 10:50 Pablo Gonzalez Ginestet & Robert Platt (McGill University)
Bayesian Adjustment for Confounding in Bayesian Propensity Score Estimation / Ajustement bayésien pour la confusion dans l'estimation bayésienne du score de propension  
- 10:50 - 11:05 Yeying Zhu (University of Waterloo), Debashis Ghosh (Colorado School of Public Health), Donna Coffman (Pennsylvania State University)
Variable Selection in the Potential Outcomes Framework / Sélection de variables dans le cadre de résultats potentiels  
- 11:05 - 11:20 Michelle Xia (Northern Illinois University), Paul Gustafson (University of British Columbia)
Bayesian Inference for Unidirectional Misclassification in Regression Models / Inférence bayésienne pour la classification erronée unidirectionnelle dans les modèles de régression  
- 11:20 - 11:35 Khokan Sikdar & Hude Quan (University of Calgary)
Estimating the Prevalence of Hypertension from Administrative Data in the Absence of a Gold Standard / Estimer la prévalence de l'hypertension à partir de données administratives en l'absence d'un étalon de référence  
- 11:35 - 11:50 Rojiar Haddadian, Yuliya Martsynyuk & Liqun Wang (University of Manitoba)
Simulation-based Estimation in Generalized Linear Models with a Binary Response Variable and Mismeasured Covariates / Estimation basée sur la simulation dans les modèles linéaires généralisés avec une variable réponse binaire et des covariables mesurées incorrectement  











10:20 - 11:50 Session 3B-C4



(abstracts/résumés - page 232)

McCain 2162**Time Series and Longitudinal Analysis****Séries chronologiques et analyse longitudinale**

Contributed Paper Session / Séance de communications libres

Chair/Présidente: Eleanor Pullenayegum (University of Toronto & Hospital for Sick Children)

- 10:20 - 10:35 Joshua Pohlkamp-Hartt & David L Riegert (Queen's University)
Tackling Data Synthesis Using a Multitaper Spectrum Estimation Technique / S'attaquer à la synthèse de données à l'aide d'une technique d'estimation spectrale à fenêtres de pondération multiples  
- 10:35 - 10:50 Nadarajah Tharshanna, Asokan Mulayath Variyath & J Concepción Loredó-Osti (Memorial University of NL)
Variable Selection for Longitudinal Data Analysis / Sélection de variables pour l'analyse de données longitudinales  
- 10:50 - 11:05 Shahedul Khan (University of Saskatchewan)
Modeling a Mixture of Linear and Change-point Trajectories for Longitudinal Time-Series Data / Modélisation d'un mélange de trajectoires linéaires et à point de changement pour des données longitudinales de séries chronologiques  
- 11:05 - 11:20 Kexin Ji & Joel Dubin (University of Waterloo)
A Bivariate Semiparametric Stochastic Mixed Model / Un modèle stochastique mixte semiparamétrique bivarié  
- 11:20 - 11:35 Wei Qian & J. N. K. Rao (Carleton University), Laura Dumitrescu (National School for Statistics and Information Analysis)
Quadratic Inference Functions for Longitudinal Survey Data / Fonctions d'inférence quadratiques pour les données d'enquête longitudinales  

11:35 - 11:50 Yukun Zhang & Tolulope Sajobi (University of Calgary), Lisa Lix (University of Manitoba)
 On Covariance Misspecification in Repeated Measures Discriminant Analysis / À propos de
 la mauvaise spécification de la covariance dans l'analyse discriminante de mesures répé-
 tées  

12:00 - 13:30 McCain 2021
Accreditation Committee
Comité d'accréditation

12:00 - 13:30 McCain 2162
Actuarial Science Section Executive Committee 2
Comité exécutif du groupe d'actuariat 2

12:00 - 13:30 McCain 2176
Biostatistics Section Executive Committee 2
Comité exécutif du groupe de biostatistique 2

12:00 - 13:30 McCain 2198
Business and Industrial Statistics Section General Meeting
Groupe de statistique industrielle et de gestion assemblée générale

12:00 - 13:30 McCain 2116
Probability Section Executive Committee 2
Comité exécutif du groupe de probabilité 2

12:00 - 13:30 McCain - Scotiabank
Program Committee
Comité du programme

12:00 - 13:30 McCain 2102
Statistical Education Section Executive Committee 2
Comité exécutif du groupe d'éducation en statistique 2

12:00 - 13:30 Rowe 1011
Student and Recent Graduate Committee
Comité des étudiants et diplômés récents

12:00 - 13:30 Rowe 1028
Treasurers
Trésoriers

13:30 - 14:30 Session 3C-A1 (abstract/résumé - page 235) **Rowe 1028**



Isobel Loutit Invited Address**Allocution invitée Isobel Loutit**

Special Session / Conférence spéciale

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

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

13:30 - 14:30 Paolo Giudici (University of Pavia)
Twitter Data Analysis in Business and Industry / Analyse des données de Twitter dans le commerce et l'industrie  

13:30 - 14:30 Session 3C-A2 (abstract/résumé - page 236) **Rowe 1011**



Survey Methods Section Presidential Invited Address**Allocution de l'invité du président du Groupe des méthodes d'enquête**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Wesley Yung (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête



13:30 - 14:30 Mary Thompson (University of Waterloo)
Big Data, Official Statistics and Survey Science / Données volumineuses, statistiques officielles et science des bases de données  

13:30 - 14:30 Session 3C-A3 (abstract/résumé - page 237) **McCain - Ondaatje**

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

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Richard Lockhart (Simon Fraser University)

13:30 - 14:30 Douglas Schaubel (University of Michigan), Hui Zhang (U.S. Food and Drug Administration), John D. Kalbfleisch & Xu Shu (University of Michigan)
Semiparametric Methods for Survival Analysis of Case-Control Data Subject to Dependent Censoring / Méthodes semi-paramétriques pour les analyses de survie de données cas-témoins sujettes à la censure dépendante  



13:30 - 14:30 Session 3C-A4 (abstract/résumé - page 238) **McCain - Scotiabank**

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

Special Session / Conférence spéciale

Chair/Président: Alexandre Leblanc (University of Manitoba)

Organizer/Responsable: Éric Marchand (Université de Sherbrooke)

13:30 - 14:30 Ying Yan (University of North Carolina)
Statistical Methods on Survival Data with Measurement Error / Méthodes statistiques sur des données de survie avec erreur de mesure  

14:30 - 15:00 Student Union Building - McInness
Coffee Break
Pause-café

15:00 - 16:30 **Session 3D-I1** (abstracts/résumés - page 239) **Rowe 1011**

Applications of Mixed-Effect Models in the Health Sciences

Applications des modèles à effets mixtes en sciences de la santé







Invited Paper Session / Présentations sur invitation

Chair/Président: Patrick Brown (University of Toronto)

Organizer/Responsable: Mahmoud Torabi (University of Manitoba)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

- 15:00 - 15:30 Sunil Rao (University of Miami), Jiming Jiang (University of California, Davis), Jie Fan (University of Miami), Thuan Nguyen (Oregon Health and Science University)
Classified Mixed Model Prediction / Préviation classifiée à l'aide d'un modèle mixte  
- 15:30 - 16:00 Cindy Feng (University of Saskatchewan)
Zero Modified Models for Modeling Length of Hospital Stay for patients with Ischaemic Heart Disease / Modèles modifiés à zéro pour la modélisation de la durée des séjours hospitaliers des patients atteints de cardiopathie ischémique  
- 16:00 - 16:30 Mahmoud Torabi (University of Manitoba)
Multivariate Mixture Spatial Generalized Linear Mixed Models / Modèles de mélange spatiaux multivariés de la classe des modèles linéaires mixtes généralisés  

15:00 - 16:30 **Session 3D-I2** (abstracts/résumés - page 241) **McCain 2198**

Limit Theorems With Applications







Théorèmes limites et applications

Invited Paper Session / Présentations sur invitation

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

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

- 15:00 - 15:30 Miklos Csorgo (Carleton University), Endre Csaki (Alfred Renyi Institute of Mathematics), Antonia Foldes (College of Staten Island, CUNY), Pal Revesz (Technische Universitat Wien)
Two-dimensional Anisotropic Random Walks: Fixed versus Random Column Configurations for Transport Phenomena / Marches aléatoires anisotropes à deux dimensions : configurations fixes et aléatoires des colonnes pour les phénomènes de transport  
- 15:30 - 16:00 Deli Li (Lakehead University), Han-Ying Liang (Tongji University, China)
A Comparison Theorem for the Law of Large Numbers / Théorème de comparaison pour la loi des grands nombres  
- 16:00 - 16:30 Priscilla Greenwood & Lawrence Ward (University of British Columbia)
Stochastic Neural Dynamics / Modèle stochastique de dynamique neuronale  



15:00 - 16:30 **Session 3D-I3** (abstracts/résumés - page 243) **McCain - Scotiabank**





Maritime Ecology

Écologie marine

Invited Paper Session / Présentations sur invitation

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

- 15:00 - 15:30 Chris Field, Hong Gu & Ximing Xu (Dalhousie University)
Modelling Microbial Multinomial Time or Space Series / Modélisation de séries multinomiales microbiennes chronologiques ou spatiales  





- 15:30 - 16:00 Michael Dowd (Dalhousie University), Paul Mattern (University of California), Jon Briggs (University of Auckland), Katja Fennel (Dalhousie University), Renate Meyer (University of Auckland)
Data Assimilation for Ocean Biology / Assimilation des données pour la biologie océanique  
- 16:00 - 16:30 Keith Thompson & Yihao Yin (Dalhousie University), Emmanuel Devred (Université Laval), Florence Birol (Laboratoire d'Etudes en Géophysique et Océanographie Spatiales)
Satellite Sensing of the Marine Environment and the Issue of Aliasing / La détection par satellite du milieu marin et le problème de repliement de spectre  

15:00 - 16:30 Session 3D-I4 (abstracts/résumés - page 245) **McCain - Ondaatje**

R Packages for Teaching

Utiliser R pour l'enseignement

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







- 15:00 - 15:35 John Fox (McMaster University)
Teaching Social Statistics Courses with R and the R Commander / Enseignement de cours de statistique sociale avec R et R Commander  
- 15:35 - 16:10 Duncan Murdoch (Western University)
R in Teaching and Testing / R dans l'enseignement et les examens  
- 16:10 - 16:30 Discussion

15:00 - 16:30 Session 3D-I5 (abstracts/résumés - page 246) **Rowe 1028**

Statistical Application and Modeling for Functional Data Analysis

Application et modélisation statistiques pour l'analyse de données fonctionnelles

Invited Paper Session / Présentations sur invitation
Organizer and Chair / Responsable et président: Haocheng Li (University of Calgary)

- 15:00 - 15:30 James Ramsay (McGill University)
Data Smoothing with Tuneable Penalties / Lissage de données avec pénalités ajustables  
- 15:30 - 16:00 Mark Koudstaal & Fang Yao (University of Toronto)
Simultaneous White Noise Models and Optimal Recovery of Functional Data / Modèles de bruit de fond simultanés et récupération optimale de données fonctionnelles  
- 16:00 - 16:30 Jason Nielsen & Nikolina Beg (Carleton University)
Model-Based Penalty Smoothing for Panel Data Under a Markov Assumption / Lissage avec pénalité basée sur un modèle pour données de panel dans une hypothèse de Markov  

16:45 - 18:00 **McCain 2021**

Board of Directors 2

Conseil d'administration 2

18:00 - 19:00 **Risley Hall, Rooms 1 & 2**

Executive Committee 2

Comité exécutif 2

Abstracts • Résumés

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

Organizer and Chair / Responsable et président: John Petkau (University of British Columbia)

Room/Salle: McCain - Ondaatje

Abstract/Résumé

XIAO-LI MENG (Harvard University)

[Monday June 15 / lundi 15 juin, 8:45]

Lifelines for Statistics (and Statisticians): Generalists, Specialists, or Your-lists?

Bouées de sauvetage pour les statistiques (et les statisticiens): généralistes, spécialistes ou vos-listes?

Statisticians are being desired and feared (Meng, August 2009 and February 2010, *The American Statistician*). The demand for statistics, from education to research, far exceeds the supply, urging us to strategize. We have a severe shortage of well-trained education generalists who can deliver effective "preventive medicine" by being the quantitative trainers of future scientists, policy makers, educators, etc, that is, by raising the general level of statistical competency. We also need many more deep-and-broad statisticians to combat complex problems requiring "medical specialists" beyond "family physicians", namely, problems about which other professions must consult us even just for their peace of minds. This talk invites second opinions, aiming to provide an accurate diagnosis of and effective prescription for our current discomfort.

Les statisticiens sont recherchés et craints (Meng, août 2009 et février 2010, *The American Statistician*). Le besoin des statistiques, de l'éducation à la recherche, dépasse largement l'offre, nous pressant d'établir une stratégie. Nous faisons face à une grave pénurie d'éducateurs généralistes qualifiés qui peuvent offrir une « médecine préventive » efficace en étant les formateurs quantitatifs des futurs scientifiques, décideurs politiques, éducateurs, etc, c'est-à-dire en élevant le niveau général de compétence statistique. Nous avons aussi besoin de plus de statisticiens qui ont des connaissances larges et approfondies pour s'attaquer à des problèmes complexes qui nécessitent des « spécialistes médicaux » au-delà des « médecins de famille », à savoir des problèmes à propos desquels d'autres professions doivent nous consulter, même si ce n'est que pour leur tranquillité d'esprit. Cet exposé invite à partager les opinions, visant ainsi à fournir un diagnostic précis de notre inconfort actuel et un remède efficace.

1B-I1: Combinatorial Probability Probabilité combinatoire

Organizer and Chair / Responsable et président: Neal Madras (York University)

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

Room/Salle: Rowe 1011

Abstracts/Résumés

JANE GAO (University of Waterloo), **MIKE MOLLOY** (University of Toronto)

[Monday June 15 / lundi 15 juin, 10:20]

From Computer Science to Statistical Physics: Solution Clustering in Random CSPs

De l'informatique à la physique statistique : solution de regroupement en cas de problèmes de satisfaction des contraintes aléatoires

While constraint satisfaction problems (CSPs) are the most classical problems in computer science, a much more recent approach by statistical physicists gave new insights into these problems. Using the cavity method, they predicted that the solution spaces of many random CSPs have a clustering phenomenon. Moreover, when the density of a random CSP increases, several phase transitions, other than clustering, will be encountered. Some of these predictions have now been rigorously proven by mathematicians. This clustering phenomenon is generic for many CSPs; it gives insight into the existence of an "algorithmic barrier" for random CSPs. In this talk, I will review research in this field and report some new results on solution clustering of random k-XORSAT instances.

Alors que les problèmes de satisfaction des contraintes sont les problèmes les plus classiques en informatique, une approche beaucoup plus récente proposée par des physiciens statisticiens a donné une nouvelle perspective à ces problèmes. Par l'utilisation de la méthode de la cavité, ils prédisent que les espaces de solutions de plusieurs problèmes de satisfaction des contraintes aléatoires ont un phénomène de regroupement. De plus, lorsque la densité de ces problèmes augmente, plusieurs transitions de phases, autres que des regroupements, surviendront. Aujourd'hui, certaines de ces prédictions ont été rigoureusement prouvées par des mathématiciens. Ce phénomène de regroupement est générique pour de nombreux problèmes de satisfaction des contraintes et donne un aperçu de la « barrière algorithmique » pour les problèmes de satisfaction des contraintes aléatoires. Dans cet exposé, j'analyserai la recherche dans ce domaine et communiquerai de nouveaux résultats sur les solutions de regroupement d'exemples de problèmes aléatoires k-XORSAT.

JEANNETTE JANSSEN & HODA CHUANGPISHIT (Dalhousie University), **MAHYA GHANDEHARI** (University of Delaware), **NAUZER KALYANIWALLA** (Dalhousie University)

[Monday June 15 / lundi 15 juin, 10:50]

Linear Embeddings of Large Graphs

Plongements linéaires de grands graphiques

Many large networks have a natural embedding in a feature space, where links are more likely to occur between nodes that are in close proximity. This leads to the question: can we recognize whether a

De nombreux grands réseaux ont un plongement naturel dans un espace des attributs, dans lequel les nœuds qui sont tout près sont plus susceptibles de se relier. Cela amène à la question suivante : pouvons-nous reconnaître si un graphe

1B-I1: Combinatorial Probability Probabilité combinatoire

given graph has such a spatial embedding? This question can be approached from many points of view. I will show how the theory of dense graph limits can be used to develop a graph parameter that measure to what extent a given graph is similar in structure to a graph produced by a random graph process with a linear embedding.

donné a un tel plongement spatial ? Cette question peut être abordée de plusieurs points de vue. Je vais montrer comment une théorie des limites des graphes denses peut être utilisée pour développer un paramètre de graphe qui mesure l'étendue de la similitude d'un graphe donné en structure par rapport à un graphe produit au moyen d'un processus de graphe aléatoire avec un plongement linéaire.

LERNA PEHLIVAN (University of Washington), **NEAL MADRAS** (York University)

[Monday June 15 / lundi 15 juin, 11:20]

Random 312 Avoiding Permutations

Permutations évitant la séquence aléatoire 312

A permutation of $1, \dots, N$ is said to avoid 312 pattern if there is no subsequence of three elements of this permutation that appears at the same relative order as 312. Monte Carlo experiments reveal some features of random 312 avoiding permutations. In light of these experiments we determine some probabilities explicitly and we obtain asymptotic approximations to these probabilities for large N .

On dit qu'une permutation de $1, \dots, N$ évite le motif 312 si aucune sous-suite des trois éléments de cette permutation n'apparaît dans le même ordre relatif que 312. Les expériences de Monte-Carlo révèlent des caractéristiques de permutations aléatoires d'évitement de 312. À la lumière de ces expériences, nous déterminons des probabilités de façon explicite et nous obtenons des approximations asymptotiques dans le contexte de ces probabilités pour les grands N .

1B-I2: Getting to Know our Students in the Electronic Age: Did the Dog Really Eat the Homework?
Apprendre à connaître nos étudiants à l'ère électronique : le chien a-t-il vraiment mangé leurs devoirs?

Chair/Président: Gary Sneddon (Mount Saint Vincent University)
Organizer/Responsable: Michele Millar (Mount Saint Vincent University)

Session sponsored by the Statistics Education Section
Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: Rowe 1028

Abstracts/Résumés

MARJORIE BOND (Monmouth College)

[Monday June 15 / lundi 15 juin, 10:20]

A Qualitative Survey's Journey: From Conception to Publication

Le parcours d'un sondage qualitatif: de la conception à la publication

Due to my work with the Survey of Attitudes Toward Statistics (SATS), I needed to learn about students' perception of the statistics prior to taking introductory statistics classes since the SATS uses the word "statistics" throughout the survey. The journey to obtain this knowledge involved focus groups and pilot surveys before finally creating a short answer survey which is administered pre- and post- course. Additionally, the results from the survey are qualitative responses which needed a coding scheme in order to analyze the data. Eventually, the results of the survey were used in a published article. This presentation will discuss a qualitative survey's journey from conception to publication.

En raison de mon travail avec l'Enquête des comportements à l'égard de la statistique (SATS), j'avais besoin d'en connaître plus sur la perception des étudiants face aux statistiques avant de prendre un cours d'introduction aux statistiques, étant donné que le SATS utilise le mot "statistiques" tout au long de l'enquête. Le parcours pour obtenir cette connaissance a comporté des groupes de consultation ainsi que des enquêtes pilotes, avant de finalement créer un sondage à réponses courtes qui fut administré avant et après le cours. De plus, les résultats du sondage étaient des réponses qualitatives qui ont eu besoin d'un système de codes pour arriver à analyser les données. Éventuellement, les résultats du sondage ont été utilisés dans un article publié. Cette présentation discutera du parcours d'un sondage qualitatif de la conception à la publication.

DANIELLE COX (Acadia University)

[Monday June 15 / lundi 15 juin, 10:45]

The Inside Scoop on Students - Experiences in the Classroom

Ce qu'il faut savoir sur les étudiants - expériences dans la salle de classe

Statistics is a required course for almost every university student. It is important to become familiar enough with our students that we discover their mathematical backgrounds and program goals. This will allow us to enhance their statistics class experience. In this presentation we will discuss

Un cours de statistique est obligatoire pour presque tous les étudiants universitaires. Il est important d'apprendre à connaître suffisamment nos étudiants pour découvrir leur base en mathématiques et les objectifs de leur programme. Ceci nous permettra d'améliorer leur expérience dans leur cours de statistique. Dans cette présentation nous discu-

1B-12: Getting to Know our Students in the Electronic Age: Did the Dog Really Eat the Homework? Apprendre à connaître nos étudiants à l'ère électronique : le chien a-t-il vraiment mangé leurs devoirs?

techniques to get to know our students and explore how we can use this knowledge to actively engage them in the learning process, whether the class is held on campus or in a virtual setting.

terons de techniques pour apprendre à connaître nos étudiants et nous explorerons comment utiliser cette connaissance pour les engager activement dans le processus d'apprentissage, que le cours se déroule sur un campus ou dans un cadre virtuel.

MICHELLE MILLAR (Mount Saint Vincent University)

[Monday June 15 / lundi 15 juin, 11:10]

Designing and Analyzing Student Surveys: Pitfalls and Rewards

Concevoir et analyser des enquêtes menées auprès d'étudiants : embûches et récompenses

Almost all statistical educators will use surveys at some point in their career. We will discuss surveys for scholarly research and less formal surveys. For example: surveys to collect data for analysis during the course; the five line exit survey to gauge student's understanding of the class topic; VARK Survey to discover student's learning styles; surveys designed by students; and formal/informal student evaluations of the course. Last but not least, surveys for scholarly research, which opens a whole new can of worms! How to reliably evaluate latent variables; Likert scores -- are they really trustworthy; qualitative methods vs quantitative methods . . .

Presque tous les enseignants en statistique vont utiliser des enquêtes à un moment au cours de leur carrière. Nous discuterons d'enquêtes pour la recherche universitaire ainsi que d'enquêtes moins formelles. Par exemple : des études pour recueillir des données aux fins d'analyse durant le cours ; le sondage de départ de cinq lignes pour mesurer la compréhension du sujet du cours par les étudiants ; le questionnaire VARK pour découvrir les styles d'apprentissage des étudiants ; des études conçues par les étudiants ; et des évaluations formelles/informelles du cours par les étudiants. Dernier exemple mais non le moindre, les enquêtes pour la recherche universitaire, ce qui ouvre une toute nouvelle boîte de Pandore ! Comment évaluer de façon fiable des variables latentes ; les résultats de basés sur l'échelle de Likert -- sont-ils vraiment dignes de confiance ; méthodes qualitatives vs méthodes quantitatives. . .

1B-I3: Recent Advances in Predictive Methods in Actuarial Science Avancées récentes dans les modèles prédictifs en science actuarielle

Organizer and Chair / Responsable et président: Manuel Morales (Université de Montréal)

Session sponsored by the Actuarial Science Section

Séance parrainée par le Groupe de science actuarielle

Room/Salle: Mona Campbell 1108

Abstracts/Résumés

CARY TSAI (Simon Fraser University)

[Monday June 15 / lundi 15 juin, 10:20]

A Linear Relational Approach to Modelling Mortality Rates

Approche relationnelle linéaire de la modélisation des taux de mortalité

In this talk, we first briefly introduce two well-known mortality models, the Lee-Carter model and the CBD model. Then a linear relational approach based on simple linear regression to modelling mortality rates is proposed, which linearly relates one mortality sequence to the other of equal length with the evidence of observations from empirical mortality data. Some variations and applications are given, and forecasting performances among models are also compared with numerical illustrations.

Nous commençons par une brève présentation de deux modèles de mortalité bien connus, le modèle Lee-Carter et le modèle CBD. Nous proposons ensuite une approche relationnelle linéaire fondée sur la régression linéaire simple pour modéliser les taux de mortalité, dans laquelle une séquence de mortalité est linéairement associée à une autre de longueur égale avec pour preuve les observations tirées des données de mortalité empiriques. Nous présentons des variations et applications, puis comparons la qualité de la prévision des modèles à l'aide d'illustrations numériques.

MACIEJ AUGUSTYNIAK (Université de Montréal), **MATHIEU BOUDREAU** (Université du Québec à Montréal), **MANUEL MORALES** (Université de Montréal)

[Monday June 15 / lundi 15 juin, 10:50]

Estimating the Markov-Switching GARCH Model with a Deterministic Particle Filter

Estimation du modèle GARCH à changement de régimes avec un filtre particulaire déterministe

The Markov-switching GARCH model allows for a GARCH structure with time-varying parameters. This flexibility is unfortunately undermined by a path dependence problem which complicates the parameter estimation process. This problem led to the development of computationally intensive estimation methods and to simpler techniques based on an approximation of the model, known as collapsing procedures. This article develops an original algorithm to conduct maximum likelihood inference in the Markov-switching GARCH model, generalizing and improving previously proposed collapsing approaches. A new relationship between

Le modèle GARCH à changement de régimes donne lieu à un problème de dépendance au chemin emprunté qui complique l'estimation du modèle. Cette difficulté a mené à l'introduction de méthodes computationnelles intensives et à des techniques plus simples basées sur une approximation du modèle, connues sous le nom de procédures d'affaissement (*collapsing procedures*). Une méthodologie originale est développée permettant d'estimer le modèle GARCH à changement de régimes par le maximum de vraisemblance, généralisant et améliorant les approches dites par affaissement développées dans les deux dernières décennies. Cette généralisation permet d'établir un lien méthodologique entre ces approches et le filtre particulaire et de justifier la validité

1B-I3: Recent Advances in Predictive Methods in Actuarial Science Avancées récentes dans les modèles prédictifs en science actuarielle

particle filtering and collapsing procedures is established which reveals that this algorithm corresponds to a deterministic particle filter. Simulation and empirical studies show that the proposed method allows for a fast and accurate estimation of the model.

de l'approche dite par affaissement pour estimer le modèle GARCH à changement de régimes.

ANNE MACKAY (ETH Zurich), **MARIO WÜTHRICH** (ETH Zürich)

[Monday June 15 / lundi 15 juin, 11:20]

Predicting Best-Estimate Yield Curves in Incomplete Bond Markets

Meilleure évaluation des courbes de rendement sur les marchés obligataires incomplets

The concept of best-estimate, prescribed by regulators to value insurance liabilities for accounting and solvency purposes, has recently been discussed extensively in the insurance industry and related academic literature. Happ, Merz and Wüthrich (2014) define best-estimates using orthogonal projections of a claim on the space of replicable payoffs. In this paper, we apply the concept of best-estimates to long-maturity claims in a market with reinvestment risk. We assume that a limited number of short-maturity bonds are traded, and derive the best-estimate price of bonds with longer maturity, thus obtaining a best-estimate yield curve. We derive expressions for the price of the long-term bond under the multifactor Vasicek model.

Le concept de la meilleure évaluation, stipulé par les organismes de réglementation pour évaluer les passifs d'assurance à des fins de comptabilité et de solvabilité, a récemment fait l'objet de discussions approfondies dans l'industrie et la littérature de l'assurance. Happ, Merz et Wüthrich (2014) définissent des meilleures évaluations en opérant la projection orthogonale d'une réclamation sur l'espace des indemnités reproductibles. Dans cet article, nous appliquons ce concept aux créances à longue échéance sur un marché avec risque de réinvestissement. Nous assumons qu'un petit nombre d'obligations à courte échéance sont négociés et dérivons la meilleure évaluation du prix des obligations à plus longue échéance, obtenant ainsi la meilleure évaluation de la courbe de rendement. Nous en dérivons des expressions pour le cours de l'obligation à longue échéance dans le modèle de Vasicek multifactoriel.

1B-14: Recursive Partitioning and Regression Trees Partitionnement récursif et arbres de régression

Organizer and Chair / Responsable et présidente: Liqun Diao (University of Waterloo)

Room/Salle: McCain - Scotiabank

Abstracts/Résumés

MICHAEL LEBLANC (Fred Hutchinson Cancer Research Center)

[Monday June 15 / lundi 15 juin, 10:20]

Refinement Rules for Prognosis and Enhanced Treatment Effect Subgroups

Règles de raffinement pour les pronostics et pour les effets de traitements améliorés dans des sous-groupes

Recursive partitioning methods including regression trees have been widely used to describe subgroups of cancer patients with differing prognosis. We describe an alternative technique based on adaptive refinement of subgroups that allows one to control the class of subset decision rules and the fraction of patients identified by the rules.

The proposed modeling strategy characterizes subgroups through a smooth basis function representation of subgroups. The strategy allows user to control the expected number of patients in the subgroup as well as the anticipated survival or treatment effect in the targeted group. We present examples from cancer clinical trials conducted by SWOG, a US national clinical trials organization.

Les méthodes de partition récursive, y compris les arbres de régression, ont largement été utilisées pour décrire des sous-groupes de patients atteints du cancer avec des pronostics divergents. Nous décrivons une autre technique basée sur le raffinement adaptatif des sous-groupes qui permet de contrôler la classe d'un sous-ensemble de règles de décision et la fraction des patients identifiés par les règles.

La stratégie de modélisation proposée caractérise des sous-groupes au moyen d'une fonction de base lisse de représentation de sous-groupes. Cette stratégie permet à l'utilisateur de contrôler le nombre prévu de patients dans un sous-groupe ainsi que l'effet de survie ou de traitement prévu dans le groupe ciblé. Nous présentons des exemples d'essais cliniques sur le cancer menés par SWOG, une organisation américaine d'essais cliniques à l'échelle fédérale.

ROBERT STRAWDERMAN (University of Rochester), **JON STEINGRIMSSON** (Cornell University), **LIQUN DIAO** (University of Waterloo)

[Monday June 15 / lundi 15 juin, 10:50]

Doubly Robust Survival Trees

Arbres de survie doublement robustes

Survival trees use recursive partitioning to separate patients into distinct risk groups. This work extends the IPCW-loss-based recursive partitioning procedures of Molinaro et al. (2004), using semi-parametric efficiency theory for missing data applied to right-censored outcomes to motivate "doubly robust" loss-based recursive partitioning procedures that make better use of available information. It is further shown how these computations lead to equivalent "full data" recursive partitioning procedures that utilize imputed survival times, allowing one to build survival trees using existing

Les arbres de survie utilisent des partitionnements récursifs pour séparer les patients dans des groupes de risques distincts. Ces travaux étendent les procédures de partitionnement récursif en fonction de la perte basée sur la probabilité inverse de pondération de la censure de Molinaro et coll. (2004). La théorie d'efficacité semi-paramétrique pour les données manquantes appliquée aux réponses censurées à droite pour motive les procédures de partitionnement récursif « doublement robustes ». L'information disponible peut ainsi être mieux utilisée. On montre comment ces calculs mènent à des procédures équivalentes de partitionnement récursif de « données complètes » au moyen de temps de

1B-14: Recursive Partitioning and Regression Trees Partitionnement récursif et arbres de régression

methods for uncensored data with minimal modification. The performance of the doubly robust survival trees, implemented as a generalization of CART through modification of the `rpart` package, is evaluated through simulation studies and through analyzing data on death from myocardial infarction.

survie imputés, ce qui permet de créer des arbres de survie avec les méthodes existantes pour des données non censurées modifiées le moins possible. Le rendement des arbres de survie doublement robustes, employés comme une généralisation des arbres de classification et de régression (CART) par la modification du paquet `rpart`, est évalué par des études de simulation et par l'analyse de données sur les décès par infarctus du myocarde.

THOMAS LOUGHIN & ANDREW HENREY (Simon Fraser University)

[Monday June 15 / lundi 15 juin, 11:20]

Heteroscedastic Regression Trees for Joint Modeling of Means and Variances

Arbres de régression hétéroscédastique pour une modélisation conjointe des moyennes et des variances

Standard regression trees as proposed by Breiman et al. (1984) are not robust in the presence of heteroscedasticity. We develop an alternative split-selection criterion that has the capacity to account for local variance in the regression in an unstructured manner. The criterion is based on a normal likelihood where both the mean and variance are allowed to change arbitrarily in the two child nodes. We use AIC to decide between the mean-only (1-parameter) and the mean+variance (2-parameter) splits at each candidate location and to prune the fully split tree. We show that the new approach performs better for mean estimation than does the standard mean-only split when variances are not constant.

Les arbres de régression standards comme ceux proposés par Breiman et coll. (1984) ne sont pas robustes en présence d'hétéroscédasticité. Nous développons un autre critère de sélection pour la division, afin de tenir compte de variances locales dans la régression d'une façon non structurée. Ce critère est basé sur une vraisemblance normale dans laquelle on permet à la moyenne et à la variance de changer arbitrairement dans les deux nœuds enfants. Nous utilisons le critère d'information d'Akaike pour décider entre les divisions ne faisant qu'intervenir la moyenne (1 paramètre) ou la moyenne + la variance (2 paramètres) à chaque endroit possible et pour élaguer l'arbre complètement fractionné. Nous montrerons que cette nouvelle approche fonctionne mieux pour l'estimation de la moyenne que celle basée sur la seule division de la moyenne lorsque les variances ne sont pas constantes.

1B-15: Statistics in Biosciences: Statistical Methods for Big Data from Health Science Statistique en biosciences : méthodes statistique pour les données volumineuses en sciences de la santé

Organizer and Chair / Responsable et présidente: Grace Yi (University of Waterloo)

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

XIHONG LIN, IAN BARNETT & RAJARSHI MURKERJEE (Harvard University)

[Monday June 15 / lundi 15 juin, 10:20]

The Generalized Higher Criticism for Testing SNP-set Effects in Genetic Association Studies

Haute critique généralisée pour tester des effets d'ensemble de SNP dans les études d'association génétique

We propose the Generalized Higher Criticism (GHC) to test for the association between a SNP set, e.g., a gene or a network, and a disease outcome in the presence of sparse alternative. The proposed GHC overcomes the limitations of the HC by allowing for arbitrary correlation structures among the SNPs in a SNP-set, while performing accurate analytic p-value calculations for any finite number of SNPs in the SNP-set. We obtain the detection boundary of the GHC test. We compared empirically using simulations the power of the GHC method with existing SNP-set tests over a range of genetic regions with varied correlation structures and signal sparsity. We apply the proposed methods to analyze the CGEM breast cancer genome-wide association study.

On propose une haute critique généralisée (HCG) pour tester l'association entre un ensemble de SNP, p. ex., un réseau de gènes, et une maladie en présence d'une alternative clairsemée. La HCG proposée surmonte les limites de la HC en permettant des structures de corrélation arbitraires parmi les SNP d'un ensemble, tout en faisant des calculs analytiques précis de valeurs p pour chaque nombre fini de SNP dans un ensemble. On obtient la borne de détection du test HCG. On compare à l'aide de simulations la puissance de la méthode HCG avec des tests d'ensemble de SNP existants sur différentes régions génétiques ayant des structures de corrélation variées et un signal clairsemé. On applique les méthodes proposées pour analyser les données de l'étude CGEM d'association pangénomique sur le cancer du sein.

HONGZHE LI (University of Pennsylvania)

[Monday June 15 / lundi 15 juin, 10:50]

Sparse Simultaneous Signal Detection and Its Applications in Genomics

Détection de signaux rares simultanés et ses applications en génomique

The increasing availability of large-scale genomic data has made possible an integrative approach to studying disease. Such research seeks to uncover disease mechanisms by combining multiple types of genomic information, which may be collected on multiple sets of patients. I focus on a study that integrates GWAS and eQTL data collected from two different sets of subjects to find transcripts potentially functionally relevant to human heart failure. I formalize a model that defines important transcripts as those whose expression levels are associated with SNPs that are simultaneously associated with

La disponibilité croissante de données à grande échelle en génomique a rendu possible une approche intégrée pour l'étude d'une maladie. Une telle recherche vise à découvrir les mécanismes de la maladie en combinant plusieurs types d'information génomique qui pourront être recueillis sur différents ensembles de patients. Je me concentre sur une étude qui intègre les données GWAS et eQTL recueillies à partir de deux ensembles différents de sujets afin de trouver des relevés potentiellement fonctionnellement pertinents à l'insuffisance cardiaque humaine. Je formalise un modèle qui définit les relevés importants comme étant ceux dont les niveaux d'expression associés à des SNP sont simultanément asso-

1B-15: Statistics in Biosciences: Statistical Methods for Big Data from Health Science Statistique en biosciences : méthodes statistique pour les données volumineuses en sciences de la santé

disease and propose a new procedure to test for detecting simultaneous signals. I show that the test statistic is asymptotically optimal under certain conditions. I present several applications and extensions.

ciés à la maladie et je propose une nouvelle procédure pour tester la détection des signaux simultanés. Je montre que la statistique du test est asymptotiquement optimale sous certaines conditions. Je présente plusieurs applications et extensions.

CHARMAINE DEAN & MARK WOLTERS (Western University & Fudan University)

[Monday June 15 / lundi 15 juin, 11:20]

Parameter Estimation in Autologistic Regression Models for Detection of Smoke in Satellite Images

Estimation des paramètres dans les modèles autologistiques de régression pour la détection de fumée dans les images satellites

Smoke from forest fires is a health hazard that is difficult to study through direct measurement. Images from earth-orbiting satellites provide a potentially valuable data source to catalogue smoke events over space and time. We are developing classifiers to segment satellite images into smoke and nonsmoke regions using the autologistic regression model, a Markov random field model with logistic regression as a special case. The large size of the images (both in terms of pixel count and number of image planes) introduces a variety of computational challenges when using this model. The talk will focus on parameter estimation, comparing alternative approaches and discussing how the goal of the study--predictive accuracy or parameter interpretation--might influence the choice of estimation method.

La fumée des incendies de forêt est un danger pour la santé qui est difficile à étudier par mesure directe. Les images provenant de satellites en orbite terrestre constituent une source de données potentiellement utiles pour faire l'inventaire d'événements de fumée dans l'espace et dans le temps. Nous développons des classificateurs permettant de segmenter des images satellites en régions de fumée et non fumée à l'aide d'un modèle de régression autologistique, un modèle de Markov à champ aléatoire avec régression logistique comme un cas particulier. La grande taille des images (à la fois en termes de nombre de pixels et de nombre de plans d'image) introduit plusieurs défis informatiques lorsque l'on utilise ce modèle. Cet exposé mettra l'accent sur l'estimation des paramètres, comparant différentes approches et discutant de la manière dont le but de l'étude -- exactitude prédictive ou interprétation du paramètre -- pourrait influencer le choix de la méthode d'estimation.

1B-C1: Biostatistics: Genetics and Imaging Topics Biostatistique : génétique et imagerie

Chair/Présidente: Zeny Feng (University of Guelph)

Room/Salle: McCain 2102

Abstracts/Résumés

YUNAN ZHU & IVOR CRIBBEN (University of Alberta)

[Monday June 15 / lundi 15 juin, 10:20]

Estimating Sparse Graphical Models

Estimation des modèles graphiques épars

Graphical models are frequently used to explore networks among a set of variables. Several methods for estimating sparse graphs have been proposed and their theoretical properties have been explored. However, their practical performance has not been studied in detail. In this work, we compare several estimation procedures (lasso, SCAD,...) and several selection criteria (AIC, BIC, CV,...) under various simulation settings. We use several evaluation criteria to compare the models and discuss in detail the superiority and deficiency of the models. In addition, we apply the techniques to a resting state fMRI data set.

Les modèles graphiques sont fréquemment utilisés pour explorer les réseaux dans un ensemble de variables. Plusieurs méthodes d'estimation des graphiques épars ont été proposées et leurs propriétés théoriques explorées. Cependant, leur performance pratique n'a pas encore été étudiée en détail. Dans cette présentation, nous comparons plusieurs procédures d'estimation (lasso, SCAD,...) et plusieurs critères de sélection (AIC, BIC, CV,...) dans divers contextes de simulation. Nous utilisons plusieurs critères d'évaluation pour comparer les modèles et discutons en détail de la supériorité et des faiblesses de chaque modèle. Enfin, nous appliquons les techniques à un ensemble de données d'IRMf à l'état de repos.

YONGLIANG ZHAI & ALEXANDRE BOUCHARD-CÔTÉ (University of British Columbia)

[Monday June 15 / lundi 15 juin, 10:35]

A Poissonian Model of Indel Rate Variation for Phylogenetic Tree Inference

Modèle poissonien de la variation du taux d'indel pour l'inférence d'arbres phylogénétiques

We introduce a stochastic process that generalizes the Poisson indel process (PIP) by allowing the insertion and deletion rate to vary in different genomic regions. This new model, called geometric Poisson indel process (GeoPIP), models long indels using high indel rates while maintaining the attractive computational properties of the PIP. The GeoPIP has the potential to help researchers identify regions of aligned sequences with fast evolutionary rates, and regions with possible long insertions and deletions. Results from simulation studies and data analysis show that modelling indel rate variation is important for accurate phylogenetic tree reconstruction.

Nous introduisons un processus stochastique qui permet de généraliser le processus d'indel de Poisson (PIP) en permettant au taux d'insertion et de délétion de varier selon la région du génome. Ce nouveau modèle, appelé processus d'indel de Poisson géométrique (GeoPIP), permet de modéliser de longs indel avec des taux d'indel élevés, tout en conservant les propriétés computationnelles intéressantes du PIP. Le modèle GeoPIP peut aider les chercheurs à identifier des régions de séquences alignées avec des taux d'évolution élevés et des régions avec des insertions et délétions potentiellement longues. Les résultats des études de simulation et de l'analyse de données montrent que la variation du taux d'indel modélisée est importante pour la reconstruction d'un arbre phylogénétique précis.

1B-C1: Biostatistics: Genetics and Imaging Topics Biostatistique : génétique et imagerie

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

[Monday June 15 / lundi 15 juin, 10:50]

Phylogenetic Inference by Divide and Conquer Sequential Monte Carlo (D&C SMC)

Inférence phylogénétique par division et conquête séquentielle de Monte-Carlo (D&C SMC)

Recently, reconstructing evolutionary histories has become a computational issue due to the increased availability of genetic sequencing data and relaxations of classical modelling assumptions. We develop a D&C SMC for phylogenetic trees to address these challenges. In phylogenetics, the tree structure used to represent evolutionary histories provides a decomposition for D&C SMC. Speciation events are used to recursively decompose the model into subproblems. Each subproblem is approximated by a population of weighted particles, which are merged and propagated to create an ancestral population. This approach provides the flexibility to relax classical assumptions on large trees by parallelizing these recursions.

Récemment, la reconstruction des histoires évolutives est devenue un problème de calcul en raison de la disponibilité accrue de données sur le séquençage génétique et de l'allègement des hypothèses classiques de modélisation. Nous développons un (D&C SMC) pour les arbres phylogénétiques afin de relever ces défis. Dans la phylogénétique, la structure de l'arbre utilisé pour représenter des histoires évolutives fournit une décomposition pour (D&C SMC). Les événements de spéciation sont utilisés pour décomposer de façon récursive le modèle en sous-problèmes. Chaque sous-problème est approximé par une population de particules pondérées, qui sont fusionnées et propagées pour créer une population ancestrale. Cette approche offre la flexibilité d'alléger les hypothèses classiques sur de grands arbres en parallélisant ces récurrences.

YUN CAI, HONG GU & TOBY KENNEY (Dalhousie University)

[Monday June 15 / lundi 15 juin, 11:05]

Supervised Non-Negative Matrix Factorization for Analysis of Microbiome Data

Factorisation de matrice non-négative supervisée pour l'analyse de données du microbiome

Non-Negative Matrix Factorization (NMF) is a very useful tool to reduce dimension of data in data mining as an unsupervised learning method. To extract more discriminant information in the training data and improve the performance of classification, we developed a new supervised NMF method. We combine feature matrices from different classes as the feature matrix for the whole data and fit a non-negative Poisson regression to calculate the weight matrix. Our method is tested on the animal dataset and moving picture dataset. The experimental results show that our supervised NMF could greatly enhance the performance of NMF for classification.

La factorisation de matrice non-négative (FMN) est un outil très utile pour réduire la dimension des données dans l'exploration de données en tant que méthode d'apprentissage non supervisée. Dans le but d'extraire plus d'information discriminante des données d'apprentissage et d'améliorer la performance de la classification, nous avons développé une nouvelle méthode FMN supervisée. Nous combinons des matrices de caractéristiques de différentes classes en tant que matrice de caractéristiques pour l'ensemble des données et ajustons une régression de Poisson non-négative pour calculer la matrice des poids. Nous testons notre méthode sur le jeu de données des animaux et celui des images en mouvement. Les résultats expérimentaux montrent que notre FMN supervisée pourrait grandement améliorer les performances de la FMN pour la classification.

BO CHEN (University of Toronto)

[Monday June 15 / lundi 15 juin, 11:20]

Statistical Analysis of Sex Chromosome in Genome-wide Association Studies

L'analyse statistique du chromosome sexuel dans les études d'association pangénomiques

1B-C1: Biostatistics: Genetics and Imaging Topics Biostatistique : génétique et imagerie

Analysis of X chromosome lags behind autosome in genome wide association studies because of its technical difficulties. Existing methods on X chromosome only test for genotype-phenotype association without allowing for separate effects of sex and sex-genotype interaction, allele frequency difference of male and female and uncertainty of X-inactivation status. For both QTL and binary traits, we use regression and meta-analysis approaches to test for genotype-phenotype association on X chromosome after considering all these issues simultaneously. A new random-effects meta-analysis approach is suggested in most realistic situations.

L'analyse du chromosome X est en décalage derrière celle des autosomes dans les études d'association pangénomiques en raison de ses difficultés techniques. Des méthodes existantes sur le chromosome X testent seulement l'association génotype-phénotype et ne permettent pas les effets séparés de sexe et d'interaction sexe-génotype, la différence de fréquence des allèles chez l'homme et la femme ainsi que l'incertitude du statut d'inactivation de l'X. Pour les locus à caractère quantitatif (QTL) et les caractères binaires, nous utilisons des méthodes d'analyse de régression et de méta-analyses pour tester l'association génotype-phénotype sur un chromosome X après avoir considéré simultanément toutes ces difficultés. Une nouvelle méthode de méta-analyse à effets aléatoires est proposée dans la plupart des situations réelles.

KEELIN GREENLAW, FAROUK NATHOO & MARY LESPERANCE (University of Victoria), **ELENA SZEFER & JINKO GRAHAM** (Simon Fraser University)

[Monday June 15 / lundi 15 juin, 11:35]

Empirical Bayes Multivariate Group-Sparse Regression for Brain Imaging Genomics

Régression multivariée bayésienne empirique à groupes clairsemés en génomique de l'imagerie cérébrale

Advances in technology for brain imaging and genotyping have motivated studies examining the influence of genetic variation on brain structure. Wang et al. (Bioinformatics, 2012) have recently developed an approach for simultaneous estimation and SNP selection based on penalized regression with a novel group $l_{2,1}$ -norm penalty. We develop a corresponding Bayesian formulation that allows for full posterior inference using Gibbs sampling with tuning parameters selected by a Monte Carlo EM algorithm. Properties of our method are investigated using simulation studies and the methodology is applied to a large dataset collected as part of the Alzheimer's Disease Neuroimaging Initiative.

Les progrès technologiques en imagerie cérébrale et génotypage ont motivé plusieurs études qui examinent l'influence de la variation génétique sur la structure du cerveau. Wang et al. (Bioinformatics, 2012) ont récemment mis au point une approche pour l'estimation simultanée et la sélection de SNP fondée sur la régression pénalisée avec une nouvelle pénalité de groupe de norme $l_{2,1}$. Nous mettons au point une formulation bayésienne correspondante qui permet une inférence a posteriori complète par échantillonnage de Gibbs avec des paramètres de réglage sélectionnés par un algorithme EM de Monte-Carlo. Nous étudions les propriétés de notre méthode à l'aide d'études de simulation et appliquons notre méthodologie à un grand ensemble de données collectées dans le cadre de l'initiative en neuroimagerie de la maladie d'Alzheimer.

1B-C2: Environmetrics Environnement

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

Room/Salle: McCain 2170

Abstracts/Résumés

DAVID J THOMSON, DAVID L RIEGERT & AARON SPRINGFORD (Queen's University)

[Monday June 15 / lundi 15 juin, 10:20]

Experiments ``Fixing'' Ground-Level Ozone

Expériences sur le rétablissement de l'ozone au sol

A quality control study of Canadian ground-level ozone measurements shows that its autocovariances are time-dependent. The data are nongaussian, have both real and instrumental outliers, plus numerous missing values. In this situation, quality control tools need to be adapted because most assume approximate stationarity. Dividing the ozone series by the solar ultraviolet insolation largely removes the annual components but the ratio is still periodically correlated. The data are also coherent over large distances, e.g., between Kejimikujik, NS and the Experimental Lakes, ON. We show how both these features can be used to help restore missing sections.

Une étude de contrôle qualitatif sur les mesures d'ozone au sol au Canada montre que ses autocovariances présentent une dépendance temporelle. Les données sont non gaussiennes et présentent des valeurs aberrantes à la fois réelles et instrumentales, ainsi que de nombreuses valeurs manquantes. Dans cette situation, il faut adapter les outils de contrôle de qualité, puisque la plupart partent d'une hypothèse de stationnarité approximative. Si l'on divise la série d'ozone par l'insolation ultraviolette solaire, on élimine en grande partie les composants annuels, mais le ratio reste périodiquement corrélé. Les données sont aussi cohérentes sur de grandes distances, par exemple entre Kejimikujik (N.-É.) et les lacs expérimentaux en Ontario. Nous montrons comment ces deux caractéristiques peuvent aider à rétablir des sections manquantes.

HERBERT NKWIMI TCHAHOU, CLAUDE GIRARD & MARTIN HAMEL (Statistique Canada)

[Monday June 15 / lundi 15 juin, 10:35]

Indicators of Database Quality for Wetlands in Canada: Example of Analysis of Environmental Data

Indicateurs de qualité des bases de données des terres humides au Canada : un exemple d'analyse de données environnementales

In order to monitor wetlands, Environment Canada (EC) has access to mega-databases that contain large quantities of information describing the Canadian territory in all its facets. To assess the quality of these databases -- that sometimes have attributes of Big Data, administrative data and survey data -- we have taken advantage of finite population survey techniques and data analysis. We present a pilot project that was implemented: the employed methodology and an overview of the obtained results.

Afin d'assurer le monitoring des terres humides, Environnement Canada (EC) dispose de méga-bases de données contenant une masse importante d'informations diverses décrivant le territoire canadien sous toutes ses facettes. Dans le but de juger de la qualité de ces bases -- qui présentent à la fois certains des attributs propres aux données volumineuses (« Big Data »), aux données administratives et aux données d'enquêtes -- nous avons exploité des techniques de sondage de populations finies et d'analyse de données. Nous présenterons un projet pilote qui a été mené : la méthodologie employée et un aperçu des résultats obtenus.

1B-C2: Environmetrics Environnement

OFIR HARARI (Simon Fraser University), **MATTHEW T. PRATOLA** (Ohio State University), **DEREK BINGHAM & GWENN E. FLOWERS** (Simon Fraser University)

[Monday June 15 / lundi 15 juin, 10:50]

Design on Non-Convex Regions: Optimal Experiments for Spatial Process Prediction

Plan d'expérience sur des régions non convexes : expériences optimales pour la prévision de processus spatiaux

Modeling a response over a non-convex design region is a common problem in diverse areas such as engineering and geophysics. The tools available to model and design for such responses are limited and have received little attention. We propose a new method for selecting design points over non-convex regions that is based on the application of multidimensional scaling to the geodesic distance. Optimal designs for prediction are described, with special emphasis on Gaussian process models, followed by a simulation study and an application in glaciology.

La modélisation d'une réponse sur une région non convexe du plan constitue un problème courant dans plusieurs domaines, notamment en ingénierie et en géophysique. Les outils disponibles pour la modélisation et la planification de telles réponses sont limités et ont suscité peu d'intérêt. Nous proposons une nouvelle méthode pour la sélection de points du plan sur des régions non convexes basée sur l'application d'une mise à l'échelle multidimensionnelle à la distance géodésique. Nous décrivons les plans optimaux pour la prévision en insistant particulièrement sur les modèles de processus gaussiens. Nous présentons enfin une étude de simulation et une application en glaciologie.

AUDREY-ANNE VALLÉE (Université de Neuchâtel), **BASTIEN FERLAND-RAYMOND** (Ministère des forêts, de la faune et des parcs du Québec), **LOUIS-PAUL RIVEST** (Université Laval), **YVES TILLÉ** (Université de Neuchâtel)

[Monday June 15 / lundi 15 juin, 11:05]

Incorporating Spatial and Operational Constraints in the Sampling Designs for Forest Inventories

Incorporation de contraintes spatiales et opérationnelles dans des plans d'échantillonnage pour des inventaires forestiers

In the province of Quebec, Canada, the forest is examined through regular inventories. Requirements for the spreading and the type of trees and for the cost are difficult to manage. We show that modern and advanced sampling techniques can be used to improve the planning of the forest inventories, even if there are many requirements. Our design includes balanced sampling, highly stratified balanced sampling and sample spreading through a two stage sample. The impact of these techniques on the satisfaction of the requirements and on the precision of survey estimates is investigated using field data from a Quebec inventory.

Dans la province du Québec, Canada, la forêt est examinée par l'entremise d'inventaires réguliers. Les exigences opérationnelles sur la répartition et le type des placettes sélectionnées et sur les coûts sont difficiles à gérer. Malgré les nombreuses exigences, les techniques d'échantillonnage modernes peuvent être utilisées pour améliorer la planification des inventaires forestiers. Notre plan de sondage inclut l'échantillonnage équilibré, l'échantillonnage équilibré hautement stratifié et l'étagement de l'échantillon à travers un plan à deux degrés. L'impact de ces techniques sur la satisfaction des exigences opérationnelles et sur la précision des estimations est étudié à l'aide des données d'un inventaire forestier québécois.

CAMILA CASQUILHO-RESENDE (University of British Columbia), **NHU D. LE** (BC Cancer Agency), **JAMES V. ZIDEK** (University of British Columbia)

[Monday June 15 / lundi 15 juin, 11:20]

Spatio-temporal Modelling of Temperature Fields in the Pacific Northwest

Modélisation spatio-temporelle des champs de température dans le Nord-Ouest Pacifique

1B-C2: Environmetrics Environnement

The modelling of temperature fields, which are crucial to understand a region's climate, can be challenging due to the topography of the study region. For instance, in the Pacific Northwest, extensive forests, mountains and proximity to the Pacific Ocean may create sudden changes in climate, contributing to the complexity of the modelling in the area. In this work, we develop a hierarchical spatio-temporal model that handles nonstationarity for characterizing daily temperature fields in that region and describe a Bayesian method that can be used for spatial prediction.

La modélisation des champs de température, qui sont essentiels pour comprendre le climat d'une région, peut poser des défis en raison de la topographie de la région étudiée. Par exemple, dans le Nord-Ouest Pacifique, les grandes forêts, les montagnes et la proximité de l'océan Pacifique peuvent engendrer des changements climatiques soudains, ce qui rend la modélisation complexe dans cet endroit. Nous développons un modèle hiérarchique spatio-temporel qui traite de la non-stationnarité pour caractériser les champs de température quotidiens dans cette région. Nous décrivons également la méthode bayésienne qui peut être utilisée pour la prédiction spatiale.

PETER STARSZYK & MARTIN LYSY (University of Waterloo)

[Monday June 15 / lundi 15 juin, 11:35]

Inferring Chemical Kinetics from Direct and Indirect Concentration Data

Inférence en cinétique chimique à partir de données de concentration directes et indirectes

Many chemical compounds used in the agricultural industry bring large amounts of arsenic into the soil. As this poses serious environmental hazards, designing safe and effective arsenic decontaminating agents is an active research area. To do this, it is crucial to understand chemical kinetics between arsenic and various geochemicals at the molecular level. However, state-of-the-art direct measurement techniques aggregate concentration measures making it impossible to infer individual reaction rates. Here we include the modelling of mass spectrometry data which serve as indirect proxies to individual chemical concentrations. The study ultimately seeks identifiability of parameters through simulated and real spectral data.

De nombreux composés chimiques utilisés dans l'industrie agricole introduisent de grandes quantités d'arsenic dans le sol. Comme cette substance pose de graves risques environnementaux, la conception d'agents de décontamination sécuritaires et efficaces constitue un domaine de recherche très actif. Cette conception exige une grande compréhension de la cinétique chimique entre l'arsenic et les diverses substances géochimiques au niveau moléculaire. Cependant, les techniques de mesure directe à la fine pointe agrègent les mesures de concentration, ce qui rend impossible l'inférence des taux de réaction individuels. Nous proposons d'inclure la modélisation de données de spectrométrie de masse qui servent d'approximations indirectes des concentrations chimiques individuelles. L'étude cherche en définitive l'identifiabilité des paramètres à l'aide de données spectrales simulées et réelles.

1B-C3: Overdispersion, Mixtures, and Related Topics Surdispersion, mélanges et sujets connexes

Chair/Président: Alexandre Leblanc (University of Manitoba)

Room/Salle: McCain 2184

Abstracts/Résumés

XINGDE DUAN & RENJUN MA (University of New Brunswick)

[Monday June 15 / lundi 15 juin, 10:20]

"Multivariate Poisson mixed models with unstructured covariance structure"

Modèles de Poisson mixtes multivariés avec une structure de covariance non structurée

Multivariate count data appear frequently in the areas of economics, engineering, epidemiology, forestry, medical and public health studies; however, modelling multivariate count data with flexible covariance structure among marginal counts remains a challenging issue in statistics. In this talk, we introduce a new model for multivariate count data by incorporating multivariate random effect into Poisson regression model. This new model allows us to handle both overdispersion and various covariance structures. Our model estimation is done using the orthodox best linear unbiased predictor approach. Our proposed methodology is illustrated with two real life examples in biology and health care.

Des données de dénombrement multivariées apparaissent fréquemment dans les domaines de l'économie, l'ingénierie, l'épidémiologie, la foresterie, les études de santé publiques et médicales; par contre, modéliser les données de dénombrement multivariées avec une structure de covariance flexible parmi des dénombrements marginaux demeure une question difficile en statistique. Dans cet exposé, nous introduisons un modèle pour les données de dénombrement multivariées en intégrant un effet aléatoire multivarié dans un modèle de régression de Poisson. Ce nouveau modèle nous permet de gérer à la fois la surdispersion et les différentes structures de covariance. L'estimation de notre modèle est effectuée à l'aide de l'approche du meilleur prédicteur linéaire non biaisé. Notre méthodologie est illustrée avec deux exemples réels en biologie et en soins de santé.

MOHAMMAD RAJIBUL ISLAM MIAN & SUDHIR PAUL (University of Windsor)

[Monday June 15 / lundi 15 juin, 10:35]

Estimation for Zero Inflated Overdispersed Generalized Linear Model with Missing Covariates as well as Response
Estimation pour les modèles linéaires généralisés surdispersés à surreprésentation de zéros avec covariables et réponses manquantes

Zero inflation, overdispersion and missing observations in the data are important features in many applied problems. We study maximum likelihood and extended quasi-likelihood estimation techniques for estimating the regression parameters as well as dispersion and zero inflation parameters of the generalized linear model for count and proportion data in the presence of missing at random observations in covariates as well as in response. We use weighted EM algorithm for the estimation procedures assuming the missing covariates are discrete as well as continuous. We use real dataset

La surreprésentation de zéros, la surdispersion et les observations manquantes dans les données sont d'importantes fonctionnalités de certains problèmes appliqués. Nous étudions les techniques d'estimation du maximum de vraisemblance et de quasi-vraisemblance élargie pour estimer les paramètres de régression et les paramètres de dispersion et de surreprésentation de zéros du modèle linéaire généralisé pour les données de comptage et de proportion, lorsque des observations aléatoires manquent pour les covariables et les réponses. Nous utilisons un algorithme EM pondéré pour les procédures d'estimation, partant de l'hypothèse que les covariables manquantes sont discrètes et continues. Nous uti-

1B-C3: Overdispersion, Mixtures, and Related Topics Surdispersion, mélanges et sujets connexes

and perform simulation study to compare the estimation procedures.

lisons un ensemble de données réelles et effectuons une étude de simulation pour comparer les procédures d'estimation.

WILLIAM H. AEBERHARD (Dalhousie University), **EVA CANTONI** (University of Geneva), **STEPHANE HERITIER** (Monash University)

[Monday June 15 / lundi 15 juin, 10:50]

Nonparametric Estimation of Mixing Distributions for Mixed Poisson Models

Estimation non paramétrique des lois mélanges pour les mélanges de Poisson

For the modeling of overdispersed count data, the class of mixed Poisson models includes many well known special cases, such as the negative binomial and the Poisson-inverse Gaussian regressions. These special cases are identified by the distribution assumed for the (unobserved) mixing variables. Such assumptions are in practice difficult to verify and yet their validity is crucial as they determine the functional form of the variance of the (observed) count response variables. We propose a semiparametric framework where the mixing variables need not be identically distributed and their distributions are estimated from the data along with the regression coefficients.

Dans le cadre de la modélisation de comptages surdispersés, la classe de modèles de mélanges de Poisson inclut de nombreux cas particuliers comme les régressions binomiale négative et Poisson-gaussienne inverse. Ces cas particuliers se distinguent par la loi postulée pour les variables de mélange (non-observées). De tels postulats sont difficiles à vérifier et pourtant leur validité est cruciale puisqu'ils déterminent la forme fonctionnelle de la variance des comptages (observés). Nous proposons un cadre semi-paramétrique dans lequel les variables de mélange ne sont pas requises d'être identiquement distribuées et leurs lois sont estimées avec les paramètres de régression.

CHRISTINE NGUYEN, JUSTYNE MOTTOLA & MEI LING HUANG (Brock University)

[Monday June 15 / lundi 15 juin, 11:05]

On Inference of a Mixture Pareto Distribution

Sur l'inférence d'une distribution Pareto de mélanges

The Pareto distribution is a heavy tailed distribution with many applications. In this paper we consider a real world example with heavy tailed observations, which leads us to propose a mixture truncated Pareto distribution (MTPD). There are difficulties in the estimation thresholds of the MTPD. We construct a cluster truncated Pareto distribution (CTPD) by using a two-point slope technique to estimate the MTPD. Finally, we apply the MTPD and CTPD to the example and compare the proposed method with existing methods. The results of goodness-of-fit tests show that the cluster method produces a good fitting distribution with real world data.

La distribution de Pareto est une distribution à ailes lourdes avec plusieurs applications. Dans cette présentation, nous considérons un exemple réel avec des observations d'ailes lourdes, qui nous mène à proposer un mélange tronqué de distribution de Pareto (MTDP). Il y a des difficultés dans les seuils d'estimation du MTDP. Nous avons construit un regroupement tronqué de distribution de Pareto (RTDP) en utilisant une technique de pente à deux points afin d'estimer le MTDP. Finalement, nous appliquons le MTDP et le RTDP à l'exemple et comparons la méthode proposée avec les méthodes existantes. Les résultats des tests d'adéquation démontrent que la méthode de regroupement produit une distribution qui cadre bien avec les données réelles.

1B-C3: Overdispersion, Mixtures, and Related Topics Surdispersion, mélanges et sujets connexes

MOHAMED AMEZZIANE (Central Michigan University)

[Monday June 15 / lundi 15 juin, 11:20]

The Exact Distribution of a Certain Class of Nonparametric Multivariate Two-Sample Test Statistics

Distribution exacte d'une certaine classe de statistiques de test multivariées non paramétriques à deux échantillons

When conducting a nonparametric two-sample test for equality of multivariate distributions, data are often visualized through a weighted bipartite graph where vertices correspond to observations and weighted edges represent the distance between observations. This particular representation gave rise to many test statistics computed as functions of the graph's weight. We derive the null distributions for two types of such test statistics: the first corresponds to a bipartite graph with a fixed degree while the second corresponds to graphs with limited edge weights. The null distributions are used to investigate the performance of these tests when the two samples are unbalanced.

Lors d'un test non paramétrique à deux échantillons vérifiant l'égalité de distributions multivariées, les données sont souvent visualisées au moyen d'un graphe biparti pondéré dans lequel les sommets correspondent aux observations et les arêtes pondérées représentent la distance entre les observations. Cette représentation particulière a mené à plusieurs statistiques de test calculées comme fonctions des poids du graphe. Nous obtenons les lois sous l'hypothèse nulle pour deux types de ces statistiques de test : le premier correspond à un graphe biparti avec un degré fixe et le second correspond à des graphes dont les poids des arêtes sont limités. Les distributions nulles sont utilisées pour étudier la performance de ces tests en présence de deux échantillons non balancés.

1B-C4: Time Series and Probability Séries chronologiques et probabilité

Chair/Président: Alexandre Bouchard-Côté (University of British Columbia)

Room/Salle: McCain 2198

Abstracts/Résumés

ABDOLLAH SAFARI & RACHEL MACKAY ALTMAN (Simon Fraser University), **BRIAN LEROUX** (University of Washington)

[Monday June 15 / lundi 15 juin, 10:20]

Estimators of Regression Coefficients in Models for Time Series of Counts

Estimateurs des coefficients de régression dans les modèles de séries chronologiques de dénombrement

Correlated count data, especially time series of counts, arise in many contexts, including the monitoring of disease incidence. Latent variable models, such as hidden Markov models and generalized linear mixed models, are one way of capturing the autocorrelation in such data. However, computing the MLEs of the regression coefficients in these models may be challenging. We consider a broad class of latent variable models and discuss three different estimators, including an estimator based on the GLM. Although the latter does not account for the autocorrelation in the data, we show that, under some conditions, it outperforms the other estimators.

Les données de dénombrement corrélées, en particulier les séries chronologiques de dénombrement, émergent dans de nombreux contextes, y compris le suivi de l'incidence des maladies. Les modèles de variables latentes, comme les modèles de Markov cachés et les modèles linéaires mixtes généralisés, sont une façon de capter l'autocorrélation de telles données. Toutefois, le calcul des EMV des coefficients de régression de ces modèles peut être difficile. Nous considérons une vaste classe de modèles de variables latentes et discutons trois estimateurs différents, y compris un estimateur basé sur le MLG. Bien que celui-ci ne tienne pas compte de l'autocorrélation dans les données, nous montrons que, sous certaines conditions, il surpasse les autres estimateurs.

JINJUN TONG & J CONCEPCIÓN LOREDO-OSTI (Memorial University of Newfoundland)

[Monday June 15 / lundi 15 juin, 10:35]

Estimation and Filtering of State Space Model with GEV marginals and Exponential Levy-type Stable Errors

Estimation et filtrage d'un modèle d'espace d'états au moyen de marges de valeur extrême généralisée (GEV) et d'erreurs exponentielles stables de type Lévy

In this presentation a non-linear non-Gaussian state space model with generalized extreme value (GEV) marginals is proposed. For heavy-tailed stationary GEV observations, parameter estimation as well as prediction and filtering are of interest. The fact that ordered statistics of GEV marginals from an AR(1) time series have the same distribution as the ordered statistics of independent variables under some weakly mixing conditions can be used to estimate tail and scale parameter via recursive linear regression. Prediction and filtering can be addressed through particle filter techniques. Different

Dans cette présentation, nous proposons un modèle d'espace d'états non gaussien et non linéaire avec des marges de valeur extrême généralisée. L'estimation de paramètres ainsi que la prédiction et le filtrage sont intéressantes pour les observations avec marges de valeur extrême généralisée stationnaires à ailes relevées. Le fait que les statistiques d'ordre des marges de valeur extrême généralisée provenant des séries chronologiques AR(1) aient la même distribution que les statistiques d'ordre de variables indépendantes dans certaines conditions faibles de mélange permet d'estimer les paramètres de l'aile et de l'échelle au moyen d'une régression linéaire récursive. La prédiction et le filtrage

1B-C4: Time Series and Probability Séries chronologiques et probabilité

importance particle filter schemes are compared on this presentation.

peuvent être effectués avec des techniques de filtrage de particules et nous en comparons différents schémas importants dans cette présentation.

TREVOR THOMSON, MELODY GHAHRAMANI & SHAKHAWAT HOSSAIN (University of Winnipeg)

[Monday June 15 / lundi 15 juin, 10:50]

Efficient Estimation for Time Series Following Generalized Linear Models

Estimation efficace des séries chronologiques qui suivent des modèles linéaires généralisés

We consider shrinkage and pretest estimation methods for time series following generalized linear models. Efficient estimation strategies are developed for when there are many covariates in the model and where some of them are not statistically significant. We investigate the relative performances of shrinkage and pretest estimators with respect to the unrestricted maximum partial likelihood estimator. Our Monte Carlo studies show that the shrinkage estimators have a significantly lower relative mean squared error as compared to maximum partial likelihood estimators when the shrinkage dimension exceeds two. The practical benefits of the proposed methods are illustrated using a real data example.

Nous étudions les méthodes d'estimation de rétrécissement et d'essai préliminaire pour les séries chronologiques qui suivent des modèles linéaires généralisés. Nous développons des stratégies d'estimation efficace pour les modèles qui incluent de nombreuses covariables dont certaines ne sont pas statistiquement significatives. Nous étudions les performances relatives des estimateurs de rétrécissement et d'essai préliminaire par rapport à l'estimateur du maximum de vraisemblance partielle sans restriction. Nos études de Monte-Carlo montrent que les estimateurs de rétrécissement présentent une erreur quadratique moyenne relative nettement inférieure aux estimateurs du maximum de vraisemblance partielle lorsque la dimension de rétrécissement est supérieure à deux. Nous illustrons les avantages pratiques des méthodes proposées par un exemple fondé sur des données réelles.

HAOSUI DUANMU, WILLIAM WEISS & JEFFREY ROSENTHAL (University of Toronto)

[Monday June 15 / lundi 15 juin, 11:05]

Nonstandard Analysis and its Application to Markov Chain Convergence Problem

Analyse non standard appliquée au problème de convergence des chaînes de Markov

We apply nonstandard analysis to construct a hyperfinite Markov chain which behaves like a finite state space Markov chain but is more widely applicable. I will show how to prove the convergence to its weak stationary distribution using the idea of infinitesimal coupling. I will also present the idea of how to construct a corresponding hyperfinite Markov chain for a standard general state space discrete time chain such that the convergence of the hyperfinite chain would imply the convergence of the original chain. This idea may be extended to more general Markov chains by considering a hyperfinite time line.

Nous appliquons une analyse non standard pour la construction d'une chaîne de Markov hyperfinie qui se comporte comme une chaîne de Markov à espace d'états fini. Cette analyse est aussi applicable à plus grande échelle. Je vais démontrer comment prouver sa convergence à sa loi stationnaire faible en utilisant l'idée de jumelage infinitésimal. Je vais aussi présenter l'idée de base qui permet de construire la chaîne de Markov hyperfinie correspondante pour une chaîne à temps discret à état d'espace général standard telle que la convergence de la chaîne hyperfinie impliquerait la convergence de la chaîne initiale. Cette idée peut être élargie à des chaînes de Markov générales en considérant une ligne de temps hyperfinie.

1B-C4: Time Series and Probability Séries chronologiques et probabilité

JINYOUNG YANG & JEFFREY ROSENTHAL (University of Toronto)

[Monday June 15 / lundi 15 juin, 11:20]

Automatically Tuned General-Purpose MCMC via New Adaptive Diagnostics

Méthodes MCMC d'usage général auto-réglées via de nouveaux diagnostics adaptatifs

Adaptive MCMC is an attempt to modify a Markov chain "on the fly" so the chain can converge quicker. As many adaptive techniques involve using information from the past iterations of the Markov chain, the chain loses its Markovian property. The algorithm introduced here employs a few adaptive rules to tune a symmetric random walk Metropolis algorithm. The adaptation stops once the algorithm diagnoses that further adaptation is unnecessary. This choice avoids the theoretical difficulties which arise in proving the convergence of chains which adapt infinitely. Once the chain stops adapting, the algorithm runs a standard non-adaptive MCMC.

Les méthodes MCMC constituent une tentative de modifier une chaîne de Markov à la volée afin qu'elle puisse converger plus rapidement. Étant donné que de nombreuses techniques adaptatives utilisent des informations tirées des itérations précédentes de la chaîne de Markov, la chaîne perd sa propriété markovienne. L'algorithme proposé ici emploie quelques règles adaptatives pour régler un algorithme de Metropolis à mouvement aléatoire symétrique. L'adaptation s'arrête une fois que l'algorithme détermine que toute adaptation supplémentaire est inutile. Ce choix évite les difficultés théoriques qui surgissent lorsqu'il s'agit de prouver la convergence de chaînes qui s'adaptent à l'infini. Une fois que la chaîne a arrêté de s'adapter, l'algorithme effectue une méthode MCMC standard non-adaptative.

SUDHAKAR ACHATH & AMRITA VISHWA VIDYAPEETHAM (University of Saskatchewan)

[Monday June 15 / lundi 15 juin, 11:35]

Imprecise Probabilities for Measuring Tracking Error Risk

Probabilités imprécises pour mesurer les risques d'erreur de suivi

This study measures tracking error for an actively managed portfolio of assets (funds) as interval-valued probability in terms of previsions (Walley and Fine, 1982). The sources of aleatory and epistemic uncertainty about the market returns is the lack of knowledge of the fund manager about the management actions of each company whose asset is listed, and stochasticity of the securities market. The empirical study using data is based on the Fundamental Theorem of Probability of de Finetti (1974), and extends earlier applications. Lower and upper previsions are found by solving sets of linear and linear-fractional programs.

Cette étude mesure l'erreur de suivi pour un portefeuille d'actifs (fonds) géré dynamiquement par un intervalle de probabilité en termes des prévisions (Walley et Fine, 1982). Les sources d'incertitude aléatoire et épistémique sur les rendements du marché est le manque de connaissance du gestionnaire de fonds sur les mesures de gestion de chaque entreprise dont l'actif est coté, et la stochasticité du marché des valeurs mobilières. L'étude empirique à partir de données est basée sur le théorème fondamental des probabilités de Finetti (1974). Des bornes inférieures et supérieures pour les prévisions se trouvent en résolvant des ensembles de programmes linéaires et linéaires-fractionnaires.

1D-I1: Advances in High-Dimensional Statistical Inference Avancées en inférence statistique à dimensions élevées

Organizer and Chair / Responsable et présidente: Yingli Qin (University of Waterloo)

Room/Salle: Rowe 1028

Abstracts/Résumés

GOURAB MUKHERJEE (University of Southern California), **LAWRENCE D. BROWN** (University of Pennsylvania), **PAAT RUSMEVICHIENTONG** (University of Southern California)

[Monday June 15 / lundi 15 juin, 13:30]

Adaptive Inventory Management Strategies for a Large Scale Newsvendor

Stratégies adaptatives de gestion des stocks pour une grande échelle de marchands de journaux

We study the inventory management problem of a Newsvendor that sells a large number of products. Based on observing noisy estimates of the underlying demands our objective is to predict the optimal stocking levels that minimize the cumulative backorder and holding costs. New perspectives in risk estimation theory are developed to design strategies that adapt to unknown data structures and have desirable properties. We construct data adaptive shrinkage strategies by minimizing Hermite polynomial based efficient risk estimates. In calculating the magnitude of shrinkage, our proposed predictive rules incorporate the non-symmetric nature of the loss function and are asymptotically optimal.

Nous étudions le problème de gestion des stocks d'un marchand de journaux qui vend un grand nombre de produits. Basés sur l'observation des estimations perturbées par des bruits des demandes sous-jacentes, notre objectif est de prévoir les stocks à des niveaux optimaux qui permettent de réduire les coûts de commandes en retard et de possession. De nouvelles perspectives dans la théorie d'estimation de risques sont développées pour créer des stratégies qui sont adaptées à des structures de données méconnues et qui ont les caractéristiques recherchées. Nous avons élaboré des stratégies de rétrécissement adaptées selon les données en minimisant les estimations de risques efficaces basées sur le polynôme de Hermite. En calculant l'ampleur du rétrécissement, nos règles prédictives proposées incorporent la nature asymétrique de la fonction de perte et sont asymptotiquement optimales.

ZHIGANG YAO (National University of Singapore), **SHIQIONG HUANG & JIASHUN JIN** (Carnegie Mellon University)

[Monday June 15 / lundi 15 juin, 14:00]

Partial Correlation Screening for Estimating Large Precision Matrices, with Applications to Classification

Filtrage par corrélation partielle afin d'estimer des matrices de grande précision, avec applications à la classification

We propose Partial Correlation Screening (PCS) as a new row-by-row approach. To estimate the i -th row of Ω , $1 \leq i \leq p$, PCS uses a Screen step and a Clean step. To estimate a row of Ω , PCS only needs a few rows (determined sequentially) of the empirical covariance matrix. This enables PCS to execute the estimation of a large precision matrix (e.g., $p = 10K$) in a few minutes, and open doors to estimating much larger precision matrices. Combining Higher Criticism Thresholding (HCT) with any ap-

Nous proposons le filtrage par corrélation partielle (FCP) comme nouvelle approche ligne par ligne. Afin d'estimer la i ème ligne de Ω , $1 \leq i \leq p$, le FCP utilise une étape Filtre et une étape Nettoyage. Afin d'estimer une ligne de Ω , le FCP n'a besoin que de quelques lignes (déterminées de manière séquentielle) de la matrice de covariance empirique. Ceci permet à FCP d'exécuter l'estimation de matrice à grande précision (p. ex., $p = 10K$) en quelques minutes, ouvrant ainsi la porte à l'estimation de matrices de précision beaucoup plus grandes. En combinant le seuil de

1D-I1: Advances in High-Dimensional Statistical Inference Avancées en inférence statistique à dimensions élevées

proach to estimating Ω gives a new classifier. We set up a general theoretical framework and show that in a broad context, PCS fully recovers the support of Ω and HCT-PCS yields optimal classification behavior.

critique supérieur (SCS) avec quelconque approche pour estimer Ω donne un nouveau classificateur. Nous établissons un cadre théorique général et démontrons que dans un contexte large, le FCP récupère pleinement le support Ω et que le SCS-FCP offre un comportement de classification optimal.

ABBAS KHALILI (McGill University), **ANAND VIDYASHANKAR** (George Mason University)

[Monday June 15 / lundi 15 juin, 14:30]

Post-Model Selection Inference for Finite Mixture of Regression (FMR) Models

Inférence post-sélection de modèles pour les modèles de mélange fini de régression

FMR models are used in situations where several sub-populations within a population exist but the sub-populations themselves are unknown. Recently developed statistical methods for parameter estimation and variable selection in FMR models involve regularization techniques such as the LASSO and SCAD. However, it is well-known that the variability due to model selection stage causes bias in the resulting statistical inference. In this presentation, we address this issue in FMR and several related mixture models.

Les modèles de mélange fini de régression sont utilisés dans des situations dans lesquelles plusieurs sous-populations dans une population existent, mais les sous-populations mêmes ne sont pas connues. Des méthodes statistiques développées récemment pour l'estimation de paramètres et la sélection de variables dans les modèles de mélange fini de régression visent des techniques de régularisation telles que LASSO et SCAD. Cependant, il est bien connu que la variabilité engendrée par le degré de sélection du modèle biaise l'inférence statistique obtenue. Dans cette présentation, nous abordons cette question dans les modèles de mélange fini de régression et dans plusieurs modèles de mélanges connexes.

1D-I2: Business and Industrial Applications Applications industrielles et de gestion

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

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: McCain - Scotiabank

Abstracts/Résumés

SHIRLEY MILLS & DAN WU (Carleton University)

[Monday June 15 / lundi 15 juin, 13:30]

Privacy Preserving Classification for Partitioned Data

Protéger la confidentialité de la classification des données partitionnées

The development of privacy-preserving data analytic methods has become increasingly necessary to protect sensitive personal information. In this presentation we consider the cases of horizontally or vertically partitioned data where the partitions are held by two different organizations that are unable to share the datasets with each other due to privacy concerns. We combine classification methods with a perturbation method presented by Du et al. [2004] to protect privacy. The resultant methodology is incorporated into R code and for illustrative purposes is applied to a medical dataset, together with comparative misclassification rates.

Le développement de méthodes analytiques de données préservant la vie privée est devenu de plus en plus nécessaire afin de protéger les renseignements personnels de nature délicate. Dans cette présentation, nous considérons les cas de données partitionnées horizontalement ou verticalement où les partitions sont détenues par deux organismes différents qui sont incapables de partager entre eux les jeux de données en raison de problèmes de confidentialité. Nous combinons des méthodes de classification avec une méthode de perturbation présentée par Du et al. [2004] pour protéger la vie privée. La méthodologie résultante est incorporée dans un code R et est appliquée à un jeu de données médicales à titre illustratif, ainsi que des taux comparatifs d'erreurs de classification.

STEFAN STEINER, TRICIA COOPER-BARFOOT & JOCK MACKAY (University of Waterloo)

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

Regulating the Bias-Variance Tradeoff in Analysis of Survey Data over Time with Application to a Customer Loyalty Survey

Réguler le compromis biais-variance dans l'analyse des données d'enquête au fil du temps avec application à une enquête de fidélisation de la clientèle

In the analysis of survey data, common objectives include estimating the population average, tracking time trends, and comparing population subgroups. When samples are taken over time, we can estimate using only the present time data or also include historical data. However, when the characteristic is drifting over time and sample sizes are small, the decision to include historical data trades precision for bias. We propose regulating the bias-

Dans l'analyse des données d'enquête, des objectifs communs incluent l'estimation de la moyenne de la population, le suivi des tendances temporelles et la comparaison de sous-groupes de la population. Lorsque les échantillons sont prélevés au fil du temps, nous pouvons estimer en utilisant seulement les données courantes ou inclure également les données historiques. Cependant, lorsque la caractéristique est à la dérive dans le temps et que les tailles d'échantillons sont petites, la décision d'inclure les données historique se

1D-I2: Business and Industrial Applications Applications industrielles et de gestion

variance tradeoff using Weighted Estimating Equations based on a suitable Generalized Linear Model that incorporates covariates. A customer loyalty survey for a smartphone vendor will be presented and resulting present time estimates of Net Promoter Score will be compared across various approaches applied to example data and simulated data.

veut un compromis entre la précision et le biais. Nous proposons de réguler le compromis biais-variance à l'aide des équations d'estimation pondérées sur un modèle linéaire généralisé approprié intégrant les covariables. Une enquête de fidélisation de la clientèle pour un fournisseur de téléphones intelligents sera présentée et les résultats des estimations en temps réel du Net Promoter Score seront comparées pour différentes approches appliquées aux données de l'exemple et à des données simulées.

1D-I3: Novel Statistical Methods for Patient-Reported Outcome Measures Nouveaux modèles statistiques pour les mesures de résultats rapportées par les patients

Organizer and Chair / Responsable et présidente: Lisa Lix (University of Manitoba)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

MYRIAM BLANCHIN, ALICE GUILLEUX, JEAN-BENOIT HARDOUIN & VÉRONIQUE SÉBILLE (University of Nantes)

[Monday June 15 / lundi 15 juin, 13:30]

Assessing Change in Longitudinal PROMs Studies: Using IRT Polytomous Models to Detect and Adjust for Response Shift

Mesurer l'évolution dans les études longitudinales utilisant des PROMs: une méthode basée sur les modèles IRT polytomiques pour la détection et la prise en compte du décalage de la réponse

Assessing change in longitudinal PROMS studies can be a challenge. Patients' standards, values, or conceptualization of what the PROMs are intended to measure can change reflecting patient adaptation to the disease. As a consequence of this "response shift", observed patient's evolutions may thus not properly reflect true perceived health changes. Item Response Theory models relies on an underlying response model relating parameters characterizing the PROMS instrument to a latent parameter interpreted as the true individual value of the concept measured by the PROMs. A change in parameters related to the PROMS instrument might reflect response shift. We will describe how Item Response Theory models could be used to detect response shift and to measure true perceived health changes.

La mesure de l'évolution de données PROMs peut s'avérer compliquée. Les références, les valeurs ou la conceptualisation qu'ont les patients du concept mesuré par les PROMs peuvent évoluer au cours du temps reflétant l'adaptation des patients à leur maladie. En raison de ce « décalage de la réponse », les évolutions observées chez les patients peuvent différer des véritables changements perçus. Les modèles de la théorie de réponse aux items (IRT) modélisent le lien entre une variable latente représentant la vraie valeur du concept mesuré et des paramètres caractérisant l'instrument de mesure. Des changements dans ces paramètres peuvent indiquer la présence de décalage de la réponse. Nous détaillerons comment utiliser les modèles IRT pour détecter le décalage de la réponse et mesurer le changement réellement perçu.

ELEANOR PULLENAYEGUM (University of Toronto)

[Monday June 15 / lundi 15 juin, 14:00]

Handling Health Utility Data: the Example of the EQ-5D

Traitement des données sur l'état de santé : l'exemple de l'EQ-5D

Health utility is a patient-reported outcome that is key to economic evaluations. The EQ-5D is a popular instrument that captures health utilities through a short questionnaire. EQ-5D health utilities pose some interesting challenges. Health utilities mea-

L'état de santé, résultat déclaré par le patient, est essentiel aux évaluations économiques. L'indice à cinq dimensions EuroQol (EQ-5D) est un outil populaire qui saisit les états de santé au moyen d'un bref questionnaire. Les états de santé (EQ-5D) posent des défis intéressants. Les états de

1D-I3: Novel Statistical Methods for Patient-Reported Outcome Measures Nouveaux modèles statistiques pour les mesures de résultats rapportées par les patients

sured using the EQ-5D have an unusual distribution that is difficult to model parametrically. Moreover, the scoring algorithm that is used to convert questionnaire responses to health utilities does so subject to prediction errors; the resulting EQ-5D health utilities are subject to sources of uncertainty that are mistakenly ignored in practice. This talk will discuss how to address these challenges.

santé mesurés par l'utilisation de l'EQ-5D ont une distribution inhabituelle difficile à modéliser paramètre par paramètre. De plus, l'algorithme de cotation utilisé pour convertir les réponses du questionnaire sur les états de santé peut donner lieu à des erreurs de prédiction. Les états de santé l'EQ-5D obtenus sont soumis à des sources d'incertitude qui sont ignorées à tort dans la pratique. Cet exposé abordera la manière pour relever ces défis.

TOLULOPE SAJOBI & YUKUN ZHANG (University of Calgary), **LISA LIX** (University of Manitoba)

[Monday June 15 / lundi 15 juin, 14:30]

On Discrimination in Multivariate Non-Normal Repeated Measures Data

Discrimination dans des données multivariées non normales de mesures répétées

Multivariate repeated measures data arise in studies in which two or more groups of individuals are repeatedly measured on several outcome variables. Repeated measures discriminant analysis models, which assume structured means and/or covariances, have been developed for predicting group membership in repeated measures data. However, these procedures may be sensitive to departures from the multivariate normality assumption. We propose repeated measures discriminant analysis procedures, which assume parsimonious covariance structures, based on maximum trimmed likelihood estimation methods for predicting group membership in multivariate non-normal repeated measures data. Data from a longitudinal health-related quality of life study are used to demonstrate the implementation of our methods. Repeated measured discriminant analysis models can be used to classify new observations into population groups.

Les données de mesures répétées multivariées surviennent dans des études dans lesquelles deux ou plusieurs groupes d'individus sont mesurés de façon répétitive en fonction de plusieurs variables des résultats. Les modèles d'analyse discriminante des mesures répétées, qui supposent que les moyennes et/ou les covariances ont une certaine structure, ont été développés pour prédire l'appartenance à un groupe dans des données de mesures répétées. Cependant, ces procédures peuvent être sensibles aux écarts par rapport à l'hypothèse de normalité multivariée. Nous proposons des procédures d'analyse discriminante de mesures répétées qui supposent des structures de covariance parcimonieuses, en fonction de méthodes tronquées d'estimation du maximum de vraisemblance pour prédire l'appartenance à un groupe dans des données de mesures répétées multivariées non normales. Les données provenant de l'étude longitudinale sur la qualité de vie liée à la santé sont utilisées pour démontrer la mise en œuvre de nos méthodes. Des modèles d'analyse discriminante de mesures répétées peuvent être utilisés pour classifier de nouvelles observations dans des groupes de population.

1D-I4: Recent Developments in Small Area Estimation Développements récents en estimation pour petites régions

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

Session sponsored by the Survey Methods Section

Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: McCain 2170

Abstracts/Résumés

MALAY GHOSH (University of Florida), **TATSUYA KUBOKAWA** & **SHONOSUKE SUGASAWA** (University of Tokyo), **SANJAY CHAUDHURI** (National University of Singapore)

[Monday June 15 / lundi 15 juin, 13:30]

Prediction in Heteroscedastic Nested Error Regression Models with Random Dispersions

Prédiction dans les modèles de régression à erreurs emboîtées hétéroscédastiques avec dispersions aléatoires

The paper addresses unit-level small area estimation problem where the population variances vary across different areas. Independent inverse gamma priors are proposed for these variances, and the corresponding small area predictors are derived. The associated unconditional and conditional mean squared errors are derived. The method is illustrated through simulation as well as with a real data set pertaining to Japanese land prices.

Cet article traite du problème d'estimation pour les petits domaines au niveau des unités où les variances de la population varient entre les différents domaines. Des lois a priori inverse-gamma indépendantes sont proposées pour ces variances et les prédicteurs pour les petits domaines correspondants sont ainsi dérivés. Les erreurs quadratiques moyennes inconditionnelles et conditionnelles qui leur sont associées sont dérivées. Cette méthode est illustrée par simulation ainsi qu'avec des données réelles se rapportant aux prix des terrains japonais.

JIMING JIANG & **ROHOSEN BANDYOPADHYAY** (University of California, Davis)

[Monday June 15 / lundi 15 juin, 14:00]

Benchmarking the Observed Best Predictor

Étalonnage du meilleur prédicteur observé

In the practice of small area estimation, it is often required that estimates of small area means satisfy certain benchmarking conditions. The problem has been extensively studied in the context of empirical best linear unbiased prediction (EBLUP). We extend two ideas of benchmarking the EBLUP to observed best prediction (OBP), a recently proposed method for small area estimation as an alternative to the EBLUP. Those ideas are (i) OBP with self-benchmarking property and (ii) adjusting the OBP. We demonstrate the benchmarking properties as well as performance of the benchmarked OBP, including a measure of uncertainty, via simulation and real data applications.

Lors de l'estimation pour petits domaines, il est souvent nécessaire que l'estimation des moyennes de petits domaines satisfasse certaines conditions de référence. Ce problème a été largement étudié dans le cadre des meilleurs prédicteurs linéaires sans biais empiriques (MPLSBE). Nous développons deux idées d'étalonnage du MPLSBE au meilleur prédicteur observé (MPO), une méthode récemment proposée pour l'estimation dans de petits domaines en tant qu'alternative au MPLSBE. Ces idées sont (i) MPO avec propriété d'auto-étalonnage et (ii) ajuster le MPO. Nous démontrons les propriétés d'étalonnage ainsi que la performance du MPO étalonné, comprenant une mesure d'incertitude, par la simulation et par des applications avec des données réelles.

1D-I4: Recent Developments in Small Area Estimation Développements récents en estimation pour petites régions

GAURI DATTA (University of Georgia), **ADRIJO CHAKRABORTY** (National Opinion Research Center), **ABHYU-DAY MANDAL** (University of Georgia)

[Monday June 15 / lundi 15 juin, 14:30]

Small Area Estimation for Normal Mixture Models

Estimation pour les petits domaines dans les modèles de mélanges normaux

We propose a new class of small area models that includes the standard random effects model of Fay and Herriot, and also the synthetic regression model following non-rejection of a test of no random effects, considered by Datta, Hall and Mandal. These models, which are formed by mixture of certain normal distributions, result in variable shrinkage of direct estimators to the regression estimators. Based on a Bayesian approach and the spike and slab priors for random effects, we propose our new method. Evaluation of the proposed method based on a U.S. Census Bureau application of estimation of poverty ratios and related simulated data shows superiority of the proposed method. We will also discuss other mixture models to handle outliers.

Nous proposons une nouvelle catégorie de modèles pour petits domaines qui inclut le modèle à effets aléatoires de Fay et Herriot, ainsi que le modèle de régression synthétique suivant le non rejet d'un test d'absence d'effets aléatoires examiné par Datta, Hall et Mandal. Ces modèles, qui sont constitués par le mélange de certaines lois normales, entraînent le rétrécissement variable des estimateurs directs à des estimateurs de régression. Nous proposons notre nouvelle méthode basée sur une approche bayésienne et la loi a priori « spike-and-slab » pour les effets aléatoires. L'évaluation de la méthode proposée, qui se base sur une application de l'estimation du ratio de pauvreté du Bureau de recensement américain et d'autres données simulées connexes, démontre la supériorité de la méthode proposée. Nous discuterons aussi d'autres modèles de mélanges pour gérer les valeurs aberrantes.

1D-I5: BFF Inference and Beyond -- Building a Unified Foundation for Statistical Inference
Inférence BFF et au-delà -- les fondements d'une inférence statistique unifiée

Chair/Président: Xiao-Li Meng (Harvard University)

Organizers/Responsables: Xiao-Li Meng (Harvard University) and/et Min-ge Xie (Rutgers University)

Session sponsored by the ICOSA – Canada Chapter

Séance parrainée par le chapitre canadien de l'ICOSA

Room/Salle: Rowe 1011

Abstracts/Résumés

NANCY REID (University of Toronto), **DAVID COX** (University of Oxford)

[Monday June 15 / lundi 15 juin, 13:30]

On some Principles of Statistical Inference

Sur certains principes de l'inférence statistique

Statistical theory aims to provide a foundation for studying the collection and interpretation of data, a foundation that does not depend on the particular details of the substantive field in which the data are being considered. This gives a systematic way to approach new problems, and a common language for summarizing results; ideally the foundations and common language ensure that statistical aspects of one study, or of several studies on closely related phenomena, can be broadly accessible. I will discuss some principles of statistical inference, to outline how these are, or could be, used to inform the interpretation of results, and to provide a greater degree of coherence for the foundations of statistics.

La théorie statistique vise à fournir une base pour l'étude de la collecte et de l'interprétation des données, une base qui ne dépend pas des détails particuliers du champ principal dans lequel les données sont étudiées. Cela fournit une manière systématique d'aborder de nouveaux problèmes et un langage commun pour résumer les résultats. Idéalement, les bases et le langage commun assurent que les aspects statistiques d'une étude ou de plusieurs études sur des phénomènes étroitement liés peuvent être largement accessibles. Je vais discuter certains principes de l'inférence statistique, décrire la manière dont ceux-ci sont, ou pourraient être, utilisés pour éclairer l'interprétation des résultats et fournir une plus grande cohérence dans les bases de la statistique.

JENNY SHI & JAN HANNIG (University of North Carolina at Chapel Hill), **THOMAS LEE & CHU SHING LAI** (University of California at Davis)

[Monday June 15 / lundi 15 juin, 14:00]

A fiducial approach to the estimation of covariance matrix

Une approche fiduciaire à l'estimation de la matrice de covariance

As a classical problem, covariance estimation has drawn much attention from the statistical community for decades. Much work has been done under the frequentist and the Bayesian frameworks. Aiming to quantify the uncertainty of the estimators without having to choose a prior, we have devel-

L'estimation de la covariance est un problème classique qui a attiré l'attention de la communauté statistique depuis des décennies. Une grande partie de ce travail a été fait dans les cadres fréquentiste et bayésien. Dans le but de quantifier l'incertitude des estimateurs sans avoir à choisir une loi a priori, nous avons développé une approche fiduciaire à

1D-I5: BFF Inference and Beyond -- Building a Unified Foundation for Statistical Inference

Inférence BFF et au-delà -- les fondements d'une inférence statistique unifiée

oped a fiducial approach to the estimation of covariance matrix. Built upon the Fiducial Bernstein-von Mises Theorem (Sonderegger and Hannig 2014), we have showed that the fiducial distribution of the covariance matrix is consistent under our framework. The samples from the fiducial distribution enable us to define a meaningful confidence region for the covariance matrix.

l'estimation de la matrice de covariance. En se basant sur le théorème fiduciaire Bernstein-von Mises (Sonderegger et Hannig, 2014), on démontre que la loi fiduciaire de la matrice de covariance est convergente selon notre cadre. Les échantillons tirés de la loi fiduciaire nous permettent de définir une région de confiance cohérente pour la matrice de covariance.

MING-GE XIE (Rutgers University)

[Monday June 15 / lundi 15 juin, 14:30]

Confidence Distribution (CD) and a Unifying Framework for Statistical Inferences

Distribution de confiance (DC) et un cadre unifié pour les inférences statistiques

Confidence distribution (CD) is any sample-dependent distribution function that can represent confidence intervals/regions of all levels for a parameter of interest. Conceptually, it is not different from a point estimator or confidence interval, but uses a sample-dependent function to estimate a parameter. It provides "simple and interpretable summaries of what can reasonably be learned from data" (Cox 2013) and also meaningful answers for all inference questions. An emerging theme is "Any approach, regardless of being frequentist, fiducial or Bayesian, can potentially be unified under the CD concept, provided it can be used to build intervals of all levels." In the talk, we articulate the logic behind and demonstrate CD can potentially serve as a unifying framework for BFF (Bayesian/fiducial/frequentist) inferences.

Une distribution de confiance (DC) est toute fonction de répartition dépendante d'un échantillon qui peut représenter des intervalles de confiance/régions de tous les niveaux pour un paramètre d'intérêt. Conceptuellement, elle n'est pas différente d'un estimateur ponctuel ou d'un intervalle de confiance, mais elle utilise une fonction dépendante de l'échantillon pour estimer un paramètre. Elle fournit des «résumés simples et interprétables de ce qui peut raisonnablement être appris au sujet des données» (Cox 2013) et aussi des réponses utiles à toutes les questions d'inférence. Un thème émergent est «toute approche, peu importe qu'elle soit fréquentiste, fiduciaire ou bayésienne, peut potentiellement être intégrée sous le concept de DC, à condition qu'elle puisse être utilisée pour construire des intervalles à tous les niveaux.» Dans cet exposé, nous en exprimons la logique et démontrons que la DC peut éventuellement servir de cadre unifié pour les inférences BFF (bayésienne / fiduciaire / fréquentiste).

1D-C1: Actuarial Science and Finance 1 Science actuarielle et finance 1

Chair/Présidente: Michelle Xia (Northern Illinois University)

Room/Salle: McCain 2184

Abstracts/Résumés

TOBY KENNEY (Dalhousie University)

[Monday June 15 / lundi 15 juin, 13:30]

Credibility Classification with Missing Data

Classification basée sur la crédibilité avec des données manquantes

Credibility theory combines the information from a small but relevant dataset and a larger but more general dataset. Here the information we wish to extract is a classifier, and different samples represent different sets of predictor variables available. Using all common variables we have a lot of data, but are missing critical predictors. Restricting attention to datapoints with a particular variable, we have less data, so overfit is possible. We apply credibility theory to create a compromise classifier. To achieve this, we develop a new variant of credibility theory combining the features of crossed-classification credibility theory and hierarchical credibility theory.

La théorie de la crédibilité combine l'information d'un petit jeu de données pertinent avec celle d'un plus gros jeu de données plus général. L'information que nous souhaitons extraire dans le cas présent est un classificateur, et divers échantillons représentent différents ensembles de variables prédictives disponibles. L'utilisation de toutes les variables communes nous fournit beaucoup de données, mais il nous manque des prédicteurs essentiels. En mettant l'accent sur les points de données avec une variable en particulier, nous obtenons moins de données, mais il existe un risque de surajustement. Nous appliquons donc la théorie de la crédibilité pour créer un classificateur de compromis. Nous développons une nouvelle variante de la théorie de la crédibilité qui combine les caractéristiques de la théorie de la crédibilité par classification croisée avec la théorie de la crédibilité hiérarchique.

ETIENNE MARCEAU, HÉLÈNE COSSETTE, ITRE MTALAI & DÉRY VEILLEUX (Université Laval)

[Monday June 15 / lundi 15 juin, 13:45]

Discrete-Time Risk Models, with Exchangeable Archimedean Copulas

Modèles de risque en temps discret avec copules archimédiennes échangeables

We consider discrete-time risk models with dependence. The dependence structure among claim amounts is defined with an exchangeable Archimedean copula. Explicit formulas and recursive relations for ruin measures are derived. Applications to the computation of premiums and dynamic risk measures are examined. Examples with Ali-Mikhail-Haq Copula and Frank copula are considered.

On considère des modèles de risque en temps discret avec dépendance. La structure de dépendance entre les montants de sinistres est définie avec une copule archimédienne échangeable. Des formules explicites et des relations récursives pour les mesures de ruine sont dérivées. Des applications au calcul des primes et des mesures de risque dynamiques sont examinées. Des exemples avec les copules d'Ali-Mikhail-Haq et de Frank sont considérés.

1D-C1: Actuarial Science and Finance 1 Science actuarielle et finance 1

ALEXANDER MELNIKOV (University of Alberta)

[Monday June 15 / lundi 15 juin, 14:00]

Financial Markets with Long-range Dependence: Partial Hedging and Statistical Study

Marchés financiers avec dépendance à long terme : couverture de risque partielle et étude statistique

The talk is devoted to a financial market governed by both a Wiener process and a fractional Brownian Motion with the Hurst index greater one half. Efficient hedging problem for a power loss function is solved for the model with the index greater three fourth by means of representation techniques. The problem is successfully studied for the models with the index greater one half using an approximation technique. This research is supported by a statistical study of such mixed markets which demonstrates that there exist a solid body of real financial markets with the high Hurst index.

Cette présentation se concentre sur un marché financier relevant à la fois d'un processus de Wiener et d'un mouvement brownien fractionnaire avec un indice de Hurst plus grand qu'une demie. Un problème de couverture de risque pour une fonction de perte de puissance est résolu avec le modèle pour un indice plus grand que trois-quarts à l'aide de représentations techniques. Le problème est étudié avec succès pour les modèles avec un indice plus grand qu'une demie à l'aide d'une approximation technique. Cette recherche est soutenue par une étude statistique de tels marchés mixtes qui démontre qu'il existe des marchés financiers réels avec un indice de Hurst élevé.

LEALAND MORIN (Capital One)

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

Keeping Diffusion Processes within Bounds: Using Information between Observations

Maintenir des processus de diffusion à l'intérieur des limites : utiliser l'information entre les observations

This paper investigates the estimation of discretely sampled, continuous time models in the presence of information bounding the series between observations. Maximum likelihood estimation of diffusion models are improved in several ways, if the process is known to remain within an interval, either by chance or by constraint. Several types of bounding information are shown to share some key features, stemming from their connection with the heat equation and the reflection principle. The advantages are realized in the form of efficiency gains, bias reduction and in the flexibility of model specification.

Cet article s'intéresse à l'estimation d'échantillonnage discret, continu de modèles de temps en présence de limites d'information pour les séries entre les observations. L'estimation du maximum de vraisemblance des modèles de diffusion est améliorée de diverses manières, si le processus est connu et demeure à l'intérieur d'un intervalle, soit par chance ou par contrainte. Plusieurs types de limites d'information sont démontrés et partagent certaines caractéristiques, provenant de leur lien avec l'équation de la chaleur et le principe de la réflexion. Les avantages sont réalisés sous forme de gain d'efficacité, baisse de biais et dans la flexibilité dans le choix de modèle.

DAVID SAUNDERS (University of Waterloo), **LUNG KWAN TSUI** (HSBC), **SATISH IYENGAR** (University of Pittsburgh)

[Monday June 15 / lundi 15 juin, 14:30]

Lower Tail Independence of Hitting Times of Two-Dimensional Diffusion Processes

Indépendance de l'aile inférieure des temps de frappe de processus de diffusion a deux dimensions

The coefficient of tail dependence measures how extreme events in one component of a bivariate copula depend on extreme events in the other component. It is well known that the Gaussian cop-

Le coefficient de dépendance des ailes mesure comment les événements extrêmes dans une composante d'une copule bivariable dépend d'événements extrêmes dans une autre composante. Il est bien connu que la copule gaussienne

1D-C1: Actuarial Science and Finance 1

Science actuarielle et finance 1

ula has zero tail dependence, a shortcoming for its application in credit risk modeling and quantitative risk management in general. We show that this property is shared by the joint distributions of hitting times of two-dimensional (uniformly elliptic) diffusion processes.

a une dépendance nulle des ailes, une faiblesse pour son application dans la modélisation du risque de crédit et de la gestion quantitative du risque en général. Nous démontrons que cette propriété est partagée par les distributions conjointes des temps de frappe de processus de diffusion à deux dimensions (uniformément elliptique).

1D-C2: Biostatistics in Action 1 Biostatistique en action 1

Chair/Présidente: Celia Greenwood (McGill University)

Room/Salle: McCain 2198

Abstracts/Résumés

JUNIOR CHUANG, WEI HAO (DAVID) DAI & WENDY LOU (University of Toronto), **KELLY SABALIAUSKAS** (Southern Ontario Centre for Atmospheric Aerosol Research, University of Toronto)

[Monday June 15 / lundi 15 juin, 13:30]

Impact of Ultra-Fine Particles on Childhood Development of Lung Functions, Results from Canadian Healthy Infant Longitudinal Development (CHILD) Study

Impact des particules ultrafines sur le développement infantile des fonctions pulmonaires, résultats de l'Étude longitudinale nationale sur le développement des enfants en santé (CHILD)

Many publications have shown the adverse effects of air pollution on lung development of young children. However the particular topic of ultra-fine particles and its association with the development of lung function in children is still not fully understood. Using Land-Use Regression and mixed models as a point of departure, the Toronto cohort from the Canadian Healthy Infant Longitudinal Development Study was combined with ultra-fine particle data over the period of 2009-2013 to identify and describe both acute and longitudinal effects of ultra-fine particles on children.

De nombreuses publications ont montré les effets néfastes de la pollution atmosphérique sur le développement pulmonaire des jeunes enfants. Cependant, le thème particulier des particules ultrafines et de leur association avec le développement de la fonction pulmonaire chez l'enfant n'est pas encore parfaitement compris. Prenant comme point de départ des modèles de régression et mixtes de l'utilisation du territoire, nous combinons la cohorte de Toronto de l'étude longitudinale sur le développement des enfants en santé et les données sur les particules ultrafines pour la période 2009-2013 afin d'identifier et de décrire les effets aigus et longitudinaux des particules ultrafines sur les enfants.

MOHAMMAD EHSANUL KARIM, PAUL GUSTAFSON, JOHN PETKAU, HELEN TREMLETT & THE BEAMS STUDY GROUP (University of British Columbia)

[Monday June 15 / lundi 15 juin, 13:45]

Comparison of Statistical Approaches Dealing with Immortal Time Bias in Drug Effectiveness Studies

Comparaison d'approches statistiques qui traitent de biais d'immortalité dans les études d'efficacité de médicaments

Immortal time bias can be minimized by using a time-dependent Cox model analysis, but the underlying assumptions may be difficult to assess. Methods, such as prescription time-distribution matching (PTDM) and the sequential Cox approach, have been proposed to avoid the need for a time-dependent Cox analysis. We compared the performance of these approaches using simulated survival data. Our simulations revealed that the PTDM approach is not effective in addressing immortal time bias, but the sequential Cox approach is more useful. All of these approaches were applied to in-

Le biais d'immortalité peut être minimisé en utilisant un modèle d'analyse Cox dépendant du temps, mais les hypothèses sous-jacentes peuvent être difficiles à évaluer. Les méthodes, telles que l'ajustement du temps de la prescription et sa distribution (ATPD) et l'approche séquentielle Cox, ont été proposées afin d'éviter le besoin d'une analyse Cox dépendante du temps. Nous avons comparé la performance de ces approches en utilisant des données de survie simulées. Nos simulations ont démontré que l'approche ATPD n'est pas efficace pour adresser le biais d'immortalité, mais que l'approche séquentielle Cox est plus utile. Toutes ces approches ont été appliquées afin d'étudier l'impact d'un traite-

1D-C2: Biostatistics in Action 1

Biostatistique en action 1

investigate the impact of beta-interferon treatment in delaying multiple sclerosis disability progression.

ment d'interféron bêta à retarder la progression de l'invalidité de la sclérose en plaques.

ALISHA ALBERT-GREEN (University of Western Ontario), **JOHN BRAUN** (University of British Columbia Okanagan), **CHARMAINE DEAN** (University of Western Ontario)

[Monday June 15 / lundi 15 juin, 14:00]

A General Framework for the Joint Modelling of Multivariate Zero-Inflated Spatial Processes

Un cadre général pour la modélisation conjointe des processus spatiaux multivariés à surreprésentation de zéros

Health and environmental data commonly consist of multivariate counts with excess zeros, where the multivariate outcomes may be spatially related. This research develops a general framework for the joint modelling of multivariate spatial zero-inflated count data where the outcomes are linked through a common random effect. Here we assume that latent spatial Gaussian random fields link various aspects of the model components, including the zero components and the mean components for each outcome, as well as across outcomes. We apply these models to zero-heavy lung and bronchus cancer counts for males and females in Ontario's public health units in 2010.

Les données sur la santé et l'environnement sont généralement constituées d'un dénombrement multivarié avec un excès de zéros, où les résultats multivariés peuvent être liés dans l'espace. Cette recherche développe un cadre général pour la modélisation conjointe de données de dénombrement multivariées spatiales à surreprésentation de zéros où les réponses sont liées par un effet aléatoire commun. Nous supposons ici que des champs gaussiens aléatoires spatiaux latents relient différents aspects des composantes du modèle, dont les zéros et la moyenne pour chaque réponse, ainsi qu'à travers les réponses. Nous appliquons ce modèle au dénombrement important en zéros des cancers du poumon et des bronches pour les hommes et les femmes des services de santé publiques de l'Ontario en 2010.

WEI HAO (DAVID) DAI (University of Toronto), **ZIHANG LU** (The Hospital For Sick Children), **WENDY LOU** (University of Toronto)

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

Composite Indices for High-Dimensional Questionnaire Data

Indices composites pour des données de questionnaire de haute dimension

In large-scale cohort studies, questionnaires are often used to capture multi-dimensional characteristics of measures of interest, such as indoor mold exposure for infants with asthma. In an attempt to develop composite indices as practical solutions for dealing with considerably large numbers of highly correlated variables and domains, we consider a few commonly used dimension-reduction approaches, including principal component analysis and canonical correlation analysis. We present empirical comparisons via an illustrating example using an indoor home environmental questionnaire from the Canadian Healthy Infant Longitudinal Development (CHILD) Study.

Dans les études de cohorte à grande échelle, les questionnaires sont souvent utilisés pour saisir des caractéristiques multidimensionnelles des mesures d'intérêt, comme l'exposition aux moisissures intérieures chez les enfants atteints d'asthme. En vue de développer des indices composites comme solutions pratiques pour gérer le nombre considérablement élevé de variables et de domaines étroitement corrélés, nous examinons quelques approches de réduction de la dimension couramment utilisées, y compris l'analyse en composantes principales et l'analyse de corrélation canonique. Nous présentons des comparaisons empiriques à l'aide d'un exemple de questionnaire portant sur le milieu de vie intérieur dans le cadre de l'Étude longitudinale nationale sur le développement des enfants en santé (CHILD).

1D-C2: Biostatistics in Action 1 Biostatistique en action 1

XUAN LI, HANNA JANKOWSKI, XIAOGANG WANG & JANE HEFFERNAN (York University)

[Monday June 15 / lundi 15 juin, 14:30]

A Method for Clustering Hemagglutinin Influenza Protein Sequences

Une méthode pour regrouper les séquences de protéines hémagglutinines de la grippe

Seasonal influenza strains contain a high level of variability in the Hemagglutinin (HA) glycoprotein. A better understanding of HA evolution over influenza seasons assists to better advice vaccine strain development. We apply a new clustering method to the highly variable influenza A H3N2 HA viral glycoprotein. Our data comprises 1960 viral protein sequences active from 1998 to 2012, and our methodology aggregates these sequences into 23 clusters. Based on these clusters, we provide an investigation of past vaccines and the dominant cluster in each influenza season. We also investigate evolutionary pressures of closely matched circulating and vaccine strains HA glycoproteins.

Les souches de la grippe saisonnière présentent une grande variabilité dans la glycoprotéine hémagglutinine (HA). Une meilleure compréhension de l'évolution de HA pendant les saisons de la grippe permet de donner de meilleurs conseils pour le développement de la souche vaccinale. Nous appliquons une nouvelle méthode de groupement à la glycoprotéine virale de la grippe extrêmement variable A H3N2 HA. Nos données incluent 1960 séquences de protéines virales actives de 1998 à 2012, et notre méthodologie rassemble ces séquences en 23 grappes. À partir de ces grappes, nous présentons une étude de vaccins antérieurs et de la grappe dominante lors de chaque saison de la grippe. Nous examinons aussi les pressions d'évolution des souches circulantes et vaccinales glycoprotéines HA étroitement appariées.

AHALYA SIVATHAYALAN & PATRICK FARRELL (Carleton University), **DANIEL KREWSKI** (University of Ottawa)

[Monday June 15 / lundi 15 juin, 14:45]

Empirical Comparisons of the Signal-to-Noise Crossover Dose and the Benchmark Dose

Comparaisons empiriques de dose de croisement signal-bruit et la dose de référence

An important aspect of dose-response analysis is the identification of an appropriate point of departure (PoD) for the development of a human exposure guidelines to toxic substances. Although the benchmark dose (BMD) is the most widely used PoD, estimation of the BMD requiring extrapolation beyond dose range can be model dependent. This limitation is overcome by signal-to-noise crossover dose (SNCD) (Sand et al., 2011). Using toxicological datasets from the US National Toxicology Program (NTP) database, empirical comparisons between the BMD and the SNCD are made, as well between lower confidence limits (BMDL, and SNCDL) on these two PoD's.

Un aspect important de l'analyse dose-réponse est l'identification d'un point de départ approprié (PoD) pour le développement de directives d'exposition humaine aux substances toxiques. Bien que la dose référence soit le point de départ le plus utilisé, l'estimation de la dose de référence nécessitant une extrapolation au-delà de l'intervalle de dose peut être dépendante du modèle. Cette limite est surmontée par la dose de croisement signal-bruit (Sand et al., 2011). En utilisant des données toxicologiques de la base de données du US National Toxicology Program (NTP), des comparaisons empiriques entre la dose référence et la dose de croisement signal-bruit sont établies, ainsi qu'entre les limites inférieures de confiance sur ces deux points de départ.

1D-C3: Biostatistics: Dealing with Incomplete Data Biostatistique : comment traiter les données incomplètes

Chair/Président: Joel Dubin (University of Waterloo)

Room/Salle: McCain 2116

Abstracts/Résumés

LIQUN DIAO, RICHARD COOK & LEILEI ZENG (University of Waterloo)

[Monday June 15 / lundi 15 juin, 13:30]

Vine Copula Models for Recurrent Exacerbations in Chronic Disease

Modèles de copules en vignes pour les exacerbations récurrentes dans les maladies chroniques

In some chronic disease settings, processes are not progressive but rather exhibit a continual risk of periodic episodic conditions. Examples include asthma, infectious disease, and psychotic disorders. For each of these examples, the recurrent events have non-ignorable durations associated with them and are better characterized as recurrent exacerbations. We assume that the dependence structure of successive gap times is induced by a D-vine copula model. We propose inverse probability weighted composite likelihood to correct biases due to induced dependent censoring, which builds a bridge to the standard practice in the context of vine copula when complete data are available.

Pour quelques maladies chroniques, les processus ne sont pas progressifs mais démontrent plutôt un risque constant d'états épisodiques périodiques. À titre d'exemple, il y a l'asthme, les maladies infectieuses et les troubles psychotiques. Pour chacun de ces exemples, les événements récurrents sont de durées non ignorables et sont mieux décrits en tant qu'exacerbations récurrentes. Nous supposons que la structure de dépendance des intervalles de temps successifs sont induits par un modèle de copules en vignes D. Nous proposons une vraisemblance composite à pondération selon la probabilité inverse pour corriger les biais dus à la censure dépendante induite, ce qui établit un lien avec la pratique standard dans le contexte des copules en vignes quand des données complètes sont disponibles.

YAYUAN ZHU, CECILIA COTTON & JERRY LAWLESS (University of Waterloo)

[Monday June 15 / lundi 15 juin, 13:45]

Analysis of Complex Event Outcome Data from Disease Clinics

Analyse des données de résultat d'un événement complexe pour des maladies cliniques

Studies based on disease clinic data often face several complications. Patients visit the clinic irregularly, and the times may depend on disease-related variables; this can cause event time outcomes to be dependently interval-censored. Event times may also be left-truncated by clinic enrolment time. Finally, treatments assigned to patients may depend on disease-related variables. A doubly weighted GEE approach will be proposed to estimate parameters in marginal structural models for outcomes given treatment, adjusting for dependent visit times and measured confounders. Simulation studies are conducted to examine the pro-

Les études fondées sur les données cliniques de maladies sont souvent confrontées à plusieurs complications. Les patients visitent la clinique sur une base irrégulière et les moments peuvent dépendre des variables liées à la maladie ; ce qui peut entraîner le fait que les résultats de l'événement dans le temps soient dépendantes censurées par intervalle. Le moment des événements peut également être tronqué à gauche selon le moment d'inscription à la clinique. Enfin, les traitements attribués aux patients peuvent dépendre des variables liées à la maladie. Une approche doublement pondérée GEE sera proposée pour estimer les paramètres des modèles structurels marginaux pour les résultats étant donné le traitement et ajustée selon les moments des visites dépen-

1D-C3: Biostatistics: Dealing with Incomplete Data Biostatistique : comment traiter les données incomplètes

posed methods. The Toronto Psoriatic Arthritis Cohort Study is used for illustration.

dantes et les facteurs confondants mesurés. Des études de simulation sont menées afin d'examiner les méthodes proposées. L'étude de cohorte de l'arthrite psoriasique de Toronto est utilisée à des fins d'illustration.

MICHAEL MCISAAC (Queen's University)

[Monday June 15 / lundi 15 juin, 14:00]

Designing Studies with Incomplete Data: An Adaptive Approach to Optimal Response-Dependent Mean Score Sampling

Concevoir des études avec des données incomplètes: une approche adaptative au score moyen d'échantillonnage dépendant de la réponse optimale

I will introduce an adaptive two-phase sampling design which incorporates an internal pilot study in order to allow for approximation of optimal sampling for mean score analysis without the need for a potentially expensive external pilot study. This adaptive procedure can be implemented in practice and retains the advantages of the theoretically optimal mean score design, which has been shown to be easy to implement and to result in efficiency gains for a variety of estimators in many settings. I derive the optimal selection probabilities and demonstrate the implementation of this design using simulation studies motivated by studies in rheumatology.

Je vais présenter un plan d'échantillonnage adaptatif en deux phases qui incorpore une étude pilote interne afin de permettre l'approximation de l'échantillonnage optimal pour une analyse de score moyen sans avoir besoin de recourir à des études pilotes externes potentiellement coûteuses. Cette procédure adaptative peut être mise en place en pratique et maintient les avantages du plan de score moyen optimal théorique, qui a été démontré facile à mettre en place et qui résulte en un gain d'efficacité pour un éventail d'estimateurs dans plusieurs situations. Je dérive les probabilités de sélection optimales et démontre la mise en place de ce plan en utilisant des études de simulations motivées par des études en rhumatologie.

CHITHRAN VASUDEVAN (Memorial University), **ASOKAN MULAYATH VARIYATH & ZHAO ZHI FAN** (Memorial University)

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

Censored Quantile Regression with the Effective Usage of Auxiliary Information

Régression quantile censurée avec l'utilisation efficace de l'information auxiliaire

In survival data analysis, it is possible to have auxiliary information about population from previous clinical trials or from existing facts. This information could be used to improve the efficiency of estimators. Using estimating equation approach, we utilize the auxiliary information to develop the empirical likelihood based data driven probability weights. Our simulation studies show that estimator have better efficiency by incorporating these weights into the linear censored quantile regression model. The proposed estimator is consistent and has an asymptotically normal distribution.

Lors d'analyse de données de survie, il est possible d'avoir des informations auxiliaires sur les populations provenant d'essais cliniques antérieurs ou de faits connus. Ces informations pourraient être utilisées afin d'améliorer l'efficacité des estimations. En utilisant l'approche d'équations d'estimation, nous utilisons l'information auxiliaire afin de développer la vraisemblance empirique basée sur les poids de probabilités selon les données. Nos études de simulation démontrent que les estimateurs ont une meilleure efficacité en incorporant ces poids dans le modèle linéaire de régression quantile censurée. L'estimateur proposé est conforme est a une loi normale asymptotique.

1D-C3: Biostatistics: Dealing with Incomplete Data Biostatistique : comment traiter les données incomplètes

YING WU & RICHARD COOK (University of Waterloo)

[Monday June 15 / lundi 15 juin, 14:30]

Assessing the Accuracy of Predictive Models with Interval-censored Data

Évaluation de la précision des modèles prédictifs sur des données censurées par intervalle

We consider the challenge of assessing the accuracy of a predictive model when response times are interval-censored. Inverse probability weighted (IPW) estimators are developed and evaluated based on the mean prediction error and the area under the receiver operating characteristic curve. The weights are estimated from a multistate model which jointly considers the event process and the inspection and censoring processes. We empirically investigate the performance of the proposed methods and illustrate their application in the context of a motivating rheumatology study in which HLA markers are used for predicting disease progression in psoriatic arthritis.

Nous considérons la difficulté d'évaluer la précision d'un modèle prédictif quand les temps de réponse sont censurés par intervalle. Les estimateurs pondérés selon la probabilité inverse (PPI) sont développés et évalués en fonction de l'erreur moyenne de prédiction et de l'aire sous la courbe ROC. Les poids sont estimés à partir d'un modèle multi-états qui considère conjointement le processus de l'événement et les processus d'inspection et de censure. Nous étudions empiriquement la performance des méthodes proposées et illustrons leur application dans le contexte d'une étude motivante de rhumatologie dans laquelle les marqueurs HLA sont utilisés pour prédire la progression de la maladie dans l'arthrite psoriasique.

CHRISTOPHER A. GRAVEL & PATRICK FARRELL (Carleton University), **DANIEL KREWSKI** (University of Ottawa), **ANUP DEWANJI** (Indian Statistical Institute)

[Monday June 15 / lundi 15 juin, 14:45]

A Validation Sampling Approach for Unbiased Estimation in Misclassified Right Censored Continuous Time Survival Data

Méthode de validation par échantillonnage pour estimation non biaisée de données de survie continues dans le temps censurées à droite mal classifiées

For right censored continuous time survival data, failing to observe the event of interest can introduce misclassification bias in risk estimation. Incorrectly observing cause-specific event types at correctly recorded event times can also introduce bias. An internal validation sampling approach is used to update a set of parametric likelihoods to produce unbiased estimates in each of these scenarios. The use of these approaches is illustrated using electronic health records data under realistic simplifying assumptions. Finally, these approaches are validated through large simulation studies.

Pour les données de survie continues dans le temps censurées à droite, la non-observation de l'événement d'intérêt peut engendrer un biais dû à l'erreur de classification dans l'estimation de risques. Une mauvaise observation des types d'événements par cause à des temps d'événements enregistrés correctement peut également engendrer un biais. Une méthode de validation interne par échantillonnage est utilisée pour mettre à jour une série de vraisemblances paramétriques permettant de produire des estimations non biaisées dans chacun de ces scénarios. L'utilisation de ces approches est illustrée par des données de dossiers de santé électroniques selon des hypothèses simplificatrices. Enfin, ces approches sont validées par de grandes études de simulation.

1D-C4: Foundations and Nonparametrics Fondements et statistique non paramétrique

Chair/Présidente: Mireille Schnitzer (Université de Montréal)

Room/Salle: McCain 2102

Abstracts/Résumés

ASMA BAHAMYIROU & ÉRIC MARCHAND (Université de Sherbrooke)

[Monday June 15 / lundi 15 juin, 13:30]

On the Discrepancy between Bayes Credibility and Frequentist Probability of Coverage

Sur le décalage entre la crédibilité bayésienne et la probabilité de couverture fréquentiste.

Suppose X is a random variable from a normal distribution with unknown mean $\mu \in [-m, m]$. Let $I_\pi(x)$ be a Bayesian credible set associated with a prior π on $[-m, m]$. The question which arises here is it possible to choose a prior π to get a frequentist coverage probability near to $1 - \alpha$? We show that this is an insoluble question while m and α are small. We exhibit situations of pronounced discrepancy between the credibility of Bayes credible regions and frequentist coverage. Frequentist confidence intervals are shown to have credibility equals one in some cases when $\mu < \Phi^{-1}(1 - \alpha)$.

Soient $X \sim N(\mu, \sigma^2)$, $\mu \in [-m, m]$ inconnue et $I_\pi(x)$ un intervalle de crédibilité bayésien associé à la loi a priori π sur $[-m, m]$. Une question est de savoir si π peut être choisie pour que la probabilité de couverture fréquentiste de $I_\pi(x)$ soit proche de $1 - \alpha$. Nous démontrerons qu'il est illusoire d'essayer de trouver une solution tant et si longtemps que m et α sont petits et nous exhiberons un décalage entre la crédibilité bayésienne et la probabilité de couverture fréquentiste. Nous montrerons que certaines méthodes fréquentistes donnent des crédibilités égales à 1 pour certains cas lorsque $\mu < \Phi^{-1}(1 - \alpha)$.

WEILI LI (Dalla Lana School of Public Health, University of Toronto), **LISA STRUG** (The Hospital for Sick Children)

[Monday June 15 / lundi 15 juin, 13:45]

Extending the Evidential Paradigm to Accommodate Composite Hypotheses

L'extension du paradigme de la preuve pour tenir compte des hypothèses composites

The Evidential Paradigm (EP) uses the likelihood ratio evaluated at two simple hypotheses to measure statistical evidence; however, the requirement of simple hypotheses can limit its applications. Bickel (2012) and Zhang and Zhang (2013) proposed to extend EP to accommodate composite hypotheses by using a Generalized Likelihood Ratio (GLR). Here we analytically derive the operational characteristics associated with GLR for normal data and fixed-dimensional parametric models asymptotically. We show that the probability of observing a large GLR in support of an incorrect hypothesis is bounded and small. We also provide guidelines for study planning, specifically, sample size calculations, under this framework.

Pour mesurer l'évidence statistique, le paradigme de la preuve utilise le rapport de vraisemblance évalué en fonction de deux hypothèses simples. Cependant, l'utilisation d'hypothèses simples peut limiter les applications du paradigme. Bickel (2012) et Zhang et Zhang (2013) ont proposé d'étendre ce paradigme pour tenir compte des hypothèses composites en utilisant le rapport de vraisemblance généralisé. Ici, nous dérivons analytiquement les caractéristiques asymptotiques opérationnelles associées au rapport de vraisemblance généralisé pour les données normales et les modèles paramétriques dimensionnels fixes. Nous montrons que la probabilité d'observer des rapports de vraisemblance généralisés supportant une hypothèse incorrecte est bornée et petite. Dans ce cadre de travail, nous fournissons également des lignes directrices pour la planification d'études, notamment pour les calculs de taille d'échantillons.

1D-C4: Foundations and Nonparametrics Fondements et statistique non paramétrique

FÉLIX CAMIRAND LEMYRE & TAOUFIK BOUEZMARNI (Université de Sherbrooke), **JEAN-FRANÇOIS QUESSY** (Université du Québec à Trois-Rivières)

[Monday June 15 / lundi 15 juin, 14:00]

Conditional Distributions from Top to Bottom.

Les distributions conditionnelles de long en large.

The stochastic behavior of a random vector might be influenced by a covariate. One way to characterize its conditional distribution is through conditional distribution functions. In this context, a nonparametric estimator can be used. However its weak limit depends on unknown quantities. To solve this problem, two multiplier bootstrap strategies are here proposed. The adaptation of the two strategies to conditional copulas is next discussed. Finally an extension of these results to the serial case is proposed.

Le comportement d'un vecteur aléatoire est parfois sujet à l'influence d'une covariable. Sa distribution conditionnelle peut être caractérisée par l'étude de sa fonction de répartition conditionnelle. Dans ce contexte, il est possible d'utiliser un estimateur non paramétrique. Toutefois, sa loi limite dépend de quantités inconnues. Pour résoudre ce problème, deux stratégies bootstrap de type multiplicateur seront proposées durant cet exposé. Leur adaptation aux copules conditionnelles est ensuite abordée. Enfin, une extension de ces résultats au cadre de données sérielles est discutée.

MIKELIS BICKIS (University of Saskatchewan)

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

Overthrowing the Ω Delusion and Making Probability Pointless

Bouleverser le délire d' Ω par la probabilité sans points

Classical probability begins with a sample space of outcomes and defines events as measurable subsets of this set. The meaning of an outcome being problematic, the event constitutes a more intuitive primitive concept. A collection of events extends to a Boolean algebra, and non-measurable sets are simply avoided. Random variables can similarly be defined as primitive objects whose intuitive meaning is an unknown number. Providing for arithmetic and order operations, random variables form a lattice algebra whose idempotents are events. Lower expectations can be defined by postulating a cone of random variables that are expected to be positive.

La théorie classique des probabilités commence avec un ensemble échantillonnal de résultats et définit les événements par des sous-ensembles mesurables. L'événement pourtant est un concept plus intuitif à cause de la signification problématique du terme « résultat ». Une collection d'événements s'étend à une algèbre de Boole, évitant ainsi les ensembles non-mesurables. On définit les variables aléatoires comme des objets primitifs avec le sens intuitif d'un nombre inconnu. Fournies des opérations d'arithmétique et d'ordre, les variables aléatoires constituent une algèbre treillis dont les éléments idempotents sont des événements. L'espérance inférieure est définie au moyen d'un cône de variables aléatoires qui sont prévues d'être positives.

ALEXANDRE LEBLANC (University of Manitoba)

[Monday June 15 / lundi 15 juin, 14:30]

Local Polynomial Regression: A Bernstein Polynomial Perspective

Régression polynomiale locale: une approche basée sur les polynômes de Bernstein

For nonparametric regression, local polynomial kernel methods are appealing and well understood. We here look at local polynomial smoothing from a

Dans un contexte de régression non paramétrique, les estimateurs à noyaux polynomiaux locaux sont populaires et ont été largement étudiés. Nous jetons ici un regard légèrement

1D-C4: Foundations and Nonparametrics Fondements et statistique non paramétrique

slightly different, yet insightful perspective. Specifically, in order to estimate the regression function g at a point x_0 , we construct a local polynomial estimator by considering the Bernstein polynomial of the remainder of the Taylor expansion of g at x_0 . This results in a local polynomial estimator based on a family of asymmetric kernels.

différent, mais instructif, sur le lissage à l'aide de polynômes locaux. Spécifiquement, nous discutons de la construction d'un estimateur par polynômes locaux d'une fonction de régression g à un point x_0 basée sur une approximation du reste d'une série de Taylor de g au point x_0 à l'aide des polynômes de Bernstein. Le résultat est un estimateur par polynômes locaux basé sur une famille de noyaux asymétriques.

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

[Monday June 15 / lundi 15 juin, 14:45]

Bernstein Conditional Density Estimation with Application to Conditional Distribution and Regression Functions
Estimation de la densité conditionnelle à l'aide des polynômes de Bernstein avec application aux lois conditionnelles et aux fonctions de régression

We first discuss smooth nonparametric estimation of a conditional probability density function based on a Bernstein polynomial representation. Our estimator can be written as a finite mixture of beta densities with data-driven weights. We derive the basic asymptotic properties of the proposed estimator. We then move on to study the distribution function and conditional mean estimators that can be derived from the considered conditional density estimator.

Nous abordons d'abord l'estimation non paramétrique lisse d'une fonction de densité de probabilité conditionnelle basée sur une représentation polynomiale de Bernstein. Notre estimateur peut être décrit comme un mélange fini de densités bêta avec une pondération fondée sur les données. Nous obtenons les propriétés asymptotiques de base de cet estimateur, puis nous étudions la fonction de répartition et les estimateurs de la moyenne conditionnelle qui peuvent être obtenus de l'estimateur de densité conditionnelle proposé.

1E-I1: Big Data Analysis: Methodology and Applications Analyse des données volumineuses : méthodologie et applications

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

Room/Salle: McCain - Scotiabank

Abstracts/Résumés

HAO CHAI & MA SHUANGGE (Yale University)

[Monday June 15 / lundi 15 juin, 15:30]

Identifying Gene-Environment Interactions using a Robust Penalization Approach

Identification des interactions gène-environnement à l'aide d'une approche de pénalisation robuste

Contaminated samples are commonly encountered in high-dimensional genetic data. Traditional least-squares based methods may fail due to outliers. We propose a robust estimator for the high-dimensional accelerated failure time model. We adopt the exponential squared loss to ensure the robustness and the penalization for the identification of genes and gene-environment interactions. The proposed method is implemented with the coordinate-descent and MM algorithms. Simulation results show that the proposed method outperforms the existing methods in terms of variable identification when the data are contaminated. We apply our method to a cancer data set with environmental variables and gene expressions. Our analysis discovers different genes and interactions from the non-robust alternatives. Some of the identified genes are also discovered by other works.

Dans les données génétiques de grande dimension, on rencontre couramment des échantillons contaminés. Or les méthodes des moindres carrés traditionnelles échouent souvent en la présence de valeurs aberrantes. Nous proposons un estimateur robuste pour le modèle du temps de défaillance accéléré de grande dimension. Nous adoptons la perte exponentielle au carré pour garantir la robustesse et la pénalisation pour l'identification des gènes et des interactions gène-environnement. La méthode proposée est appliquée avec les algorithmes de descente de coordonnées et MM. Les résultats de la simulation montrent que cette méthode est supérieure aux méthodes existantes pour ce qui est de l'identification des variables quand les données sont contaminées. Nous appliquons notre méthode à un ensemble de données sur le cancer avec des variables environnementales et des expressions génétiques. Notre analyse permet de découvrir des gènes et des interactions différents des alternatives non robustes. Certains gènes ainsi identifiés avaient aussi été découverts par d'autres travaux.

FAROUK NATHOO (University of Victoria)

[Monday June 15 / lundi 15 juin, 16:00]

A High-Dimensional State-Space Model for the Joint Analysis of EEG and MEG Data

Modèle spatiotemporel de grande dimension pour l'analyse commune de données EEG et MEG

I present a state-space model for solving the neuroelectromagnetic inverse problem based on a joint model combining EEG and MEG data. This inverse problem is ill-posed and leads to a system of underdetermined dynamic models parameterized over a high-dimensional state-space. We assume linear non-Gaussian dynamics for the neural activity that incorporate spatial dependence across neighbouring brain locations as well as bilateral dependence

Je présente un modèle spatiotemporel permettant de résoudre le problème inverse neuroélectromagnétique fondé sur un modèle conjoint combinant des données EEG et MEG. Ce problème inverse est mal posé et conduit à un système de modèles dynamiques sous-déterminés paramétrés sur un espace-temps de grande dimension. Nous parlons de l'hypothèse d'une dynamique non gaussienne linéaire pour l'activité neuronale, avec une dépendance spatiale entre régions voisines du cerveau et une dépendance

1E-I1: Big Data Analysis: Methodology and Applications

Analyse des données volumineuses : méthodologie et applications

corresponding to symmetric locations on opposite hemispheres of the brain. Modeling neural activity over a large number of spatial locations and time points leads to challenges for Bayesian computation. I will discuss solutions based on mean-field variational approximations incorporating the Kalman filter along with parallel computing implementations.

bilatérale entre les régions symétriques des hémisphères du cerveau. La modélisation de l'activité neuronale sur de nombreuses régions spatiales et à de nombreux moments pose des problèmes pour le calcul bayésien. Je discute de solutions qui s'appuient sur des approximations variationnelles de champ moyen avec le filtre de Kalman et des mise en œuvre en calcul parallèle.

ANAND VIDYASHANKAR (George Mason University), **AHMED EJAZ** (Brock University)

[Monday June 15 / lundi 15 juin, 16:30]

Robust Inference for Conditionally Specified Models via Divergences

Inférence robuste pour les modèles à spécifications conditionnelles via la divergence

In several instances of statistical practice it is common, that a pre-test of a subset of parameters is made and based on the results, an appropriate model is chosen. However, when evaluating the resulting inferential procedure the variability induced by model-selection need to be taken into account. Pre-test estimators have been shown to strike a balance between bias and efficiency in the context of maximum likelihood. We revisit this issue from the perspective of model mis-specification and provide a systematic analysis of pre-test procedure via divergence theory. We establish the properties of the pre-test minimum divergence estimators and provide analyses of robustness aspects. We illustrate the usefulness of data-splitting methods and describe links to post-model selection inference.

Dans la pratique statistique il est commun d'effectuer un test préliminaire sur un sous-ensemble de paramètres puis de choisir un modèle approprié en fonction des résultats de ce test. Cependant, pour évaluer la procédure d'inférence résultante, il faut tenir compte de la variabilité induite par la sélection d'un modèle. Il a été démontré que les estimateurs des tests préliminaires permettent d'établir un équilibre entre biais et efficacité dans le contexte du maximum de vraisemblance. Nous réexaminons cette question sous l'angle des erreurs de spécification du modèle et donnons une analyse systématique de la procédure de test préliminaire par la théorie de la divergence. Nous déterminons les propriétés des estimateurs de la divergence minimum du test préliminaire et offrons des analyses de l'aspect robustesse. Nous illustrons l'utilité des méthodes de fractionnement de données et décrivons les liens avec l'inférence après la sélection d'un modèle.

1E-I2: Methodological developments that could impact the practice of statistics Avancées méthodologiques pouvant influencer la pratique de la statistique

Organizer and Chair / Responsable et président: Fernando Camacho (Damos Inc.)

Session sponsored by the Accreditation Committee

Séance parrainée par le Comité d'accréditation

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

RICHARD COOK (University of Waterloo)

[Monday June 15 / lundi 15 juin, 15:30]

Adaptive Clinical Trials: Statistical Methods and Practical Considerations.

Essais cliniques adaptatifs: méthodes statistiques et considérations pratiques.

The need for rigorous and efficient evaluation of experimental interventions has led to increased interest in adaptive designs of randomized clinical trials. In this talk I will review adaptive allocation schemes, methods for adaptive sample size determination, group sequential methods, and other forms of adaptive design. The assumptions, statistical theory and practical issues will be discussed with illustrations provided.

Le besoin d'une évaluation rigoureuse et efficace des interventions expérimentales a donné lieu à un intérêt accru pour les structures adaptatives des essais cliniques aléatoires. Dans cet exposé, je vais passer en revue des schémas d'allocation adaptatifs, des méthodes adaptatives pour la détermination de la taille d'un échantillon, des méthodes séquentielles groupées, et d'autres formes de structures adaptatives. Les hypothèses, la théorie statistique et les problèmes pratiques seront discutés et examinés à l'aide d'illustrations.

LEHANA THABANE (McMaster University)

[Monday June 15 / lundi 15 juin, 16:15]

Issues in Vaccine Trials: Recent Debate, Methodological Challenges and the Ethics

Enjeux dans les essais de vaccin: débat récent, défis méthodologiques et éthique

With emergence of measles and recent outbreak of ebola in West Africa, vaccination to protect individuals and populations has been a topical issue in Canada and around the world. Using vaccine trials as a case study, I will discuss some key methodological issues in trials including: the ethical dilemmas in vaccine trials; selection of outcomes -- measures of vaccine protection or efficacy; the choice of design (superiority vs non-inferiority); and application of intention-to-treat principle in analysis of phase III trials. I will also explore some of the recent key controversies regarding measles and ebola vaccine trials.

Suite à l'émergence de la rougeole et l'épidémie d'Ebola en Afrique de l'Ouest, la vaccination pour la protection des individus et des populations est devenue un enjeu actuel au Canada et autour du monde. En utilisant les essais de vaccin comme étude de cas, je vais discuter de quelques questions méthodologiques essentielles dans les essais : les dilemmes moraux dans les essais de vaccin ; la sélection des réponses -- mesures de protection vaccinale ou efficacité ; le choix de conception (supériorité vs non-infériorité) ; et l'application du principe de l'intention de traiter dans l'analyse des essais de phase III. Je vais aussi explorer les principales controverses récentes concernant les essais de vaccin pour la rougeole et l'Ebola.

1E-I3: Novel Assessment Methods in Undergraduate Education Nouvelles méthodes d'évaluation dans le premier cycle d'études

Chair/Présidente: Alison Gibbs (University of Toronto)

Organizer/Responsable: Bruce Dunham (University of British Columbia)

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: Rowe 1011

Abstracts/Résumés

SUSAN KINNIBURGH, BREE WILTON & SUSAN CHEN (Camosun College)

[Monday June 15 / lundi 15 juin, 15:30]

Outcomes of an Introductory Statistics Course Using WeBWork Online Assessment

Résultats d'un cours d'introduction à la statistique utilisant l'évaluation en ligne WeBWork

This year we introduced WeBWork, a free, open-source online homework application, to two introductory statistics courses. The application provides a bank of questions that can be easily modified to fit individual course needs. WeBWork replaces a costly publisher-based site and supplements hand-written assignments, allowing students to regularly practice material and receive immediate feedback. This presentation will discuss results from a student satisfaction survey, comparison to publisher-based sites, as well as general observations and experiences with the application.

Cette année, nous avons introduit WeBWork, une application de devoirs en ligne, gratuite, de code libre à deux cours d'introduction à la statistique. L'application fournit une banque de questions pouvant facilement être modifiées pour répondre aux besoins de cours individuels. WeBWork remplace un site coûteux d'un éditeur et complète les devoirs écrits, permettant aux étudiants de s'exercer régulièrement et de recevoir une rétroaction immédiate. Cette présentation portera sur les résultats d'une enquête de satisfaction des étudiants, la comparaison aux sites des éditeurs ainsi que des observations générales et des expériences avec l'application.

BRUCE DUNHAM (University of British Columbia)

[Monday June 15 / lundi 15 juin, 16:00]

Using Two-Stage Examinations in Undergraduate Courses

Utilisation d'examens à deux degrés dans les cours de premier cycle

A two-stage examination comprises an individual component and a test that students complete in small groups. The group test may be identical to that completed individually. The rationale for using such an assessment tool is explored, along with suggestions for implementation. Advantages and disadvantages of the approach are discussed, as are examples and anecdotes from three upper-level courses.

Un examen à deux degrés comprend une composante individuelle et un test que les élèves complètent en petits groupes. Le test de groupe peut être identique à celui complété individuellement. La justification de l'utilisation d'un tel outil d'évaluation est étudiée ainsi que des suggestions pour la mise en œuvre. Les avantages et inconvénients de l'approche sont discutés, de même que des exemples et anecdotes de trois cours de niveau supérieur.

1E-I3: Novel Assessment Methods in Undergraduate Education Nouvelles méthodes d'évaluation dans le premier cycle d'études

DOUGLAS WHITAKER & TIM JACOBBE (University of Florida)

[Monday June 15 / lundi 15 juin, 16:30]

LOCUS: Assessing Conceptual Understanding of Statistics

LOCUS : Évaluation de la compréhension conceptuelle de la statistique

The Levels of Conceptual Understanding in Statistics (LOCUS) assessments are an NSF-funded (DRL-1118168) family of assessments designed to measure conceptual -- rather than procedural -- understanding in statistics and are aligned with the American Statistical Association's GAISE Framework (Franklin et al., 2007). This paper introduces the LOCUS assessments with emphasis on their potential for research use. An overview of the development process, including the validity argument for their use with students in grades 6-12, college students, and teachers is provided, along with sample items and findings about students' current knowledge. LOCUS is freely available at <http://locus.statisticseducation.org>.

Les évaluations des niveaux de compréhension conceptuelle de la statistique (LOCUS) sont une famille d'évaluations financées par la NSF (DRL-1118168). Elles sont conçues pour mesurer la compréhension conceptuelle - plutôt que procédurale - de la statistique et coïncident avec le cadre GAISE de l'American Statistical Association (Franklin et al., 2007). Cet article présente les évaluations LOCUS en mettant l'accent sur leur potentiel d'utilisation en recherche. Un aperçu du processus de développement est fourni, y compris l'argument de validité de leur utilisation auprès des élèves des niveaux 6 à 12, des étudiants universitaires et des enseignants, ainsi que des exemples et les résultats sur les connaissances actuelles des élèves. LOCUS est disponible gratuitement au <http://locus.statisticseducation.org>.

1E-I4: Statistical Methods in Molecular Evolution Méthodes statistiques en évolution moléculaire

Organizer and Chair / Responsable et présidente: Liangliang Wang (Simon Fraser University)

Room/Salle: Rowe 1028

Abstracts/Résumés

CÉCILE ANÉ (University of Wisconsin-Madison), **CLAUDIA SOLÍS-LEMUS** (University of Wisconsin - Madison)

[Monday June 15 / lundi 15 juin, 15:30]

Statistical Methods to Detect Reticulate Evolution

Méthodes statistiques pour réseaux phylogénétiques avec hybridations

Phylogenetic networks are often necessary to represent the history of a group of species, where a tree is expanded by hybridization edges to explain gene flow or horizontal gene transfer. However, methods to infer phylogenetic networks are still limited. I will first present a new method to test the goodness-of-fit of the coalescent process on a particular tree (or network), to explain the variation in gene trees obtained from molecular sequences. Second, I will present a statistical method to infer phylogenetic networks from sequence data in a pseudolikelihood framework. Our method has a computational advantage over other methods, because calculations are broken down into smaller problems that are easily parallelizable.

Les réseaux phylogénétiques sont souvent nécessaires pour représenter l'histoire d'un groupe d'espèces, en utilisant un arbre augmenté par des arêtes supplémentaires pour représenter des hybridations ou transferts de gènes horizontaux. Cependant, les méthodes d'inférence de réseaux phylogénétiques sont encore limitées. Je présenterai une nouvelle méthode pour tester l'adéquation du processus de coalescence le long d'un arbre (ou d'un réseau) donné, pour expliquer la variabilité des arbres de gènes. Ensuite, je présenterai une méthode statistique pour l'inférence de réseaux phylogénétiques à partir de séquences moléculaires, en utilisant une pseudovraisemblance. Notre méthode est rapide, parce que les calculs sont séparés en sous-problèmes, eux-même facilement parallélisables.

ALEXANDRE BOUCHARD-CÔTÉ (University of British Columbia)

[Monday June 15 / lundi 15 juin, 16:00]

Inferring the Phylogeny of Cancer Cells

La reconstruction de la phylogénèse de cellules cancéreuses

Cancer cells accumulate large amounts of genetic aberrations as they proliferate, inducing an evolutionary process inside the tumour. Recent advances in sequencing technology allow us to probe intra-tumour heterogeneity, with the objective of elucidating the evolutionary dynamics of cancer.

I will outline the unique challenges of cancer phylogenetics and describe two approaches we have developed in this area. The first approach uses a Bayesian nonparametric model to deconvolve cancer cell subpopulations from bulk sequencing data. The second approach is a tree reconstruction method tailored to cancer evolution.

Les cellules cancéreuses accumulent une profusion d'aberrations génétiques durant leur prolifération, ce qui induit un processus évolutif au sein de la tumeur. Le progrès des technologies de séquençage nous permet de fouiller l'hétérogénéité tumorale afin d'élucider la dynamique évolutive du cancer.

Je vais exposer les grandes lignes des défis uniques de la phylogénèse du cancer et décrire deux approches que nous avons développées dans ce domaine. La première approche utilise un modèle bayésien non paramétrique pour faire la déconvolution de données de séquençage en vrac. La deuxième approche est une méthode de reconstruction d'arbres faite sur mesure pour l'analyse évolutive du cancer.

1E-I4: Statistical Methods in Molecular Evolution Méthodes statistiques en évolution moléculaire

FORREST CRAWFORD (Yale University), **MARC SUCHARD** (University of California, Los Angeles), **VLADIMIR MININ** (University of Washington)

[Monday June 15 / lundi 15 juin, 16:30]

Estimation for General Birth-Death Processes in Molecular Evolution

Estimation des processus généraux de naissance et de mort en évolution moléculaire

Birth-death processes (BDPs) are continuous-time Markov chains that are widely used as models for stochastic counting processes in molecular evolution. However, estimating mutation parameters in these models is remarkably difficult when observation of the process is discrete since likelihoods cannot be computed analytically. We derive EM algorithms for estimating mutation rates in BDPs and show that the E-step can be expressed as a convolution of transition probabilities for any general BDP with arbitrary rates. We use a continued fraction representation of the Laplace transforms of the transition probabilities to derive fast EM algorithms. This allows rapid estimation in a large class of previously intractable stochastic processes. We apply the method to estimate of mutation parameters in microsatellite evolution.

Les processus de naissance et de mort sont des chaînes de Markov en temps continu largement utilisés comme modèles de processus stochastiques de dénombrement en évolution moléculaire. Cependant, il est très difficile d'estimer les paramètres de mutation dans ces modèles quand l'observation du processus est discrète, puisque les vraisemblances ne peuvent être calculées de façon analytique. Nous obtenons des algorithmes EM pour l'estimation des taux de mutation des processus de naissance et de mort, puis nous montrons que l'étape de l'espérance peut être exprimée comme une convolution de probabilités de transition pour tout processus général avec des taux arbitraires. Nous utilisons une représentation en fraction continue des transformées de Laplace des probabilités de transition pour dériver des algorithmes EM rapides. Cette méthode permet une estimation rapide pour une grande classe de processus stochastiques auparavant insolubles. Nous appliquons notre méthode à l'estimation de paramètres de mutation en évolution microsatellite.

1E-15: Stochastic Processes and Applications Processus stochastiques et applications

Organizer and Chair / Responsable et présidente: Gail Ivanoff (University of Ottawa)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Room/Salle: McCain 2170

Abstracts/Résumés

GENNADY SHAIKHET & EVANGELOS KRANAKIS (Carleton University)

[Monday June 15 / lundi 15 juin, 15:30]

Asymptotically Tight Bounds for Sensor Allocation Problems on the Real Line

Bornes asymptotiquement étroites pour les problèmes de répartition de capteurs sur la droite des réels

A large number n of sensors (finite connected intervals) are placed randomly on the real line so that the distances between the consecutive centers are independent random variables with expectation inversely proportional to n . In this work we address two fundamental sensor allocation problems. Interference problem tries to reallocate the sensors from their initial positions so as to eliminate overlaps. Coverage problem, on the other hand, allows overlaps, but tries to eliminate uncovered spaces between the originally placed sensors. Both problems seek to minimize the total sensor movement while reaching their respective goals. Using tools from queueing theory, Skorokhod reflections and weak convergence, we investigate asymptotical behaviour of the optimal costs as n increases to infinity.

Un grand nombre n de capteurs (intervalles connectés finis) sont placés de façon aléatoire sur la droite des réels de sorte que les distances entre les centres consécutifs soient des variables indépendantes avec une espérance inversement proportionnelle à n . Dans cette étude, nous abordons deux problèmes fondamentaux d'allocation de capteurs. Le problème d'interférence tente de réaffecter les capteurs pour éliminer les chevauchements. D'autre part, le problème de couverture permet le chevauchement mais essaie d'éliminer les espaces non couverts entre les emplacements initiaux des capteurs. Ces deux problèmes essaient de minimiser le mouvement total des capteurs tout en parvenant à atteindre leur but respectif. À partir d'outils de la théorie des files d'attente, les réflexions de Skorokhod et la faible convergence, nous explorons le comportement asymptotique des coûts minimaux quand n croît à l'infini.

AARON SMITH (University of Ottawa), **NATESH PILLAI** (Harvard University)

[Monday June 15 / lundi 15 juin, 16:00]

Mixing Time of the Constrained Ising Process on the Lattice at Low Density

Temps de mélange du processus d'Ising sur le treillis à faible densité

The kinetically constrained Ising process (KCIP) was introduced by Fredrickson and Andersen to study the glass transition, and was proposed as a model for computer networks by Aldous. I will present recent work on the mixing properties of the KCIP at low density, which is the natural regime for the network model. We point out that Aldous' conjecture for the mixing time of this process requires

Le processus cinétiqument contraint d'Ising (PCCI) fut introduit par Fredrickson et Anderson pour étudier la transition vitreuse et fut proposé comme modèle pour les réseaux informatiques par Aldous. Je vais présenter une étude récente sur les propriétés de mélange du PCCI à faible densité, ce qui est le régime naturel pour le modèle de réseau. Nous soulignons que la conjecture d'Aldous pour le temps de mélange de ce processus nécessite un simple terme de cor-

1E-15: Stochastic Processes and Applications Processus stochastiques et applications

a simple correction term, and show that this corrected version of the conjecture is true up to logarithmic factors for \mathbb{Z}_n^d , $d \neq 2$. I introduce the model, survey some fantastic recent work on related problems, and give a short proof sketch that emphasizes the relationship between the KCIP and other interacting particle systems.

rection et nous démontrons que cette version corrigée de la conjecture est vraie jusqu'à des facteurs logarithmiques pour \mathbb{Z}_n^d , $d \neq 2$. Je présente le modèle, j'étudie de fantastiques études récentes sur des problèmes connexes, et je donne une brève ébauche de preuve qui souligne la relation entre le PCCI et d'autres systèmes de particules en interaction.

LOUIS-PIERRE ARGUIN (Université de Montréal), **DAVID BELIUS** (NYU), **ADAM HARPER** (Cambridge)
[Monday June 15 / lundi 15 juin, 16:30]

Probabilistic Approach for the Maxima of the Riemann Zeta Function on the Critical Line

Approche probabiliste pour le maximum de la fonction zeta de Riemann sur la droite critique

A recent conjecture of Fyodorov, Hiary and Keating states that the maxima of the Riemann Zeta function on a bounded interval of the critical line behave similarly to the maxima of a specific class of Gaussian fields, the so-called log-correlated Gaussian fields. These include important examples such as branching Brownian motion and the 2D Gaussian free field. In this talk, we will highlight the connections between the number theory problem and the probabilistic models. We will outline the proof of the conjecture in the case of a randomized model of the Zeta function. We will discuss possible approaches to the problem for the function itself. This is joint work with D. Belius (NYU) and A. Harper (Cambridge).

Une conjecture récente de Fyodorov, Hiary et Keating stipule que les maxima de la fonction zeta de Riemann dans un intervalle borné de la droite critique ont des statistiques semblables à une classe spécifique de champs gaussiens dits log-corrélés. Le mouvement brownien branchant et le champs libre gaussien en 2D sont des exemples de tels champs. Dans cet exposé, nous soulignerons les connexions entre le problème sur zeta et les modèles probabilistes. Nous esquisserons la preuve de la conjecture pour un modèle-jouet de la fonction zeta. Ce travail est conjoint avec D. Belius (NYU) et A. Harper (Cambridge).

1E-C1: Complex Correlation Structures and Zero-Heavy Data Structures de corrélation complexes et données à surreprésentation de zéros

Chair/Présidente: Ying Zhang (Acadia University)

Room/Salle: McCain 2198

Abstracts/Résumés

TIAGO PELLEGRINI, TARIQUL HASAN & RENJUN MA (University of New Brunswick)

[Monday June 15 / lundi 15 juin, 15:30]

Tweedie Mixed Models for Spatial Areal Data

Modèles mixtes de Tweedie pour des données géospatiales

The analysis of spatial areal data is important to various application fields to monitor spatial trends, risk and public assessment. However, spatial data has shown difficult due to its complex correlation structure among neighboring areas, which often violates the assumptions of classical statistical models. We propose a new spatial model that takes into account correlated areas and responses through the Tweedie family of distributions. Our model enjoys flexibility while modelling different types of skewed, positive and semi-continuous data when latent processes are also of interest. Data examples will be presented in order to illustrate applications of our modelling approach.

L'analyse de données géospatiales est importante dans divers champs d'application pour surveiller les tendances spatiales et l'évaluation des risques et du public. Cependant, les données spatiales comportent des défis en raison de leur structure de corrélation complexe parmi les zones avoisinantes, ce qui enfreint souvent les hypothèses des modèles statistiques classiques. Nous proposons un nouveau modèle spatial qui tient compte des zones et des réponses corrélées dans la famille des distributions de Tweedie. Notre modèle est flexible et peut traiter différents types de données asymétriques, positives et semi-continues quand les processus latents sont aussi étudiés. Nous présentons des exemples de données pour illustrer des applications de notre modèle.

SHIJIA WANG & NOEL CADIGAN (Memorial University of Newfoundland), **HUGUES BENOIT** (Fisheries and Oceans Canada)

[Monday June 15 / lundi 15 juin, 15:45]

Inference About Regression Parameters Using Highly Stratified Survey Count Data with Overdispersion and Repeated Measurements

Inférence sur les paramètres de régression par l'utilisation de données de dénombrement de sondages hautement stratifiées présentant une surdispersion et des mesures répétées

We study methods to estimate regression and variance parameters for overdispersed and correlated count data from highly stratified surveys. We develop a computationally efficient profile generalized estimating equation method and compare it to marginal maximum likelihood estimation (MLE) and restricted MLE (REML) methods. We use REML to address MLE bias and inaccurate confidence intervals caused by the large number of nuisance parameters involved with highly stratified surveys. The marginal MLE and REML approaches involve

Nous étudions des méthodes permettant d'estimer les paramètres de régression et de variance des données de dénombrement surdispersées et corrélées à partir de sondages hautement stratifiés. Nous développons une méthode profilée d'équations d'estimation généralisées informatiquement efficace et la comparons avec une estimation marginale du maximum de vraisemblance et avec des méthodes d'estimation du maximum de vraisemblance restreint. Nous utilisons une estimation du maximum de vraisemblance restreint pour éviter le biais de l'estimation du maximum de vraisemblance et pour éviter des intervalles de confiance imprécis causés

1E-C1: Complex Correlation Structures and Zero-Heavy Data Structures de corrélation complexes et données à surreprésentation de zéros

intractable integrals and we used a new R package that is designed for estimating complex nonlinear models that may include random effects.

par un grand nombre de paramètres de nuisance qui découlent des sondages hautement stratifiés. Les approches de l'estimation marginale du maximum de vraisemblance et de l'estimation du maximum de vraisemblance restreint font appel à des intégrales insolubles. Nous utilisons un nouveau paquet R conçu pour estimer les modèles non-linéaires complexes qui peuvent inclure des effets aléatoires.

TARIQUL HASAN (University of New Brunswick), **GARY SNEDDON** (Mount St. Vincent University), **RENJUN MA** (University of New Brunswick)

[Monday June 15 / lundi 15 juin, 16:00]

Simultaneous Modelling of Clustered Marginal Counts and Multinomial Proportions with Zero-Inflation with Application to Analysis of Osteoporotic Fractures Data

Modélisation simultanée de données de dénombrement marginal en grappes et de proportions multinomiales avec surreprésentation de zéros appliquée à l'analyse de données de fractures ostéoporotiques

Osteoporotic fractures are known to be highly recurring. We investigate bone-dependent and bone-independent risk factors of osteoporotic fracture frequency and relative proportions at various body locations using the data from the osteoporotic fracture study conducted by the National Health and Nutrition Examination Survey (NHANES), 2007-2008. We propose a new zero-inflated baseline-category multinomial mixed model to characterize the clustered count responses and multinomial proportions by subject simultaneously while taking account of zero-inflation and randomness of cluster sizes. Our approach gives additional insights into the risk factors of osteoporotic fracture frequency occurring at various body locations.

Les fractures ostéoporotiques sont reconnues pour leur récurrence élevée. Nous étudions les facteurs de risque qui dépendent ou non des os pour la fréquence et la proportion relative des fractures ostéoporotiques à divers endroits du corps à l'aide des données issues de l'étude portant sur ce type de fracture menée par le National Health and Nutrition Examination Survey (NHANES) en 2007-2008. Nous proposons un nouveau modèle mixte multinomial à surreprésentation de zéros avec catégorie de référence pour caractériser simultanément les variables de dénombrement par grappes et les proportions multinomiales par sujet, tout en tenant compte de la surreprésentation de zéros et de la nature aléatoire de la taille des grappes. Notre approche offre un nouvel éclairage sur les facteurs de risque relatifs à la fréquence des fractures ostéoporotiques qui surviennent à divers endroits du corps.

CHUNLIN WANG, PENGFEI LI & PAUL MARRIOTT (University of Waterloo)

[Monday June 15 / lundi 15 juin, 16:15]

Semiparametric Inferences on the Means of Groups with Excess Zero Observations

Les inférences semi-paramétriques sur les moyennes de groupes avec des observations à surreprésentation de zéros

In this talk, we propose empirical likelihood-based approaches for making statistical inferences on the means of positively skewed populations with excess zero observations. We propose inference procedures about the difference and ratio of means based on the empirical likelihood ratio (ELR) and Wald-type statistics under the semiparametric density ratio model. The proposed ELR method pro-

Dans cet exposé, nous proposons des approches empiriques fondées sur la vraisemblance pour effectuer des inférences statistiques sur les moyennes de populations présentant une asymétrie positive avec des observations à surreprésentation de zéros. Nous proposons des procédures d'inférence concernant la différence et le rapport des moyennes basées sur le rapport de vraisemblance empirique et les statistiques de type Wald dans un modèle de rapport de den-

1E-C1: Complex Correlation Structures and Zero-Heavy Data Structures de corrélation complexes et données à surreprésentation de zéros

vides a unified framework that covers a wide range of inference problems through the use of general estimating equations. Large sample properties of both methods are studied and simulation results demonstrate that the proposed methods are well adapted to the data with skewness and excess zeros.

sités semi-paramétrique. La méthode du rapport de vraisemblance fournit un cadre de travail unifié qui couvre de nombreux problèmes d'inférence par l'utilisation d'équations d'estimation générales. Nous examinons les propriétés des deux méthodes dans un cadre de grands échantillons et démontrons que les méthodes proposées sont bien adaptées aux données à surreprésentation de zéros présentant une asymétrie positive.

ERIN LUNDY & CHARMAINE DEAN (Western University), **ELIZABETH JUAREZ-COLUNGA** (University of Colorado Denver)

[Monday June 15 / lundi 15 juin, 16:30]

Joint Models for Multivariate Longitudinal Zero-Heavy Count Data Analysis

Modèles conjoints pour l'analyse de données de dénombrement multivariées longitudinales à surreprésentation de zéros

Regression models for zero-inflated count data often need to accommodate within-subject correlation and between-individual heterogeneity; random effects models may be utilized for incorporating complex correlation structures. We consider several longitudinal zero-heavy count outcomes jointly. A novel feature is the use of a multistage approach in situations where zero counts arise from distinct sources. In the context of a major study on criminal behaviour patterns, we address an additional complication in that some outcomes are prohibited during specific exposure times leading to some of the zero-heavy nature of the data accounted for in a structural manner based on an exposure variable.

Les modèles de régression pour les données de dénombrement à surreprésentation de zéros doivent souvent tenir compte de la corrélation à l'intérieur un même sujet et de l'hétérogénéité entre individus. Des modèles à effets aléatoires peuvent être utilisés pour incorporer des structures de corrélation complexes. Nous considérons conjointement plusieurs réponses de dénombrement longitudinales à surreprésentation de zéros. Une nouvelle caractéristique consiste à utiliser une approche en plusieurs étapes dans des situations dans lesquelles les dénombrements de zéro surviennent à partir de sources distinctes. Dans le contexte d'une étude importante sur les types de comportements criminels, nous traitons une complication supplémentaire dans laquelle certaines réponses sont impossibles pendant des durées spécifiques d'exposition, ce qui amène prendre en compte des données de nature à surreprésentation de zéros d'une manière structurelle en fonction d'une variable d'exposition.

HAOCHENG LI (University of Calgary), **RAYMOND CARROLL** (Texas A&M University)

[Monday June 15 / lundi 15 juin, 16:45]

Joint Modeling Methods for Functional Data Mixed with Zero-and-One Inflated Continuous Proportions and Skewed Continuous Measurements

Méthodes de modélisation conjointe pour des données fonctionnelles mélangées avec des proportions continues à surreprésentation de zéros et de uns et des mesures continues asymétriques

A functional data approach is proposed for outcomes mixed with a proportion and continuous measurements. The continuous proportions are featured by excess values at zero and one, and the continuous measures are skewed. A three-

Nous proposons une approche par analyse fonctionnelle pour des réponses mélangées avec une proportion et des mesures continues. Les proportions continues sont caractérisées par un surnombre de zéros et de uns, et les mesures continues sont asymétriques. Nous présentons un modèle

1E-C1: Complex Correlation Structures and Zero-Heavy Data

Structures de corrélation complexes et données à surreprésentation de zéros

part functional data modeling is introduced. In this three-part model, the regression structures are specified as smooth curves measured at various time-points with random effects that have a correlation structure. The difficulties in handling the binary and continuous proportions variables are solved by using a quasiliikelihood type approximation. We develop an efficient algorithm, which involves the selection of the number of principal components.

d'analyse fonctionnelle en trois parties dont les structures de régression sont définies comme des courbes lisses mesurées à divers temps et présentant des effets aléatoires qui ont une structure de corrélation. Nous réglons les difficultés liées au traitement des variables des proportions binaires et continues à l'aide d'une approximation de type quasi vraisemblance. Nous développons un algorithme efficace qui implique la sélection du nombre de composantes principales.

1E-C2: Experimental Design, Simulation, and Reliability Plan expérimental, simulation et fiabilité

Chair/Président: Stefan Steiner (University of Waterloo)

Room/Salle: McCain 2116

Abstracts/Résumés

HUGH CHIPMAN & PRITAM RANJAN (Acadia University)

[Monday June 15 / lundi 15 juin, 15:30]

Simulation Studies for Statistical Procedures: Why Can't We Practice What We Preach?

Études de simulation pour des procédures statistiques : pourquoi ne pouvons-nous pas faire ce que nous prêchons?

When a new statistic, model, or learning algorithm is developed, simulation studies are often used to examine sampling behaviour or other performance measures. These studies are experimental designs. Yet the most common designs are, to put it politely, unsophisticated: either varying a single factor at a time, or a time-consuming full-factorial study. On the analysis side, basic tools like ANOVA are often forgotten. Instead, voluminous tables are presented! We demonstrate simple concepts from the design and analysis of experiments. Such tools can do for statistical simulations what statisticians promise to other scientists: save time and make efficient use of data.

Lorsqu'un nouvel algorithme statistique, de modèle, ou d'apprentissage est développé, des études de simulation sont souvent utilisées pour analyser un comportement d'échantillonnage ou d'autres mesures de rendement. Ces études sont des plans expérimentaux. Désormais, les plans les plus communs ne sont pas sophistiqués : soit c'est une étude qui fait varier un seul facteur à la fois, soit c'est une étude factorielle complète qui prend du temps. Du point de vue analytique, on oublie souvent des outils de base comme ANOVA. À la place, on présente de nombreux tableaux volumineux ! Nous démontrons des concepts simples provenant de plans et d'analyses d'expériences. Ces outils permettent ce que promettent les statisticiens à d'autres scientifiques pour des simulations statistiques : gagner du temps et utiliser efficacement les données.

FATIMAH AL AHMAD & HUGH CHIPMAN (Acadia University)

[Monday June 15 / lundi 15 juin, 15:45]

Assessing the Accuracy of BART Credible Intervals for Prediction

Évaluation de la précision des intervalles de crédibilité à des fins prédictives

Bayesian Additive Regression Trees (BART) is an accurate and flexible modern supervised statistical learning model. BART provides Bayesian credible intervals for predictions. This talk shall illustrate the coverage and accuracy of BART credible intervals. Fractional factorial design will be a main part of conducting this experiment to study the influence of factors such as sample size and predictor correlation on BART credible intervals.

Les arbres additifs de régression bayésienne (BART) forment un modèle d'apprentissage statistique moderne précis et flexible. BART donne des intervalles de crédibilité bayésiens à des fins prédictives. Cet exposé illustrera la couverture et la précision des intervalles de crédibilité BART. Le plan factoriel fractionnel sera une composante importante de cette expérimentation pour étudier l'influence de certains facteurs tels que la taille de l'échantillon et la corrélation entre prédicteurs sur les intervalles de crédibilité BART.

NATHANIEL STEVENS, STEFAN STEINER & JOCK MACKAY (University of Waterloo)

[Monday June 15 / lundi 15 juin, 16:00]

Quantifying Agreement Between Two Methods of Measurement

Quantification de l'adéquation entre deux méthodes de mesure

1E-C2: Experimental Design, Simulation, and Reliability Plan expérimental, simulation et fiabilité

The comparison of two measurement systems is important in a variety of contexts. A common goal of such a comparison is to decide if a new measurement system agrees suitably with an existing one, and hence whether the two can be used interchangeably. Various methods for assessing agreement exist, but none are without challenges. In this talk we introduce the probability of agreement, a novel method that more effectively and transparently quantifies agreement between two measurement systems, than do existing analysis methods. We also make recommendations for a study design that facilitates precise estimation of the probability of agreement.

La comparaison de deux systèmes de mesure est importante dans divers contextes. L'objectif commun d'une telle comparaison consiste à déterminer si un nouveau système de mesure concorde correctement avec un système existant et, par conséquent, si les deux systèmes peuvent être utilisés de façon interchangeable. Diverses méthodes existent pour l'évaluation de l'adéquation, mais aucune n'est entièrement dépourvue de défis. Dans cet exposé, nous présentons la probabilité d'adéquation, une nouvelle méthode qui quantifie l'adéquation entre deux systèmes de mesure de façon plus efficace et transparente que les méthodes d'analyse existantes. Nous formulons des recommandations pour un modèle de planification d'étude qui permet d'améliorer la précision de l'estimation de la probabilité d'adéquation.

CANDEMIR CIGSAR (Memorial University of Newfoundland)

[Monday June 15 / lundi 15 juin, 16:15]

Carryover Effects in Repairable Systems

Effets de report dans les systèmes réparables

Reliability data analysis of repairable systems typically involves recurrent events. An important aspect of recurrent event processes that may result in some forms of clustering of events is the presence of carryover effects. This phenomenon can occur when repairs to address a failure in a hardware or software system may not resolve the problem or may even introduce new problems. Therefore, carryover effects provide a natural definition for imperfect repairs in repairable systems. In this study, we discuss testing such effects in repairable system settings. We provide simple but intuitive tests for carryover effects.

L'analyse des données de fiabilité des systèmes réparables implique généralement des événements récurrents. L'un des aspects importants des processus d'événements récurrents qui donnent lieu à un certain regroupement d'événements est la présence d'effets de report. Ce phénomène peut se produire lorsque les réparations visant à remédier à la défaillance d'un système matériel ou logiciel ne résolvent pas le problème (ou en introduisent de nouveaux). Ainsi, les effets de report offrent une définition naturelle de la réparation imparfaite d'un système réparable. Dans cette étude, nous discutons des tests qui permettent d'identifier ces effets dans les systèmes réparables. Nous proposons des tests simples mais intuitifs pour détecter les effets de report.

HENSLEY HUBERT MARIATHAS & CANDEMIR CIGSAR (Memorial University of Newfoundland)

[Monday June 15 / lundi 15 juin, 16:30]

The Effect of Heterogeneity on the Assessment of Imperfect Repairs

Effet de l'hétérogénéité sur l'évaluation des réparations imparfaites

Repairable systems are subject to imperfect repairs. In such settings, the event intensity function may temporarily change after an event or condition occurs when repairs undertaken to deal with a failure may not resolve the problem. Although models that can reflect such behaviour are attractive, unobserved heterogeneity could be an important obstacle for their use when multiple processes are under

Les systèmes réparables font parfois l'objet de réparations imparfaites. Dans de telles situations, la fonction d'intensité d'événement risque de changer temporairement après un événement ou une condition si les réparations effectuées pour réparer une défaillance du système n'ont pas résolu le problème. Il existe des modèles intéressants capables de refléter ce comportement, mais l'hétérogénéité non détectée peut constituer un obstacle majeur à leur utilisation lorsque

1E-C2: Experimental Design, Simulation, and Reliability Plan expérimental, simulation et fiabilité

observation. The main objective of this talk is to investigate the effect of heterogeneity in the analysis of repairable systems when repairs are imperfect. We discuss this issue through simulations and analytically.

de multiples processus sont sous observation. L'objectif principal de cette présentation est d'étudier l'effet de l'hétérogénéité sur l'analyse des systèmes réparables lorsque les réparations sont imparfaites. Nous discutons de ce problème via des simulations et par analyse.

PO YANG (University of Manitoba), **CHANG-YUN LIN** (National Chung Hsing University)

[Monday June 15 / lundi 15 juin, 16:45]

Response Surface Methodology using Split-Plot Definitive Screening Designs

Méthodologie de surface de réponse avec plans de criblage définitifs à parcelles divisées

Definitive screening designs are a new class of three-level designs. We investigate the performance of such designs in split-plot structures. The result of the projection eligibility and the study of D -efficiency and I -efficiency show that split-plot definitive screening designs perform well when the number of important factors is small. To reduce the risk of being unable to fit second-order models for response surfaces, we provide the column indexes of projections. Experimenters can assign potentially important factors to those columns to avoid ineligible projections. An example is presented to demonstrate how to analyze data using the split-plot definitive screening design.

Les plans de criblage définitifs sont une nouvelle classe de plans à trois niveaux. Nous étudions la performance de ces plans dans les structures à parcelles divisées. Le résultat de l'éligibilité des projections et l'étude de l'efficacité D et I montrent que les plans de criblage définitifs à parcelles divisées donnent de bons résultats lorsque le nombre de facteurs importants est limité. Afin de réduire le risque de ne pas pouvoir ajuster des modèles de second ordre aux surfaces de réponse, nous donnons les indices de colonne des projections. Les expérimentateurs peuvent alors attribuer les facteurs potentiellement importants à ces colonnes pour éviter des projections inéligibles. Nous présentons un exemple qui montre comment analyser les données à l'aide du plan de criblage définitif à parcelles divisées.

1E-C3: Statistical Science for Ocean Ecosystems Science statistique pour les écosystèmes océaniques

Chair/Président: Duncan Murdoch (Western University)

Room/Salle: McCain 2184

Abstracts/Résumés

JOANNA MILLS FLEMMING (Dalhousie University)

[Monday June 15 / lundi 15 juin, 15:30]

Statistical Methods for Assessing Impacts of Overfishing

Méthodes statistiques pour l'évaluation des répercussions de la surpêche

The oceanic ecosystem is by far the largest on Earth, covering more than 70% of the planet. Human impacts on this ecosystem like overfishing are of serious conservation concern. In fact, overfishing is the primary threat to elasmobranch species (sharks, skates and rays) around the world. Abundance is a key quantity for conservation and management strategies but remains challenging to reliably estimate particularly from the limited data available for many such endangered marine species. Here we present statistical tools for properly estimating both probability of presence and abundance as well for identifying 'hotspots' of occurrence through space and time.

L'écosystème océanique est de loin le plus grand sur terre, couvrant plus de 70 % de la planète. L'incidence de l'activité humaine sur l'écosystème, comme la surpêche, pose de sérieux problèmes de conservation. En fait, la surpêche est la menace principale aux espèces des élastombranchés (requins, pocheteaux et raies) dans le monde entier. L'abondance est une quantité essentielle pour les stratégies de gestion et de conservation, mais continue à constituer un défi pour effectuer une estimation fiable, notamment à partir des données disponibles limitées pour un grand nombre de ces espèces marines en voie de disparition. Dans cet exposé, nous présentons des outils statistiques pour estimer correctement la probabilité de la présence et de l'abondance, ainsi que pour identifier les « points chauds » d'occurrences dans l'espace et le temps.

STUART CARSON & JOANNA MILLS FLEMMING (Dalhousie University)

[Monday June 15 / lundi 15 juin, 15:45]

*Spatio-Temporal Statistical Modelling for Atlantic Cod (*Gadus morhua*) Abundance on the Scotian Shelf using Sample Survey Data*

*Modélisation statistique spatio-temporelle de l'abondance de la morue de l'Atlantique (*Gadus morhua*) sur le plateau néo-écossais à l'aide de données d'enquête*

We present a spatio-temporal model for Atlantic cod on the Scotian shelf based upon sample survey data. The discussion highlights the spatial and temporal aspects in the period of decline experienced. We demonstrate how the density in areas of preferred habitat remained steady despite changes in the stock level but that the spatial extent of distribution changed. We discuss how the use of a spatio-temporal model allows these aspects of the species behaviour to be easily visualized and graphically and intuitively presented. We discuss further work;

Nous présentons un modèle spatio-temporel pour la morue de l'Atlantique sur le plateau néo-écossais à l'aide de données d'enquête. La discussion met en lumière les aspects spatiaux et temporels dans la période de déclin connu. Nous démontrons comment la densité dans les zones d'habitat préféré est restée stable malgré les changements dans le niveau des stocks, mais que l'étendue spatiale de la distribution a changé. Nous discutons comment l'utilisation d'un modèle spatio-temporel permet de visualiser facilement et graphiquement ces aspects du comportement des espèces et des les présenter intuitivement. Nous discutons de la

1E-C3: Statistical Science for Ocean Ecosystems Science statistique pour les écosystèmes océaniques

a spatio-temporal analysis of a predator and a prey species is jointly modelled.

suite des travaux ; une modélisation conjointe d'une analyse spatio-temporelle d'un prédateur et d'une proie.

NOEL CADIGAN & JON FISHER (Marine Institute of Memorial University of Newfoundland), **DARRELL MULLOWNEY & EARL DAWE** (Fisheries and Oceans Canada)

[Monday June 15 / lundi 15 juin, 16:00]

A Spatio-Temporal Stock Assessment Model for Snow Crab in Newfoundland and Labrador

Un modèle spatio-temporel d'évaluation des stocks de crabe des neiges à Terre-Neuve-et-Labrador

Fish stock size declines throughout the fishing season as fishery catches increase and this relationship can be sometimes used to estimate stock size based on the rate of decline in a catch per unit effort (CPUE) time-series. However, this decline can also be affected by changes in the spatial distribution of the fishing fleet. We propose a spatio-temporal depletion model that adjusts for bias caused by changes in the fleet distribution. This model is applied to data from 2000-2012 to improve parameter estimation for the economically important snow crab fishery in Newfoundland and Labrador.

La taille des stocks de poissons diminue tout au long de la saison de la pêche au fur et à mesure que les prises de pêche augmentent. Cette relation peut parfois être utilisée pour estimer la taille des stocks selon le taux de déclin à l'aide d'une série chronologique de prise par unité d'effort (PPUE). Toutefois, cette baisse peut également être affectée par des changements dans la distribution spatiale de la flotte de pêche. Nous proposons un modèle de diminution spatio-temporel qui ajuste le biais causé par les changements dans la répartition de la flotte. Nous appliquons ce modèle à des données de 2000 à 2012 dans le but d'améliorer l'estimation des paramètres pour la pêche de crabe des neiges à Terre-Neuve-et-Labrador qui est importante du point de vue économique.

LIN QIN & MARTIN LYSY (University of Waterloo)

[Monday June 15 / lundi 15 juin, 16:15]

Detecting Sea Water Migration in Canadian Arctic Fish

Détection de la migration des poissons d'eau de mer en Arctique canadien

Recent studies revealed increasing rates of mercury exposure in Canadian Inuit populations. This is due to high contamination of freshwater fish which constitutes a primary component of the traditional northern diet. However, sea-run fish have much lower mercury than their freshwater counterparts. Thus, we need to determine the migratory patterns of various fish species to suggest safer food source. In order to do this, high frequency multivariate data is collected on the fish ear bones that record the chemical composition of the sample throughout lifetime. We propose the multiple change-point search methods to understand measurement error, seasonal trends, individual-level effects.

Des études récentes révèlent des taux d'exposition au mercure en augmentation dans les populations inuites du Canada. Ceci s'explique par une contamination élevée du poisson d'eau douce qui constitue le composant principal du régime alimentaire nordique. Cependant, le poisson de mer contient beaucoup moins de mercure que les poissons d'eau douce. Ainsi, nous devons déterminer les courants migratoires des diverses espèces de poissons pour suggérer une source alimentaire plus sécuritaire. À cette fin, des données multivariées à haute fréquence sont recueillies sur les osselets d'oreilles de poissons qui enregistrent la composition chimique de l'échantillon dans un cycle de vie. Nous proposons plusieurs méthodes de recherche avec points de rupture pour comprendre les erreurs de mesures, les tendances saisonnières et les effets individuels.

1E-C3: Statistical Science for Ocean Ecosystems Science statistique pour les écosystèmes océaniques

DANIEL DINSDALE & MATÍAS SALIBIÁN-BARRERA (University of British Columbia)

[Monday June 15 / lundi 15 juin, 16:30]

Stochastic Modelling of Marine Mammal Tag Data

Modélisation stochastique des données d'étiquetage des mammifères marins

Current marine mammal tags provide researchers with a wealth of data about animal behaviour, location and their environment. The information collected by these tags can, for example, be used to describe habitat and population dynamics. In this work we use a flexible stochastic state space model to represent the movement of Northern fur seals in the Eastern Bering Sea. Our model incorporates uncertainty in the GPS locations. If time permits we will also discuss the problem of preferential sampling that may be present in environment data (e.g. water temperature) collected by animal-mounted sensors.

Les étiquettes actuelles des mammifères marins fournissent une multitude de données aux chercheurs sur le comportement, la localisation et l'environnement de l'animal. L'information recueillie par ces étiquettes peut, par exemple, être utilisée pour décrire l'habitat et les dynamiques de la population. Dans cette étude, nous utilisons un modèle spatial d'état stochastique flexible pour représenter le mouvement des otaries à fourrure du Nord dans la partie est de la mer de Béring. Notre modèle incorpore de l'incertitude dans les localisations GPS. Si le temps le permet, nous discuterons aussi du problème d'échantillonnage préférentiel qui peut se présenter dans des données environnementales (par exemple, la température de l'eau) recueillies par des capteurs fixés sur un animal.

CONNIE STEWART (University of New Brunswick Saint John)

[Monday June 15 / lundi 15 juin, 16:45]

Measuring Distance between Compositional Data Containing Zeros

Mesure de la distance entre des données compositionnelles comportant des zéros

Distance measures play an important role in a variety of applications involving multivariate data. When the data is compositional with zeros, traditional distance measures either cannot be applied directly or do not satisfy fundamental properties of compositional data analysis. In this talk we compare three distance measures that are capable of handling zeros, but do not satisfy the principle of subcompositional coherence. We present the results of a simulation study in which we attempted to measure the subcompositional incoherence of the distance measures, as well as an application of the measures to real-life grey seal data which motivated this work.

Les mesures de distance jouent un rôle important dans une foule d'applications comportant des données multivariées. En présence de données compositionnelles avec des zéros, les mesures de distance classiques ne peuvent pas être appliquées directement ou ne respectent pas les propriétés fondamentales de l'analyse de données compositionnelles. Dans cet exposé, nous comparons trois mesures de distance qui peuvent traiter des zéros, mais qui ne respectent pas le principe de cohérence sous-compositionnelle. Nous présentons les résultats d'une étude de simulation dans laquelle nous tentons de mesurer l'incohérence sous-compositionnelle des mesures de distance, ainsi qu'une application des mesures à des données réelles sur le phoque gris, qui ont motivé ces travaux.

1E-C4: Statistics in Action: A Smörgåsbord Statistique en action : un «pot-pourri »de présentations

Chair/Présidente: Wendy Lou (University of Toronto)

Room/Salle: McCain 2102

Abstracts/Résumés

YIZHOU FANG & MARTIN LYSY (University of Waterloo)

[Monday June 15 / lundi 15 juin, 15:30]

Fast and Flexible Non-linear Modeling of Financial Assets

Modélisation non linéaire rapide et souple des actifs financiers

The last few decades have featured numerous examples of delicately inter-related financial products. Stochastic models have gone very far in the financial modeling field. However, many of the complex multi-factor models used at present incur a large computational burden which increases dramatically with the size of the data. We propose a nonlinear framework including observable proxies for latent volatility components. Inference for the proposed framework is very fast and scalable to high frequency data, and therefore amenable to large-scale model comparisons and goodness-of-fit assessments. A number of case studies illustrating the methodology will be presented.

Au cours des dernières décennies, de nombreux produits financiers subtilement liés ont vu le jour. Les modèles stochastiques ont fait de gros progrès dans le domaine de la modélisation financière. Cependant, bon nombre des modèles multifactoriels complexes utilisés actuellement exigent un lourd traitement computationnel, qui ne fait que croître avec le volume de données. Nous proposons un cadre non linéaire qui inclut des substituts observables pour les éléments de volatilité latents. Dans le cadre proposé, l'inférence est très rapide et peut s'étendre à des données de haute fréquence, permettant ainsi des comparaisons de modèles de plus grande échelle et une évaluation de l'ajustement des modèles. Nous présentons plusieurs études de cas qui illustrent notre méthodologie.

PHILIPPA SWARTZ, DEREK BINGHAM, MICHAEL GROSSKOPF & TIM SWARTZ (Simon Fraser University)

[Monday June 15 / lundi 15 juin, 15:45]

The Quality of Pitches in Major League Baseball

La qualité des lanceurs dans la Ligue majeure de baseball

We investigate the quality of pitches thrown in Major League Baseball (MLB). The approach is based on characterizing pitches based on pitch count and a descriptor. The descriptor may include information on the type of pitch, the speed of the pitch, the location of the pitch, etc. This is a big data problem with roughly 750,000 pitches thrown in a single season of MLB. If we are able to assess the quality of pitches, then a host of practical questions may be addressed. For example, can we detect when a pitcher's performance begins to deteriorate prior to bad results?

Nous étudions la qualité des lanceurs de balles dans la Ligue majeure de baseball. Notre approche est basée sur la caractérisation des lanceurs en fonction du nombre de lanceurs et d'un descripteur. Celui-ci peut contenir de l'information sur le type de lancer, la vitesse et le lieu du lancer, etc. Le nombre de données constitue un problème, car il y a environ 750 000 lanceurs en une seule saison dans la Ligue majeure de baseball. Si nous sommes en mesure d'analyser la qualité des lanceurs, une foule de questions pratiques peut être abordée. Par exemple, pouvons-nous détecter quand l'efficacité du lanceur commence à se détériorer avant que les résultats ne deviennent mauvais ?

1E-C4: Statistics in Action: A Smörgåsbord Statistique en action : un «pot-pourri »de présentations

MICHAEL GUERZHOY & NATHAN TABACK (University of Toronto)

[Monday June 15 / lundi 15 juin, 16:00]

Hierarchical Bayesian Models for Uncertainty-Quantified Ranking of Restaurant Chains by Food Safety Compliance
Modèles bayésiens hiérarchiques pour classement avec incertitude quantifiée des chaînes de restaurants selon la conformité aux normes sanitaires alimentaires

For an episode of CBC Marketplace (April 2014), data has been collected on the number of food safety violations found during inspections at hundreds of locations of 13 restaurant chains in 5 Canadian cities. We describe several methods to produce an uncertainty-quantified nationwide ranking of the chains by food safety compliance, addressing the issue of differing standards for what constitutes a violation in different cities. We demonstrate a ranking based on the coefficients of a Poisson regression. To obtain more precise rankings with less uncertainty, and gain insights about the data, we fit a series of Bayesian overdispersed Poisson models.

Pour un épisode de CBC Marketplace (2014), des données furent recueillies sur le nombre de violations de normes sanitaires trouvées lors des inspections, à de centaines d'emplacements de 13 chaînes de restaurants, dans cinq villes canadiennes. Nos méthodes produisent un classement des chaînes au Canada, avec incertitude quantifiée, selon leur conformité aux normes sanitaires, abordant la question des différentes définitions de « violation » dans différentes villes. Nous démontrons un classement basé sur les coefficients d'une régression de Poisson. Pour obtenir un classement plus précis et pour examiner les données, nous adaptons une série de modèles de Poisson surdispersés bayésiens.

ANITA BROBBEY & ASOKAN MULAYATH VARIYATH (Memorial University of Newfoundland)

[Monday June 15 / lundi 15 juin, 16:15]

Analysis of Ordered Categorical Data from Designed Experiments
Analyse de données catégorielles ordonnées de plans d'expériences

The responses of interest in many designed experiments are ordered categorical. Taguchi Accumulation analysis (AA), Nair's Scoring Scheme (1986) and Jeng's weighted Probability Scoring Scheme are existing approaches for analyzing ordered categorical data from industrial experiments for optimal factor settings in quality improvement studies. In this work, we discuss analyses of ordered categorical data using cumulative link models. This approach models both the location and dispersion effects and it is easy to implement. Relative Performances of the four approaches are compared through simulation studies and analysis of real data set.

La réponse d'intérêt dans de nombreux plans d'expériences est catégorielle ordonnée. L'analyse d'accumulation (AA) de Taguchi, le schéma de notation de Nair (1986) et le schéma de notation de probabilité pondérée de Jeng sont des approches existantes d'analyse des données catégorielles ordonnées provenant d'expériences industrielles visant à établir les réglages optimaux des facteurs dans les études d'amélioration de la qualité. Dans ce travail, nous discutons des analyses de données catégorielles ordonnées qui utilisent des modèles cumulatifs liés. Cette approche modélise à la fois les effets de l'emplacement et de la dispersion et elle est facile à mettre en œuvre. Les performances relatives des quatre approches sont comparées à l'aide d'études de simulation et d'analyses sur un jeu de données réel.

1E-C4: Statistics in Action: A Smörgåsbord Statistique en action : un «pot-pourri »de présentations

THUVA VANNIYASINGAM, GARY FOSTER, LEHANA THABANE & ALFONSO IORIO (McMaster University)

[Monday June 15 / lundi 15 juin, 16:30]

Determining Individual-Specific Terminal Half-Life Estimates and Optimal Sampling Time Points for Patients with Hemophilia B Administered rFIXfc

Détermination des estimations individuelles spécifiques de demi-vie terminale et échantillonnage optimal des points dans le temps pour des patients atteints d'hémophilie B à qui on a administré rFIXfc

Individual-specific estimates have not been determined for patients with hemophilia B who are receiving prophylactic treatment for Alprolix. We aim to (i) determine individual-specific terminal half-life estimates and (ii) identify informative sampling points to reduce the required number of blood samples, without loss of accuracy in pharmacokinetic estimates. Using nonlinear mixed-effects regression, we identified a three-compartment model to describe plasma concentrations after an IV bolus injection, while simultaneously estimating interindividual variability, intraindividual variability, and covariate effects. A limited sampling strategy will be used to determine optimal sampling time points to reduce the number of required concentration measures and improve overall prophylactic management.

Des estimations individuelles spécifiques n'ont pas été déterminées pour les patients atteints d'hémophilie B qui reçoivent un traitement prophylactique pour Alprolix. Nous visons à (i) déterminer des estimations individuelles spécifiques de demi-vie terminale et (ii) identifier des points d'échantillonnage informatifs pour réduire le nombre d'échantillons de sang requis, sans perdre de la précision dans les estimations pharmacocinétiques. En utilisant une régression non linéaire à effets mixtes, nous identifions un modèle à trois compartiments pour décrire les concentrations de plasma suite à une injection du bolus IV, tout en estimant simultanément la variabilité d'un individu à l'autre, la variabilité à l'intérieur d'un même individu, et les effets covariables. Une stratégie d'échantillonnage limitée sera utilisée pour déterminer l'échantillonnage optimal des points dans le temps pour réduire le nombre de mesures de concentration requis et pour améliorer la gestion globale du prophylactique.

CHARLES BORDET & LOUIS-PAUL RIVEST (Université Laval)

[Monday June 15 / lundi 15 juin, 16:45]

A Stochastic Pella Tomlinson Model and its Maximum Sustainable Yield

Le modèle stochastique de Pella Tomlinson et son rendement équilibré maximal

This presentation investigates the biological reference points, such as the maximum sustainable yield, for the Pella Tomlinson surplus production model in the presence of a multiplicative environmental noise. This model is used in fisheries stock assessment as a firsthand tool for the elaboration of fishing strategies. The consideration of random noise leads to more conservative harvesting target than deterministic models.

Cette présentation s'intéresse aux points de référence biologique, tels que le rendement équilibré maximal, pour le modèle de surplus de production de Pella et Tomlinson alloué d'un bruit environnemental multiplicatif. Ce modèle est utilisé dans l'évaluation des stocks pour les pêcheries comme premier outil lors de l'élaboration des stratégies de pêche. La prise en compte d'un bruit aléatoire implique un taux de pêche optimal plus conservateur que dans le cas déterministe.

2A-A: Gold Medal Address
Allocution du récipiendaire de la Médaille d'or

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

Room/Salle: McCain - Ondaatje

Abstract/Résumé

JIAHUA CHEN (University of British Columbia)

[Tuesday June 16 / mardi 16 juin, 8:45]

EM-Test for Finite Mixture Models

Test EM pour les modèles de mélanges finis

In scientific investigations, a population is often suspected of containing several more homogeneous sub-populations. Such a population structure is most accurately described by a finite mixture model. The evidence for mixture is best examined through a rigorous statistical hypothesis test. Developing valid and effective statistical inference methods on is an important and challenging research problem. Classical procedures when applied to mixture models often have sophisticated asymptotic properties which render them useless in applications. For many finite mixture models, we have successfully designed corresponding EM-tests whose limiting distributions are easier to derive mathematically and simpler for implementation in data analysis. This talk illustrates the ideas behind the EM-tests, their elegant asymptotic properties and other related issues.

Dans les études scientifiques, une population est souvent soupçonnée de contenir plusieurs sous-populations homogènes. Une telle structure de la population est décrite le plus précisément par un modèle de mélange fini. La preuve supportant le mélange peut être examinée à travers un test d'hypothèse statistique. Le développement de méthodes d'inférence statistique valides et efficaces est un problème de recherche important et complexe. Les procédures classiques ont souvent des propriétés asymptotiques sophistiquées quand elles sont appliquées à des modèles de mélanges ce qui les rend inutile en pratique. Pour plusieurs modèles de mélanges finis, nous avons conçu avec succès des tests EM correspondants dont les lois limites sont plus faciles à dériver mathématiquement et plus facile à implanter dans l'analyse de données. Cet exposé illustre les idées derrière les tests EM, leurs élégantes propriétés asymptotiques et autres questions connexes.

2B-11: Modelling of Temporally Correlated Multivariate Counts Modélisation de décomptes multivariés corrélés temporellement

Organizer and Chair / Responsable et président: Tariqul Hasan (University of New Brunswick)

Room/Salle: Mona Campbell 1108

Abstracts/Résumés

DIMITRIS KARLIS (Athens University of Economics and Business)

[Tuesday June 16 / mardi 16 juin, 10:20]

On Multivariate Time Series for Counts Based on Multivariate Integer Autoregressive models

Séries chronologiques multivariées pour les dénombrements basées sur des modèles autorégressifs d'entiers multivariés

While the literature on multivariate time series is abundant for continuous data, there is much less work on the case of multivariate time series for counts. Such data can occur in several different disciplines, including, marketing, criminology, seismology, biostatistics and many others. Observation driven models for such data can be constructed by extending the so called integer Autoregressive (INAR) models based on extensions of binomial (and related) thinning operations. In this talk we aim to present new results towards the applicability of such models in larger dimensions and also examine problems of model selection. Real data application for earthquake data will be also considered.

Tandis qu'il existe une littérature abondante sur les séries chronologiques multivariées pour les données continues, il y a bien moins de travaux sur le cas des séries chronologiques multivariées pour les dénombrements. Or de telles données sont communes dans des disciplines aussi variées que le marketing, la criminologie, la séismologie ou la biostatistique. On peut construire des modèles axés sur les observations pour ces données en étendant les modèles dits autorégressifs d'entiers (INAR) en fonction des extensions des opérations d'affinage binomiales (et connexes). Nous présentons de nouveaux résultats concernant l'applicabilité de ces modèles dans de plus grandes dimensions et examinons les problèmes de sélection de modèles. Nous étudions une application réelle pour les données sur les séismes.

RENJUN MA (University of New Brunswick), **EDWARD HUGHES** (Edward Hughes Consulting)

[Tuesday June 16 / mardi 16 juin, 10:50]

Analysis of Multivariate Longitudinal Count Data

Analyse des données de dénombrement longitudinales multivariées

We present a new approach to analysis of longitudinal multivariate count data with which possible negative cross correlation can be modelled and assessed. To handle overdispersion, serial and cross covariance structures simultaneously, we incorporate temporally correlated multivariate random effects into Poisson regression models. An orthodox best linear unbiased predictor approach has been developed in the estimation of our model. Our main results depend only on the first- and second-moment assumptions of unobserved random effects. Our proposed methodology is applicable to

Nous présentons une nouvelle approche de l'analyse des données de dénombrement longitudinales multivariées qui permet de modéliser et d'évaluer l'éventuelle corrélation croisée négative. Afin de traiter simultanément la sur-dispersion et les structures de covariance sérielle et croisée, nous intégrons des effets aléatoires multivariés corrélés dans le temps dans des modèles de régression de Poisson. Pour estimer notre modèle, nous avons mis au point une approche orthodoxe du meilleur prédicteur linéaire sans biais. Nos résultats principaux dépendent uniquement des hypothèses sur les premier et deuxième moments des effets aléatoires non observés. La méthode que nous proposons peut

2B-11: Modelling of Temporally Correlated Multivariate Counts Modélisation de décomptes multivariés corrélés temporellement

various multivariate longitudinal count data arising from biology, business, economics, forestry, health and environmental sciences.

s'appliquer à diverses données de dénombrement longitudinales en biologie, commerce, économie, foresterie, santé ou sciences environnementales.

GARY SNEDDON (Mount Saint Vincent University), **TARIQUL HASAN & RENJUN MA** (University of New Brunswick)

[Tuesday June 16 / mardi 16 juin, 11:20]

A State-Space Model for Multivariate Count Time Series with Extra Zeros

Modèle d'espace d'états pour séries chronologiques de dénombrement multivariées avec surreprésentation de zéros

Multivariate discrete time series counts with excessive zeros frequently occur in a variety of disciplines. An appropriate analysis of such data has to account for serial correlation over time, correlation between the multivariate responses, and excessive zeros. Our proposed model uses a serially correlated random effect series shared by each set of responses to account for a common temporal trend. A key feature of our model is the incorporation of a compound Poisson random effect series into a Poisson model for each response to capture excessive zeros and additional variation. We will discuss parameter estimation for this model and illustrate its use in the analysis of air pollution-related emergency room visit data.

Dans diverses disciplines, on retrouve des dénombrements de séries chronologiques multivariées discrètes à surreprésentation de zéros. Toute analyse appropriée de ces données doit tenir compte de l'autocorrélation dans le temps, de la corrélation entre les réponses multivariées et de la surreprésentation de zéros. Nous proposons un modèle qui utilise une série à effets aléatoires autocorrélée partagée par chaque ensemble de réponses pour tenir compte d'une tendance temporelle commune. L'une des fonctionnalités fondamentales de notre modèle est qu'il intègre pour chaque réponse une série de Poisson composée à effets aléatoires dans un modèle de Poisson afin de tenir compte de la surreprésentation de zéros et de toute autre variation éventuelle. Nous discutons de l'estimation des paramètres pour ce modèle et en illustrons l'utilisation pour des données sur les visites aux urgences liées à la pollution atmosphérique.

2B-I2: Recent Advances in Copula Dependence Modelling Avancées récentes en modélisation de la dépendance au moyen de copules

Chair/Présidente: Debbie Dupuis (HEC Montréal)
Organizer/Responsable: Elif Acar (University of Manitoba)

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

LOUIS-PAUL RIVEST (Université Laval), **FRANÇOIS VERRET** (Statistique Canada), **SOPHIE BAILLARGEON** (Université Laval)

[Tuesday June 16 / mardi 16 juin, 10:20]

Copulas and Small Area Estimation

Copules et estimation dans des petits domaines

This talk presents a new application of multivariate exchangeable copulas. They are used to construct estimators of the means of a variable of interest Y in several small domains of a population. In survey sampling, this problem is known as small area estimation (Rao, 2003). The standard method uses a normal regression model (Battese, Harter and Fuller, 1988) featuring random intercepts that create a within area dependency. Exchangeable copulas provide an alternative way of modeling this dependency for data that is not normal, a common occurrence in survey sampling. Several issues will be discussed, including model selection, mean square error estimation and non-linear link functions.

Cet exposé présente une nouvelle application des copules multivariées échangeables. Elles sont utilisées pour construire des estimateurs de la moyenne d'une variable d'intérêt Y dans plusieurs petits domaines de la population. C'est le problème de l'estimation dans des petits domaines en échantillonnage (Rao, 2003). La méthode standard utilise un modèle de régression normal (Battese, Harter et Fuller, 1988) avec des ordonnées à l'origine aléatoires pour créer de la dépendance dans les petits domaines. Les copules échangeables donnent une autre façon de modéliser la dépendance pour des données non-normales que l'on retrouve fréquemment en échantillonnage. Plusieurs points seront abordés, entre autres la sélection de modèle, l'estimation de l'erreur quadratique moyenne et des fonctions de liens non normales.

BRUNO RÉMILLARD (HEC Montréal), **CHRISTIAN GENEST & JOHANNA NEŠLEHOVÁ** (McGill University)

[Tuesday June 16 / mardi 16 juin, 10:50]

Copulas for Discrete or Mixed Data and Applications

Copules pour données discrètes ou mixtes et applications

In this talk I will introduce the multilinear empirical copula for discrete or mixed data and its asymptotic behavior will be studied. This result will then be used to construct inference procedures for multivariate data. Applications for testing independence will be presented.

Dans cet exposé, on introduira la copule empirique multilinéaire pour des données discrètes ou mixtes, et son comportement asymptotique sera étudié. Ce résultat servira ensuite à construire des procédures inférentielles pour des données multidimensionnelles. Des applications pour les tests de dépendance seront présentées.

ELIF ACAR (University of Manitoba)

[Tuesday June 16 / mardi 16 juin, 11:20]

Non-simplified Vine Copula Models

Modèles de copules en vignes non simplifiées

2B-I2: Recent Advances in Copula Dependence Modelling

Avancées récentes en modélisation de la dépendance au moyen de copules

Vine copulas are highly flexible multivariate dependence models built from a cascade of bivariate copulas via iterative conditioning. A large number of estimation techniques for these models have been proposed, but almost all make the assumption that conditional pair copulas depend on the conditioning variables only through their arguments. This talk will present a nonparametric strategy that relaxes this so-called simplifying assumption. In order to accommodate multiple conditioning variables in conditional pair copulas, we consider an additive formulation of the copula parameter and estimate smooth component functions associated with each conditioning variable via a local likelihood backfitting algorithm. The finite sample performance of the proposed approach will be demonstrated using simulated and real data.

Les copules en vignes sont des modèles multivariés de dépendance très flexibles construits à partir de copules bivariées via un conditionnement itératif. Un grand nombre de techniques d'estimation pour ces modèles ont déjà été proposées, mais presque toutes présument que les copules par paires conditionnelles dépendent des variables de conditionnement seulement par leurs arguments. Cet exposé présentera une stratégie non paramétrique qui assouplit cette soi-disant hypothèse de simplification. Pour accommoder plusieurs variables de conditionnement dans les copules par paires conditionnelles, nous considérons une formulation d'additif du paramètre de la copule et nous estimons des fonctions à composantes lisses associées à chaque variable de conditionnement via un algorithme de rétro-ajustement de la vraisemblance. Le rendement en échantillon fini de l'approche proposée sera démontré à l'aide de données réelles et simulées.

2B-I3: Recent Advances in Infectious Disease Epidemiology Avancées récentes et épidémiologie des maladies infectieuses

Organizer and Chair / Responsable et président: Rob Deardon (University of Calgary)

Room/Salle: Rowe 1020

Abstracts/Résumés

ELAINE NSOESIE (Harvard University)

[Tuesday June 16 / mardi 16 juin, 10:20]

Modeling the Chikungunya Epidemic in the Caribbean

Modélisation de l'épidémie de chikungunya dans les Caraïbes

Since identification of the chikungunya virus in 1953, outbreaks of chikungunya have mostly been restricted to Africa and Asia, where it is endemic in some countries. In December 2013, two cases of chikungunya were confirmed in Saint Martin, an island in the Caribbean. The outbreak quickly spread to several nations in the Americas, affecting 1 million cases (confirmed and suspected) within a year. We performed spatio-temporal analysis to describe the early spread of chikungunya virus in Dominica. We also mapped the occurrence and assessed environmental suitability of chikungunya globally.

Depuis l'identification du virus chikungunya en 1953, les épidémies de la maladie ont essentiellement été circonscrites à l'Afrique et l'Asie, où le virus est endémique dans certains pays. En décembre 2013, deux cas d'infection au virus chikungunya ont été confirmés à Saint-Martin, une île des Caraïbes. L'épidémie s'est rapidement propagée dans plusieurs pays des Amériques, affectant un million de personnes (cas confirmés et suspects) en moins d'un an. Nous examinons une analyse spatio-temporelle décrivant la propagation rapide du virus chikungunya en Dominique. Nous présentons aussi une cartographie de l'événement et une évaluation globale de l'adéquation environnementale du virus chikungunya.

PATRICK BROWN (University of Toronto)

[Tuesday June 16 / mardi 16 juin, 10:50]

Having Fun with the Spatio-Temporal Log-Gaussian Cox Process

Amusons-nous avec le processus Cox log-gaussien spatio-temporel

The locations and dates of murders in Toronto are explored using a Log-Gaussian Cox Process (LGCP). The LGCP is the simplest model for data of this sort which is both inhomogeneous (intensity varies with population) and allows for spatio-temporal dependence (cases grouping together in "clusters"). An LGCP models an intensity surface as a log-linear combination of covariates and offsets, and adds a spatio-temporally correlated Gaussian random field. Bayesian inference can be used to estimate model parameters (i.e. strength of correlation in space and time) and detect 'hot spot' areas with abnormally high or low numbers of cases. As a conditionally independent model, the LGCP is undoubtedly the wrong model to use for

Les lieux et dates des meurtres à Toronto sont explorés en utilisant un processus Cox log-gaussien spatio-temporel (PCLG). Le PCLG est le modèle le plus simple pour les données de ce genre qui est à la fois hétérogène (l'intensité varie en fonction de la population) et inclut la dépendance spatio-temporelle (regroupement en « grappes »). Un modèle PCLG construit une surface d'intensité comme combinaison log-linéaire de covariables et compensations, et ajoute un champ aléatoire gaussien spatio-temporelle. L'inférence bayésienne estime les paramètres du modèle et détecte les « points chauds » avec des nombres anormalement élevés ou faibles de cas. Parce que c'est un modèle conditionnellement indépendant, le PCLG est le mauvais modèle à utiliser pour les données de maladies infectieuses. Il pourrait être utile néanmoins.

2B-I3: Recent Advances in Infectious Disease Epidemiology Avancées récentes et épidémiologie des maladies infectieuses

infectious disease data. It could be useful nonetheless.

RAZVAN ROMANESCU (University of Guelph), **ROB DEARDON** (University of Calgary)

[Tuesday June 16 / mardi 16 juin, 11:20]

Explaining Patterns in Influenza Count Data via Epidemic Network Models

Explications des tendances dans les données de dénombrement de la grippe à l'aide de modèles épidémiologiques en réseau

Network models are versatile at capturing heterogeneity in populations during epidemics. It is known to epidemiologists that highly connected individuals tend to get infected at a higher rate and sooner during an outbreak than those with less connections. It is this feature that we investigate on an influenza dataset from the CDC. We start with an SIR model for influenza-like illness (ILI) counts, and discuss to what extent it is possible to make simple modifications to this model to reflect the changing network structure during an outbreak. Our motivation is to find network features that might account for the observed peaks and drops in transmissibility during the flu season.

Les modèles en réseau sont polyvalents pour saisir l'hétérogénéité des populations lors d'épidémies. Les épidémiologistes savent que les personnes qui ont de nombreux contacts sont plus souvent et plus rapidement infectées lors d'une flambée que celles qui entretiennent moins de relations. Nous étudions cet aspect au moyen d'un jeu de données grippales du CDC. Nous présentons d'abord un modèle SIR pour les dénombrements de maladies pseudo-grippales, puis nous examinons dans quelle mesure il est possible d'apporter de simples modifications à ce modèle pour refléter la structure changeante du réseau au cours d'une flambée. Notre objectif est de trouver des caractéristiques du réseau qui peuvent expliquer les pics et les chutes de la transmissibilité pendant la saison grippale.

2B-14: Statistical Inference in the Presence of Missing Survey Data Inférence statistique en l'absence de certaines données de sondage

Chair/Présidente: Jenny Thompson (US Census Bureau)
Organizer/Responsable: David Haziza (Université de Montréal)

Session sponsored by the Survey Methods Section
Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: Rowe 1028

Abstracts/Résumés

JAE-KWANG KIM (Iowa State University), **SHU YANG** (Harvard University)

[Tuesday June 16 / mardi 16 juin, 10:20]

A Note on Multiple Imputation for General-Purpose Estimation

Note sur l'imputation multiple pour l'estimation ordinaire

Multiple imputation is a popular imputation method for general-purpose estimation. The variance estimation formula of Rubin (1987) is very simple and easy to apply. However, the validity of the multiple imputation inference requires the congeniality condition of Meng (1994). This paper presents a theoretical justification for the upward asymptotic bias of Rubin's variance estimator when the Method-of-Moment estimator is used as a complete-sample estimator in the multiple imputation procedure. A simple modification of Rubin's variance estimator is proposed to provide asymptotically valid inference for general purpose estimation.

L'imputation multiple est une méthode d'imputation populaire pour l'estimation ordinaire. La formule d'estimation de la variance de Rubin (1987) est très simple et facile à appliquer. Cependant pour être valide, l'inférence de l'imputation multiple exige la condition de compatibilité de Meng (1994). Cet article présente une justification théorique du biais asymptotique vers le haut de l'estimateur de la variance de Rubin lorsque l'estimateur de la méthode des moments est utilisé comme estimateur de l'échantillon complet dans la procédure d'imputation multiple. Nous proposons une modification simple de l'estimateur de la variance de Rubin qui permet une inférence asymptotiquement valable de l'estimation ordinaire.

CHANGBAO WU, MIN CHEN & MARY THOMPSON (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 10:50]

Empirical Likelihood Methods for Two Survey Samples with Data Missing-by-Design

Méthodes de vraisemblance empirique pour deux échantillons de sondage avec données délibérément omises

In this talk we first present empirical likelihood methods for analyzing data from pretest-posttest studies where baseline information can effectively be used to provide more powerful tests on the response variable, which is missing by design. We then discuss methods for analyzing survey data obtained through two different modes for data collections. This is a two-sample problem but with unique features and challenges for analysis. Some limited simulation results and an analysis of the

Nous commençons par présenter des méthodes de vraisemblance empirique permettant d'analyser les données d'études pré-test/post-test, dans lesquelles des informations de base peuvent être utilisées pour donner des tests plus performants sur la variable réponse, qui est délibérément omise. Nous discutons ensuite de méthodes permettant d'analyser les données de sondage obtenues par deux modes de collecte de données différents. Il s'agit d'un problème de deux échantillons, mais avec des caractéristiques et des problèmes d'analyse uniques. Nous présen-

2B-I4: Statistical Inference in the Presence of Missing Survey Data Inférence statistique en l'absence de certaines données de sondage

International Tobacco Control (ITC) Four Country Survey data will be presented.

tons quelques résultats par simulation et analysons des données tirées du sondage international d'évaluation de la lutte antitabac (ITC).

JONGHO IM, JAE-KWANG KIM & WAYNE A. FULLER (Iowa State University)

[Tuesday June 16 / mardi 16 juin, 11:20]

Two-Per-Stratum Sampling for Fractional Hot Deck Imputation

Échantillonnage à deux échantillons par strate pour imputation hot-deck fractionnaire

Hot deck imputation is popular for handling item nonresponse in survey sampling. Fractional hot deck imputation of Kim and Fuller (2004) is extended to multivariate missing data. The joint distribution of the study items is nonparametrically estimated using a discrete approximation. The discrete transformation serves to create imputation cells. The fractional imputation procedure first assigns missing items to the cells and then imputes the real observations within each imputed cell. Replication variance estimation is discussed and results from a limited simulation study presented.

L'imputation hot deck est souvent utilisée pour compenser la non-réponse partielle dans l'échantillonnage. Nous étendons l'imputation hot deck fractionnaire de Kim et Fuller (2004) aux données manquantes multivariées. La distribution conjointe des éléments de l'enquête est estimée de manière non paramétrique à l'aide d'une approximation discrète. Nous utilisons une transformation discrète pour créer des cellules d'imputation. La procédure d'imputation fractionnaire commence par affecter les éléments manquants aux cellules, puis impute les observations réelles à chaque cellule imputée. Nous discutons de l'estimation de la variance de réplication, avant de présenter les résultats d'une étude de simulation limitée.

2B-I5: Uncertainty Quantification in Ocean and Environmental Models Quantification de l'incertitude dans les modèles océanographiques et environnementaux

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

Session sponsored by the Business and Industrial Statistics Section
Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: McCain - Scotiabank

Abstracts/Résumés

OXANA CHKREBTII (Ohio State University), **DAVID CAMPBELL** (Simon Fraser University), **MARK GIROLAMI** (University of Warwick), **BEN CALDERHEAD** (Imperial College London)

[Tuesday June 16 / mardi 16 juin, 10:20]

Quantifying Numerical Uncertainty in Models of Fluid Dynamics

Quantification de l'incertitude numérique dans les modèles de dynamique des fluides

Exact inference for differential equation models requires the ability to evaluate system states explicitly across the parameter space. For complex models of fluid dynamics such solutions are rarely available in closed form leading existing inferential tools to rely on discretization and the resulting approximate likelihood. We explore the use of probability models to quantify discretization uncertainty when evaluating the model, defining a formal trade-off between accuracy and computational expenditure. The implications of this approach are illustrated on a problem of inference on a canonical model of fluid dynamics.

Afin d'établir une inférence exacte dans les modèles d'équations différentielles, il faut pouvoir évaluer explicitement les états de système dans l'ensemble de l'espace paramètres. Pour les modèles de dynamique des fluides complexes, de telles solutions sont rarement disponibles explicitement, si bien que les outils inférentiels existants doivent recourir à la discrétisation et à la probabilité approximative résultante. Nous explorons l'utilisation de modèles de probabilité pour quantifier l'incertitude de discrétisation lors de l'évaluation du modèle, ce qui permet d'établir un compromis formel entre précision et coût informatique. Nous illustrons les implications de cette approche avec un problème d'inférence tiré d'un modèle de dynamique des fluides traditionnel.

MICHAEL DOWD (Dalhousie University), **PAUL MATTERN** (University of California), **JON BRIGGS** (University of Auckland), **KATJA FENNEL** (Dalhousie University), **RENATE MEYER** (University of Auckland)

[Tuesday June 16 / mardi 16 juin, 10:50]

Uncertainty Quantification for Biological Ocean Models

Quantification de l'incertitude dans les modèles océaniques biologiques

Biological ocean models are numerical models based on discretized spatio-temporal partial differential equations that describe geophysical fluid dynamics coupled with interacting biological populations. These are computationally demanding computer codes, and usually run as deterministic simulators. In this talk, I will review some recent work in uncertainty quantification that makes use of efficient and effective surrogates for these com-

Les modèles océaniques biologiques sont des modèles numériques fondés sur des équations différentielles partielles spatiotemporelles discrétisées qui décrivent la dynamique des fluides géophysiques couplée aux populations biologiques associées. Il s'agit de codes informatiques exigeants sur le plan des calculs qui sont généralement utilisés comme des simulateurs déterministes. Dans cette présentation, je passe en revue de récents travaux en quantification de l'incertitude qui utilisent des substituts efficaces et efficaces à

2B-15: Uncertainty Quantification in Ocean and Environmental Models

Quantification de l'incertitude dans les modèles océanographiques et environnementaux

plex biological ocean models. These are explored in the context of real applications including: the use of polynomial chaos emulators for uncertainty quantification; and the problem of stochastic simulation for data assimilation making use of multivariate distributions constructed using copulas. Some current work on Gaussian process emulators for fully Bayesian ocean data assimilation will also be discussed.

ces modèles océaniques biologiques complexes. Je les étudie dans le contexte d'applications réelles comme l'utilisation d'émulateurs de chaos polynomial pour la quantification de l'incertitude, ou encore le problème de la simulation stochastique aux fins de l'assimilation des données, en utilisant des distributions multivariées construites à l'aide de copules. Je discute également des travaux actuels sur les émulateurs de processus gaussiens utilisés pour l'assimilation entièrement bayésienne des données océaniques.

WILLIAM WELCH, TIANJI SHI & DOUW G. STEYN (University of British Columbia)

[Tuesday June 16 / mardi 16 juin, 11:20]

Dynamic Space-Time Modelling of Ozone Processes

Modélisation spatiotemporelle dynamique des processus d'ozone

Monitoring data on an atmospheric pollutant, e.g., ground-level ozone, are collected over a geographical region of interest and over time, say hourly measurements. Similarly, a numerical air quality model produces simulated output over a region of grid cells and over time. We first decompose space-time data into simpler components relating to either space or time. We model the components using Gaussian processes (GPs); computation is reduced by working with the simpler components. Recombining the modelled components gives forecasts of dynamic processes. The main purpose, however, is to evaluate a numerical air-quality model, i.e., compare it with network data, via the GP components. We demonstrate the methodology using ozone episodes in the Lower Fraser Valley of British Columbia, Canada.

Des données de surveillance d'un polluant atmosphérique comme l'ozone au sol sont collectées sur une région géographique d'intérêt au cours du temps, par exemple toutes les heures. En parallèle, un modèle numérique de la qualité de l'air produit des résultats simulés sur une région d'une grille de cellules pour une durée déterminée. Nous commençons par décomposer les données spatiotemporelles en composantes plus simples relatives à l'espace ou au temps. Nous modélisons ces composantes à l'aide de processus gaussiens (PG); nous réduisons l'effort de calcul en travaillant avec ces composantes plus simples. En recombinaison des composantes modélisées, nous produisons des prévisions de processus dynamiques. L'objectif principal, cependant, est d'évaluer un modèle numérique de la qualité de l'air, c'est-à-dire de le comparer aux données réseaux, via les éléments PG. Nous démontrons notre méthodologie au moyen d'épisodes d'ozone dans la vallée du Bas-Fraser en Colombie-Britannique, Canada.

2B-C1: Biostatistics: Methodological Innovation 1 Biostatistique : innovation méthodologique 1

Chair/Président: Jason Nielsen (Carleton University)

Room/Salle: McCain 2198

Abstracts/Résumés

HAI YAN LIU, PIERRE-JÉRÔME BERGERON & IOANA SCHIOPU-KRATINA (University of Ottawa)

[Tuesday June 16 / mardi 16 juin, 10:20]

Conditional GEE for Recurrent Gap Times

Équation d'estimation généralisée conditionnelle pour événement récurrents

As in Murphy et al. (1995), Clement and Strawderman (2009), we adopt the conditional estimating functions(EF) approach to analyse recurrent events subject to censoring. While our initial EF's are similar to those of Murphy et al., we actually estimate the censored gap time using only the observed data, by solving estimating equations (EE) that are asymptotically unbiased. The roots of our EE's form a sequence of estimators of the parameters indexing the means and variances of the gap times. We prove their strong consistency using modern analytical techniques, and illustrate our results through simulations. Key words: Conditional estimating functions, gap time, consistency.

Comme dans les travaux de Murphy et coll. (1995), Clement et Strawderman (2009), nous adoptons une approche fondée sur des fonctions d'estimation conditionnelle pour analyser les événements récurrents qui font l'objet de censure. Alors que nos fonctions d'estimation de base sont similaires à celles de Murphy et coll., nous estimons en fait l'intervalle de temps censuré en utilisant seulement les données observées par la résolution d'équations d'estimation qui sont asymptotiquement non biaisées. Les racines de nos équations d'estimation forment une suite d'estimateurs des paramètres qui indexent les moyennes et les variances des intervalles de temps. Nous démontrons leur convergence forte au moyen de techniques analytiques modernes, et nous illustrons nos résultats par des simulations. Mots clés : fonctions d'estimation conditionnelle, intervalles, convergence.

NATHALIE MOON, RICHARD COOK & LEILEI ZENG (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 10:35]

Estimating Disease Incidence with Truncated and Misclassified Response Data

Estimer l'incidence de maladie avec des réponses mal classées et tronquées

Our interest lies in characterizing the risk of psoriatic arthritis (PsA) in patients with psoriasis using two registry cohorts of patients with psoriasis and PsA respectively. The former cohort yields left-truncated PsA onset times and the latter yields right-truncated PsA onset times. We combine data from these two cohorts using likelihood and inverse probability weighting (IPW) methods, as well as an expectation-maximization algorithm. An adaptation is developed to deal with the fact that the PsA outcome status is uncertain for some individuals in the psoriasis cohort since their final assessment is based on an inaccurate screening tool.

En se basant sur des registres de personnes atteintes de psoriasis et d'arthrite psoriasique, nous visons à caractériser l'incidence d'arthrite psoriasique chez les individus atteints de psoriasis. La première cohorte consiste de temps d'apparition d'arthrite psoriasique tronqués à gauche tandis que la deuxième consiste de temps tronqués à droite. Nous proposons deux méthodes pour combiner ces cohortes : une méthode de vraisemblance et une méthode de pondération de probabilité inverse. Nous considérons aussi le fait que le temps d'apparition d'arthrite psoriasique est classifié incorrectement pour certains individus parce que leur évaluation finale est basée sur un test de dépistage inexact.

2B-C1: Biostatistics: Methodological Innovation 1 Biostatistique : innovation méthodologique 1

OMIDALI AGHABABAEI JAZI, MASOUD ASGHARIAN & ABBAS KHALILI (McGill University)

[Tuesday June 16 / mardi 16 juin, 10:50]

Simultaneous Parameter Estimation and Variable Selection in Cox Proportional Hazards Model for Length-Biased Data

Estimation de paramètres et sélection de variables simultanées dans les modèles de risques proportionnels de Cox pour des données à biais de longueur

Follow-up studies on prevalent cohorts are commonly conducted by epidemiologists and medical researchers when logistics or other constraints preclude the possibility of follow-up studies on incident cohorts. It is well-known that time-to-event data collected on prevalent cases do not form a representative sample from the target population. They constitute a biased sample. A type of such bias that has attracted much attention is called "length-bias". In this talk, we will discuss a class of penalized estimating equations for simultaneous parameter estimation and variable selection in the Cox proportional hazards model under length-biased and right censored data.

Les études de suivi sur des cohortes prévalentes sont généralement menées par les épidémiologistes et les chercheurs médicaux lorsque la logistique ou d'autres contraintes entravent la possibilité d'études de suivi sur des cohortes d'incidentes. Il est reconnu que les données de temps d'événement, recueillies sur des cas prévalents, ne constituent pas un échantillon représentatif de la population cible. Elles constituent un échantillon biaisé. Un tel type de biais ayant beaucoup attiré l'attention est appelé «biais de longueur». Dans cet exposé, nous discuterons d'une classe d'équations d'estimation pénalisées pour l'estimation des paramètres et la sélection de variables simultanées dans le modèle de risques proportionnels de Cox pour des données à biais de longueur et censurées à droite.

BO DENG & KEUMHEE CHOUGH CARRIERE (University of Alberta)

[Tuesday June 16 / mardi 16 juin, 11:05]

Testing Simultaneous Marginal Homogeneity in Clustered Matched-Pair Multinomial Data

Test d'homogénéité marginale simultanée avec données multinomiales groupées par paires

For matched-pair multinomial data, the Stuart-Maxwell and the Bhapkar tests are commonly used to test the marginal homogeneity. However, many studies for assessing toxicity and safety consider multiple polytomous outcomes to fully capture the treatment effects. To test the simultaneous marginal homogeneity (SMH) of such clustered matched-pair multinomial data, three tests are proposed. Furthermore, for the ordinal outcomes, three ordinal statistics which test SMH against stochastic order are proposed. The Monte Carlo simulation shows the score-type tests perform well regarding the nominal size and power even under small sample size settings. For ordinal outcomes, the ordinal statistics provide larger power.

En présence de données multinomiales par paires, les tests de Stuart-Maxwell et Bhapkar sont utilisés couramment pour tester l'homogénéité marginale. Cependant, de nombreuses études portant sur l'évaluation de la toxicité et de la sécurité tiennent compte des variables réponses polytomiques multiples pour saisir la totalité des effets du traitement. Nous proposons trois tests d'homogénéité marginale simultanée pour des données multinomiales groupées par paires, ainsi que trois statistiques ordinales pour les variables réponses ordinales qui permettent de tester l'homogénéité marginale simultanée contre un ordre stochastique. Une simulation de Monte-Carlo montre que les tests de type score offrent de bons résultats pour la taille et la puissance nominales, même en présence d'échantillons de petite taille. Pour les variables réponses ordinales, les statistiques ordinales sont plus puissantes.

2B-C1: Biostatistics: Methodological Innovation 1

Biostatistique : innovation méthodologique 1

PARISA GAVANJI (Queen's university), **BINGSHU CHEN & WENYU JIANG** (Queen's University)

[Tuesday June 16 / mardi 16 juin, 11:20]

Penalized Likelihood Ratio Test for Biomarker Threshold Model with an Unknown Cut-Point in Clinical Trials
Test du rapport des vraisemblances pénalisé pour modèle de seuil de biomarqueurs avec seuil inconnu lors d'essais cliniques

In clinical trials, new treatments may tend to benefit a subset of patients more. It is of interest to test biomarker effects in Logistic model to see if the new treatment benefits all patients in the same way. When the subset is defined by a continuous biomarker with an unknown cut-point, regularity conditions of the ordinary likelihood ratio test are not satisfied. We propose a penalized likelihood method to overcome irregularities and show that the null distribution of the test statistic converges to a mixture of $0.5\chi_2^2 + 0.5\chi_3^2$. The proposed method works well in simulation and real data application.

Dans les essais cliniques, de nouveaux traitements semblent davantage profiter à un sous-ensemble de patients. Il est intéressant de tester les effets des biomarqueurs dans le modèle logistique pour voir si le traitement profite à tous les patients de la même façon. Lorsque le sous-groupe est défini par un biomarqueur continu avec un seuil inconnu, les conditions de régularité du test du rapport des vraisemblances ordinaire ne sont pas satisfaites. Nous proposons une méthode de vraisemblance pénalisée pour surmonter les irrégularités et nous montrons que la distribution nulle de la statistique de test converge vers un mélange $0.5\chi_2^2 + 0.5\chi_3^2$. La méthode proposée fonctionne bien lors de simulations et dans l'application sur des données réelles.

2B-C2: Methodological Advances: A Smörgåsbord Progrès méthodologiques : un «pot-pourri »de présentations

Chair/Président: Linglong Kong (University of Alberta)

Room/Salle: McCain 2102

Abstracts/Résumés

FARNOOSH ABBAS AGHABABAZADEH, MAYER ALVO & DAVID R. BICKEL (University of Ottawa)

[Tuesday June 16 / mardi 16 juin, 10:20]

Local False Discovery Rate Estimate Based on Reference Set

Estimation du taux local de fausses découvertes à partir d'un ensemble de référence

Modern scientific techniques consider thousands hypothesis testing problems. The relevant statistical methods, for instance a local false discovery rate, tend to be applied to entire set of problems at hand. Such combination of hypothesis testing may increase the insignificant cases while the significant ones may be hidden. The separated analysis is proposed in estimating the local false discovery rate. Under this model, certain assumptions hold only locally in a symmetric window of each case. The performances of separation and combination analysis are compared theoretically and by considering several simulation studies.

Les techniques scientifiques modernes étudient des milliers de problèmes sur la vérification d'hypothèse. Les méthodes statistiques pertinentes, par exemple le taux local de fausses découvertes, ont tendance à être utilisées pour l'ensemble complet des problèmes existants. Une telle combinaison de tests d'hypothèse peut augmenter les cas non significatifs tout en cachant ceux qui sont significatifs. L'analyse séparée est recommandée pour le taux local de fausses découvertes. Dans ce modèle, certaines hypothèses tiennent seulement localement dans une fenêtre symétrique de chaque cas. Les performances des analyses de séparation et de combinaison sont comparées théoriquement et en étudiant plusieurs simulations.

IRAJ YADEGARI & ÉRIC MARCHAND (Université de Sherbrooke)

[Tuesday June 16 / mardi 16 juin, 10:35]

Post-Selection Point Estimation and Predictive Inference

Estimation et densité prédictive après sélection

Consider a set of k independent normal populations with unknown means, where we study the problem of selecting the population with the largest mean and to estimate its features. We propose to construct efficient predictive density estimators for the density of the selected population, addressing both the issues of selection bias and the principle of variance expansion which we will describe. Risk comparisons are performed under Kullback-Leibler, Hellinger, and more generally α -divergence losses, and we obtain improvements on usual benchmark procedures such as the maximum likelihood predictive density estimator.

À partir de k échantillons issus de lois normales avec moyennes inconnues, nous étudions le problème de sélectionner la population avec la plus grande moyenne et d'estimer ses caractéristiques. En particulier, nous nous intéressons à obtenir une densité prédictive pour la population choisie et d'évaluer cette densité sous un coût α -divergence, tel les coûts Kullback-Leibler et Hellinger, et du risque fréquentiste associé. Des améliorations par rapport à la densité prédictive de vraisemblance maximale sont obtenues en travaillant sur le biais de sélection, ainsi qu'en introduisant une inflation de la variance.

2B-C2: Methodological Advances: A Smörgåsbord Progrès méthodologiques : un «pot-pourri »de présentations

AZIZ LMOUDDEN & ÉRIC MARCHAND (Université de Sherbrooke), **WILLIAM STRAWDERMAN** (Rutgers University), **OTHMANE KORTBI** (United Arab Emirates University)

[Tuesday June 16 / mardi 16 juin, 10:50]

Estimation of Predictive Density for a Constrained Gamma Model

Estimation de densité prédictive pour un modèle Gamma sous contrainte

We consider the problem of estimation of a predictive density $q(y|\beta) \sim \text{Gamma}(\alpha_2, \beta)$ from $X \sim \text{Ga}(\alpha_1, \beta)$ where the parameters α_1, α_2 are known and $\beta \in [a, b]$ is unknown. Under Kullback-Leibler loss and the associated frequentist risk, we obtain Bayesian improvements on target predictive densities such as the best equivariant predictive density or of type "plug-in". We obtain minimax procedures and address the performance of predictive

densities under integrated L_1 loss.

Keywords: predictive density, Kullback-Leibler, Bayesian methods, L_1 loss, plug-in

Nous considérons le problème d'estimation d'une densité prédictive $q(y|\beta) \sim \text{Gamma}(\alpha_2, \beta)$ à partir de $X \sim \text{Ga}(\alpha_1, \beta)$ où les paramètres α_1, α_2 sont connus et $\beta \in [a, b]$ est inconnu. Sous le coût Kullback-Leibler et le risque fréquentiste associé, nous obtenons des améliorations bayésiennes sur des densités prédictives cible tels la meilleure densité prédictive équivariante ou du type "plug-in". Nous obtenons des procédures minimax et abordons la performance de densités prédictives sous la perte L_1 intégrée.

Mots-clés : Densité prédictive, Kullback-Leibler, Méthodes bayésienne, Perte L_1 , Plug-in.

NASSER SADEGHKANI & ÉRIC MARCHAND (Université de Sherbrooke)

[Tuesday June 16 / mardi 16 juin, 11:05]

Predictive Density Estimation with Additional Information

Estimation d'une densité prédictive avec information additionnelle

We consider the problem of predictive density estimation for a bivariate normal model under Kullback-Leibler (KL) and Hellinger loss functions in the presence of additional information. Specially let $\mathbf{X} = (X_1, X_2)^T \sim \text{N}_2((\theta_1, \theta_2)^T, \sigma_X^2 \mathbf{I}_2)$ and $Y_1 \sim \text{N}(\theta_1, \sigma_Y^2)$ be independent with $\theta_1 \geq \theta_2$.

We describe Bayes predictive density estimators of Y_1 for truncated normal priors, which involve skewed normal distributions. We show that the generalized Bayes predictive density estimator for the uniform prior on the restricted parameter space dominates the minimum risk equivariant predictive density estimator $\hat{q}_{mre} \sim \text{N}(x_1, \sigma_X^2 + \sigma_Y^2)$.

Nous étudions le problème qui vise à préciser une densité prédictive pour la densité de $Y_1 \sim \text{N}(\theta_1, \sigma_Y^2)$ à partir de $\mathbf{X} = (X_1, X_2)^T \sim \text{N}_2((\theta_1, \theta_2)^T, \sigma_X^2 \mathbf{I}_2)$ et avec l'information additionnelle $\theta_1 \geq \theta_2$. Sous les pertes Kullback-Leibler et Hellinger, nous décrivons les densités prédictives bayésiennes associées à des lois a priori de lois binormales tron-

quées. Ces dernières font intervenir des lois normales asymétriques. Enfin, nous démontrons que la densité prédictive de Bayes généralisée associée à la loi uniforme sur l'espace tronqué des paramètres domine la densité prédictive $\hat{q}_{mre} \sim \text{N}(x_1, \sigma_X^2 + \sigma_Y^2)$ qui est la meilleure équivariante.

ADAM RAHMAN & WAYNE OLDFORD (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 11:20]

Minimal Spanning Tree Based Data Configurations -- Construction and Distributions

Configurations de données basées sur des arbres couvrants minimaux -- Construction et distributions

Interesting low dimensional data structure can be determined from the geometric configuration of minimal spanning trees of projected data, as pro-

Une structure de données de dimension faible intéressante peut être déterminée à partir d'une configuration géométrique d'arbres couvrant minimaux de données projetées,

2B-C2: Methodological Advances: A Smörgåsbord Progrès méthodologiques : un «pot-pourri »de présentations

posed by the "scagnostic" measures of Wilkinson and Wills (2008). We explore the null distributions of these measures and, for non-null comparisons, propose and illustrate an algorithm that generates data configurations embedded in 2d having a given set of minimal spanning tree distances (and optionally constrained to exhibit other visually interesting structure). These configurations are solutions of a more general constrained Euclidean Distance Matrix Completion Problem. Applications and the problem of embedding in a higher dimension will also be discussed.

comme le proposent Wilkinson et Wills (2008) avec les mesures provenant de diagnostics de nuages de points. Nous explorons les lois sous l'hypothèse nulle de ces mesures, et pour des comparaisons non nulles, nous proposons et illustrons un algorithme qui produit des configurations de données intégrées en 2D qui ont un ensemble de distances d'arbres couvrant minimaux données (et si nécessaire, qui doivent montrer d'autres structures visuelles intéressantes). Ces configurations constituent des solutions en cas de problème pour la réalisation d'une matrice de distance euclidienne contrainte plus générale. On discutera également des applications et du problème d'intégration dans une dimension plus élevée.

MATTHEW PENCER (McGill University)

[Tuesday June 16 / mardi 16 juin, 11:35]

Post-Selection Inference in Multinomial Logistic Regression

Inférence post-sélection en régression logistique multinomiale

With Big Data becoming more prominent, variable selection techniques such as the Lasso are increasing in popularity. In this talk, I will first demonstrate the invalidity of classical significance tests performed on models selected through regularization. I will then review recently developed methods to implement statistically proper post-selection inference in linear and general linear models. Finally, I will explain how multinomial models add an extra layer of difficulty when assessing significance, and I will present my research on how to perform inference in these situations.

À l'ère des données volumineuses, les techniques de sélection de variables comme le Lasso gagnent en popularité. Dans cet exposé, je démontre d'abord l'invalidité des tests d'hypothèses classiques effectués sur des modèles choisis par régularisation. Puis, j'examine les nouvelles méthodes de mise en œuvre d'une inférence post-sélection adéquate d'un point de vue statistique dans des modèles linéaires et linéaires généralisés. Enfin, j'explique comment les modèles multinomiaux ajoutent à la complexité pour l'évaluation de la signification et je présente mes travaux portant sur la façon de faire de l'inférence dans ces situations.

2B-C3: Statistical Science for Genetics and Genomics Science statistique pour la génétique et la génomique

Chair/Présidente: Karen Kopciuk (Cancer Epidemiology and Prevention Research, AHS/University of Calgary)

Room/Salle: McCain 2170

Abstracts/Résumés

SHOFIQL ISLAM, SONIA ANAND, JEMILA HAMID, LEHANA THABANE & JOSEPH BEYENE (McMaster University)

[Tuesday June 16 / mardi 16 juin, 10:20]

Comparing the Performance of Linear and Nonlinear Principal Components in the Context of High Dimensional Genomic Data Integration

Comparer la performance des composantes principales linéaires et non-linéaires dans le contexte de l'intégration des données génomiques de grande dimension

We integrate phenotype, gene expression and miRNA expression data to classify tumor or death of Lung Cancer patients. We apply linear and kernel principal component analysis (PCA) to reduce the dimension and subsequently integrate three domains based on several nonlinear approaches including logistic regression. We also conducted an extensive simulation to compare the performances. We observe that a reduced set of kernel principal component carries more information from a genetic process but failed to show better predictive ability in this occasion. Reducing dimension with linear PCA and a logistic regression model for classification seems to be sufficient for this purpose.

Nous intégrons phénotype, expression des gènes et données d'expression miARN pour classer la tumeur ou la mort de patients atteint du cancer du poumon. Nous appliquons une analyse en composantes principales (ACP) linéaire et de noyau afin de réduire la dimension et d'intégrer ensuite trois domaines en se basant sur plusieurs approches non linéaires, y compris la régression logistique. Nous avons également effectué une vaste simulation pour comparer les performances. Nous observons qu'un nombre réduit de composantes principales de noyau renferme plus d'information sur un processus génétique, mais a échoué, dans ce cas, à montrer une meilleure capacité prédictive. À cette fin, il semble suffisant de réduire la dimension avec une ACP linéaire et un modèle de régression logistique.

GUN HO JANG, LINCOLN STEIN & JOHN MCPHERSON (Ontario Institute for Cancer Research)

[Tuesday June 16 / mardi 16 juin, 10:35]

A Copy-Number Recalibration Method using Object Integration for Next Generation Sequencing Data

Une méthode de recalage du nombre de copies utilisant l'intégration objet pour séquencer les données de prochaine génération

Both point mutations and copy number alterations are widespread characteristics of tumour. Whole genome sequencing is one of the most popular technologies to detect both of them. Many methods have been developed for calling copy-number alteration using reads counts and non-reference allele frequencies. The differences in normalization method and presumed statistical models result in different outcomes. An object integrated copy-

Les mutations ponctuelles et les modifications du nombre de copies sont des caractéristiques répandues des tumeurs. Le séquençage complet du génome est l'une des technologies les plus populaires pour les détecter. Plusieurs méthodes ont été développées pour désigner la modification du nombre de copies en utilisant les comptes lus et les fréquences d'allèles non-référence. Les différences entre la méthode de normalisation et les modèles statistiques présumés aboutissent à des résultats différents. Une méthode de recalage

2B-C3: Statistical Science for Genetics and Genomics Science statistique pour la génétique et la génomique

number recalibration method is proposed to evaluate and improve copy-number calls. The performance of the proposed method is presented and compared with several copy-number callers including HMMcopy, TitanCNA and Control-FREEC.

du nombre de copies utilisant l'intégration objet est proposée afin d'évaluer et d'améliorer les appels du nombre de copies. La performance de la méthode proposée est présentée et comparée à plusieurs appelants du nombre de copies, y compris HMMcopy, TitanCNA et Control-FREEC.

LI LI, HONG GU & TOBY KENNEY (Dalhousie University)

[Tuesday June 16 / mardi 16 juin, 10:50]

A New Approach to Detect Genetic Recombination in Phylogeny

Nouvelle approche pour détecter la recombinaison génétique dans la phylogénèse

Genetic Recombination is a process where parts of different genes are combined to form a new gene. Recombination has large effects on the phylogenetic analysis of DNA sequences, and needs to be identified correctly. We propose a new algorithm for recombination detection based on the site log-likelihood matrix. The method is computationally efficient and shows great ability to detect recombination events. Performance of the method is evaluated on both simulated and real data examples. Results are compared with other methods. abstract

La recombinaison génétique est un processus dans lequel les parties de différents gènes sont combinées pour former un nouveau gène. La recombinaison a de nombreux effets sur l'analyse phylogénétique de séquences d'ADN, et doit être identifiée correctement. Nous proposons un nouvel algorithme permettant de détecter la recombinaison en fonction de la matrice locale du logarithme de la vraisemblance. Cette méthode informatique est efficace et montre de grandes capacités à détecter des recombinaisons. Le rendement de cette méthode est évalué au moyen d'exemples de données simulées et réelles. Les résultats sont comparés avec ceux d'autres méthodes.

KUN LIANG (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 11:05]

A Hidden Markov Tree Model for Multiple Hypotheses Testing of Gene Ontology Gene Sets

Modèle d'arbre de Markov caché pour tester les nombreuses hypothèses de l'ontologie des gènes sur des ensembles de gènes

Testing predefined gene categories has become a common practice for scientists analyzing high throughput transcriptome data. A systematic way of testing gene categories leads to testing hundreds of null hypotheses that correspond to nodes in a directed acyclic graph. The relationships among gene categories induce logical restrictions among the corresponding null hypotheses. Under a hidden Markov tree model, we develop a computationally efficient method to incorporate the dependence information among the null hypotheses. Our method provides more powerful results than existing methods that honor the logical restrictions.

Les tests sur des catégories de gènes prédéfinies sont devenus une pratique commune pour les scientifiques qui analysent les données à débit élevé provenant de transcriptomes. Une façon systématique de tester les catégories de gènes amène à tester des centaines d'hypothèses nulles qui correspondent à des nœuds dans un graphe acyclique orienté. Les relations entre les catégories de gènes entraînent des restrictions logiques entre les hypothèses nulles correspondantes. Selon un modèle d'arbre de Markov caché, nous développons une méthode informatique efficace pour incorporer l'information de dépendance parmi les hypothèses nulles. Notre méthode fournit des résultats plus puissants que ceux provenant des méthodes existantes qui respectent les restrictions logiques.

2B-C3: Statistical Science for Genetics and Genomics Science statistique pour la génétique et la génomique

FARHAD SHOKOOHI & AURÉLIE LABBE (McGill University), **DAVID STEPHENS** (McGill University)

[Tuesday June 16 / mardi 16 juin, 11:20]

DNA Methylation Analysis; A Thorough Comparison of Available Analytic Tools

Analyse de données de méthylation de l'ADN : une comparaison des outils d'analyse disponibles

DNA methylation studies have increased in number recently thanks to recent advances in next-generation sequencing and microarray technology, providing many data sets at high resolution, enabling researchers to understand methylation patterns and their regulatory roles in biological processes and diseases. Diverse methods and software create ample opportunities for researchers to do quantitative analysis, however, this also makes it difficult for practitioners to choose one that is suitable and efficient in analyzing the data. The goal of this talk is to compare the available analytic tools using real and simulated data sets to give insights into their validity, efficiency and utility.

Le nombre d'études portant sur la méthylation de l'ADN a beaucoup augmenté depuis quelques années, grâce aux progrès technologiques. Ces études engendrent la production de données à haut débit, permettant aux chercheurs de comprendre les profils de méthylation ainsi que leur rôles régulateurs dans les processus biologiques menant aux maladies. Plusieurs logiciels permettent d'analyser ces données, mais il est encore difficile pour les praticiens de choisir la méthode la plus efficace. Le but de cette présentation est de comparer les outils statistiques disponibles au travers de données réelles et simulées, afin d'établir leur efficacité, leur validité et leur utilité.

THIERRY CHEKOUO TEKOUANG, FRANCESCO STINGO, MICHELE GUINDANI & KIM-ANH DO (UT MD Anderson Cancer Center)

[Tuesday June 16 / mardi 16 juin, 11:35]

A Bayesian Predictive Model for Imaging Genetics with Application to Schizophrenia

Un modèle prédictif bayésien pour la génétique d'imagerie avec application à la schizophrénie

In this talk, I'll introduce an integrative Bayesian risk prediction model that allows to discriminate between schizophrenic patients and healthy controls, based on a sparse set of discriminatory ROIs (region of interests) and SNPs (single-nucleotide polymorphism). Inference on a regulatory network between SNPs and ROI intensities (ROI-SNP network) is used in a single modeling framework to inform the selection of the discriminatory ROIs and SNPs. Using simulation studies, we assess the performance of our method, and we apply it to a schizophrenia dataset. We confirm that some biomarkers involved in the ROI-SNP network are more likely to be discriminating.

Dans cet exposé, j'introduis un modèle bayésien intégratif de prédiction du risque qui permet de distinguer les patients schizophrènes des contrôles sains, basé sur un ensemble clairsemé de ROI (régions d'intérêt) et de SNP (polymorphisme d'un seul nucléotide) discriminantes. Une inférence sur un réseau régulateur entre les ROI et les SNP est utilisée dans une modélisation unifiée pour informer la sélection de ROI et de SNP discriminantes. En se servant d'études de simulations, nous évaluons la performance de notre méthode, et nous l'appliquons à des données de schizophrénie. Nous confirmons que certains bio-marqueurs impliqués dans le réseau ROI-SNP sont potentiellement discriminants.

2B-C4: Statistics Education Éducation en statistique

Chair/Président: Jeffrey Picka (University of New Brunswick)

Room/Salle: McCain 2184

Abstracts/Résumés

SOHEE KANG (University of Toronto Scarborough)

[Tuesday June 16 / mardi 16 juin, 10:20]

Do Collaborative Learning Strategies Improve Students' Learning in Statistics Course?

Est-ce que les stratégies d'apprentissage coopératif améliorent l'apprentissage des étudiants dans un cours de statistique?

Educational research has shown that collaborative learning, or group work often results in a higher level of learning and achievement than individual work. In order to explore how these teaching methods should be translated to the statistics classroom, the experiment was conducted in Statistics II course. In addition, the Immediate-Feedback Assessment Tools (IF-AT) are adapted for assessment. In the control group, students write a quiz individually. The experiment group adapts the collaboration strategy of "Think/Pair/Share". Students' final grades are compared as a measure of learning and also comments from course evaluations are shared as qualitative measures.

La recherche en éducation a démontré que l'apprentissage coopératif, ou le travail d'équipe, mène souvent à un niveau plus élevé d'apprentissage et de réussite que le travail individuel. Afin d'explorer comment ces méthodes d'enseignement devraient être amenées dans la salle de classe de statistique, une expérience a été effectuée dans le cours Statistique II. De plus, les outils de mesure de rétroaction immédiate (OM-RI) sont adaptés pour cette évaluation. Dans le groupe contrôle, les étudiants répondent à un quiz individuellement. Le groupe expérimental adopte la stratégie coopérative de "Penser/Pairs/Partager". Les notes finales des étudiants sont comparées afin de mesurer l'apprentissage et les commentaires des évaluations de cours sont partagés comme mesures qualitatives.

JENNIFER THORNTON (Mount Saint Vincent University)

[Tuesday June 16 / mardi 16 juin, 10:40]

Online Teaching and Assessing Student Learning

Enseignement en ligne et évaluation de l'apprentissage des étudiants

Online learning offers many advantages to both students and faculty, but that comes with disadvantages too. Students need to be taught the material, but also need guidance for submitting work electronically. Professors need to adapt to new and changing technology, and the affect it might have on how we teach the material. I use BlackBoard and Moodle, and through practice I have learned some useful tools for evaluation. The purpose of my talk is to share what I have used that works, and hear from others about what they have done as well.

L'apprentissage en ligne offre de nombreux avantages à la fois pour les étudiants et les professeurs, mais il présente aussi certains inconvénients. Il faut certes transmettre un contenu aux étudiants, mais également leur enseigner comment soumettre leurs travaux par voie électronique. Les professeurs doivent s'adapter à une nouvelle technologie en pleine évolution, ainsi qu'aux effets que celle-ci peut avoir sur nos méthodes d'enseignement. J'utilise BlackBoard et Moodle et j'ai tiré de mon expérience des leçons bien utiles concernant l'évaluation. Dans ma présentation, j'explore ce qui s'avère efficace et ce que pratiquent d'autres collègues.

2B-C4: Statistics Education Éducation en statistique

ASOKAN MULAYATH VARIYATH (Memorial University of Newfoundland)

[Tuesday June 16 / mardi 16 juin, 11:00]

Teaching Statistics to First Year Undergraduate Students - Challenges

Enseigner les statistiques aux étudiants de première année du premier cycle - Défis

Teaching statistics to first year undergraduate students is always a challenge. The main strategy is to develop statistical thinking so that effective use of statistical methods can be ensured. We developed a systematic approach to teaching statistics at undergraduate level based on HSC (Hearing, Seeing and Understanding) concept. We developed statistical applets for better statistical thinking and understanding the basic concepts. We also introduced "projects" as part of the course. The approach was implemented in our newly developed first year statistics course.

Enseigner les statistiques aux étudiants de première année du premier cycle est toujours un défi. La principale stratégie est de développer la pensée statistique afin d'assurer l'utilisation efficace des méthodes statistiques. Nous avons développé une approche systématique à l'enseignement des statistiques au premier cycle basée sur le concept EVC (Entendre, Voir et Comprendre). Nous avons développé des applets statistiques afin d'améliorer la réflexion et la compréhension des concepts de base en statistique. Nous avons aussi introduit des « projets » dans le cadre du cours. Cette approche a été mise en place dans notre cours de statistique de première année nouvellement développé.

SWARNA WEERASINGHE, KEVIN WILSON & ADAM MUIR (Dalhousie University)

[Tuesday June 16 / mardi 16 juin, 11:20]

Use of Social Media as a Secure and Supportive Learning Environment for a Graduate Level Biostatistics Course in Medicine

Utilisation des médias sociaux comme un milieu d'apprentissage sécuritaire et favorable pour un cours de biostatistique de deuxième cycle en médecine

Recently, statistical educators have been focussing attention on teaching statistics to non-mathematically oriented students using multimedia platforms. Constant global reforms in university level statistical education have taken place. These pedagogical reforms had placed more attention to promote basic statistical literacy. Little attention has been paid in how social engagement can be developed into a supervised peer learning process. Facebook learning environment is viewed as one that merges social and academic lives. In this presentation I will present the social network analysis results from the pilot data collected from discussion based learning in a peer learning activity on Facebook.

Récemment, des éducateurs en statistique ont porté de l'attention à l'enseignement des statistiques aux étudiants sans orientation mathématique à l'aide de plateformes multimédias. Des réformes globales et constantes ont eu lieu dans l'enseignement de la statistique au niveau universitaire. Ces réformes pédagogiques ont mis l'emphase sur la promotion d'une compréhension de base des statistiques. Peu d'attention a été accordée à savoir comment l'engagement social peut se transformer en un processus d'apprentissage supervisé par des pairs. L'environnement d'apprentissage de Facebook est considéré comme l'un qui fusionne la vie sociale et académique. Dans cet exposé, je présenterai les résultats d'analyse de réseau social d'après des données pilotes recueillies à partir d'apprentissage basé sur les discussions dans une activité d'apprentissage avec des pairs sur Facebook.

2C-P: Poster Session Séance d'affichage

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)

(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

Room/Salle: Student Union Building - McInness

Abstracts/Résumés

OLU AWOSOGA & BONNIE LEE (University of Lethbridge)

[Tuesday June 16 / mardi 16 juin, 12:00]

Multi-site Pilot Randomized Control Trial of Congruence Couple Therapy for Problem Gamblers

Étude pilote randomisée contrôlée multi-sites pour la thérapie de couple basée sur la congruence pour les problèmes de jeu

This study was conducted in Ontario and Alberta, Canada from 2009-2011 to compare the status of couples in treatment to control condition with usual treatment and non-treatment. Treatment couples received 12-week CCT while control couples received 3 brief check-ins over 12 weeks. Baseline mean DSM-IV gambling score = 8.7/10. Retention of the treatment couples was 89% at 2-month follow-up. Retention of control couples was 78%. Of interest is the non-intended positive effects of couple research participation on control participants. The sample (N=30; 15 couples) consisted of 66% male gamblers and 34% female. CCT treatment was found to be well accepted.

Cette étude a été menée de 2009 à 2011 en Ontario et en Alberta, au Canada, en vue de comparer le statut de couples en traitement avec des couples contrôles sans traitement ou suivant le traitement habituel. Les couples en traitement ont suivi une thérapie de couple de 12 semaines basée sur la congruence alors que les couples contrôles ont été vus lors de trois brèves rencontres réparties sur 12 semaines. La moyenne initiale du score de jeu DSM-IV était de 8,7/10. Le taux de rétention des couples traités au suivi de 2 mois était de 89 %, comparativement à 78 % pour les couples non traités. Il est intéressant de noter un effet positif inattendu sur les couples du groupe contrôle en raison de la participation à une étude. L'échantillon (N=30 ; 15 couples) était constitué à 66 % de couples dont l'homme a un problème de jeu et à 34 % de femmes ayant le même problème. La thérapie de couple basée sur la congruence s'est avérée bien acceptée.

ENTAO CHEN, YING ZHANG, HOLGER TEISMANN & KHURRAM NADEEM (Acadia University)

[Tuesday June 16 / mardi 16 juin, 12:00]

A Beta Distribution Based Hierarchical Multinomial Model for Age Structure Analysis of Wildlife Animal

Modèle multinomial hiérarchique basé sur la loi bêta pour l'analyse de la structure d'âges d'animaux sauvages

Nested with a modified Beta distribution for estimating animal survivorship shape, a hierarchical multinomial model is developed to estimate the underlying yearly age distribution by sole usage of yearly age-specific harvest counts. Evaluation of model performance and robustness, by conducting extensive Monte Carlo simulation studies, indicates the accuracy and coverage of estimation under various simulation scenarios. A further analysis of perfor-

Une loi bêta modifiée servant à estimer la forme de la survie des animaux est emboîtée dans un modèle multinomial hiérarchique développé pour estimer la distribution sous-jacente des âges annuels par usage exclusif du dénombrement annuel par âge des récoltes. L'évaluation de la performance et de la robustesse du modèle est effectuée à l'aide d'études de Monte-Carlo et montre la précision et le taux de couverture selon différents scénarios de simulation. Une analyse plus approfondie de la performance est obtenue par

2C-P: Poster Session Séance d'affichage

mance will be provided by the real application on two focal age-at-harvest datasets.

une application réelle à deux jeux de données d'âge à la récolte.

HAO (NELSON) CHEN (University of British Columbia)

[Tuesday June 16 / mardi 16 juin, 12:00]

Tree-based Methods for Emulation of a Complex Computer Model

Méthodes basées sur les arbres pour l'émulation d'un modèle informatique complexe

Many complex phenomena are difficult to investigate through controlled physical experiments. Instead, computer models become important alternatives to provide insights into such phenomena. A Gaussian Process (GP) is commonly used as a statistical surrogate for the input-output relationship of a computer model. However, a GP has a strong assumption of stationarity of the output. This drawback can be overcome by tree-based methods, which split the output space and fit separate statistical surrogates within each subregion. In this poster, we first review several tree-based methods. A comparison between different tree-based methods is conducted via simulation and the optimal method is identified.

Plusieurs phénomènes complexes sont difficiles à investiguer à l'aide d'expériences physiques contrôlées. Les modèles informatiques deviennent alors des solutions de rechange importantes pour comprendre de tels phénomènes. Un processus gaussien est souvent utilisé en tant que substitut statistique pour la relation entre les entrées et sorties d'un modèle informatique. Cependant, les processus gaussiens impliquent une forte hypothèse de stationnarité des sorties. Cet inconvénient peut toutefois être corrigé avec des méthodes à base d'arbres qui découpent l'espace de sortie et ajustent des substituts statistiques différents dans chaque sous-région. Cette affiche présente une revue de plusieurs méthodes basées sur les arbres. Une comparaison entre différentes méthodes basées sur les arbres est ensuite entreprise par simulation et la méthode optimale est identifiée.

JINGJIA CHU, REG KULPERGER & HAO YU (University of Western Ontario)

[Tuesday June 16 / mardi 16 juin, 12:00]

Modelling the Common Risk among Equities: A Multivariate Time Series Model with an Additive GARCH Structure

Modélisation du risque commun entre les actions : un modèle multivarié de séries chronologiques avec une structure GARCH additive

The DCC GARCH models (Engle and Sheppard, 2001) have been well studied to describe the conditional covariance and correlation matrixes while the common risk among series cannot be captured intuitively by the existing multivariate GARCH models. A new class of multivariate time series model with an additive GARCH structure is proposed. The dynamic conditional covariances between series are aggregated by a common risk term which has been the key to characterize the conditional correlation. The model we proposed here can be implemented to apply to institutional portfolio management to determine the weights on equities and fixed income securities.

Les modèles GARCH DCC (Engle et Sheppard, 2001) ont été bien étudiés pour décrire les matrices de covariance conditionnelle et de corrélation alors que le risque commun dans les séries ne peut pas être saisi intuitivement par les modèles GARCH multivariés existants. Une nouvelle classe de modèles de séries chronologiques multivariés avec une structure additive GARCH est proposée. La covariance conditionnelle dynamique entre les séries est agrégée par un terme de risque commun qui a été la clé pour caractériser la corrélation conditionnelle. Le modèle que nous avons proposé ici peut être mis en œuvre pour s'appliquer à la gestion de portefeuilles institutionnels afin de déterminer les poids sur les actions et les titres à revenu fixe.

2C-P: Poster Session Séance d'affichage

HYUKJUN GWEON (University of Waterloo), **LARS KACZMIREK** (GESIS–Leibniz Institute for the Social Sciences), **MATTHIAS SCHONLAU** (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 12:00]

Automated Occupational Coding Using Machine Learning

Codification automatisée des professions par l'apprentissage automatique

Occupational coding refers to categorizing a survey respondent's text answer into one of hundreds of occupation codes. Instead of manual coding, we propose automated coding using machine learning. Automated coding is a challenging problem because answers usually consist of only a few words while there are hundreds of possible categories. Nevertheless, the use of machine learning approaches may be helpful for a fraction of records. We include simulation results using occupational data from a German agency.

La codification des professions fait référence à la catégorisation d'une réponse narrative d'un répondant à une enquête dans l'une des centaines de codes des professions. À la place d'une codification manuelle, nous proposons une codification automatisée par l'apprentissage automatique. La codification automatisée constitue un défi, car les réponses consistent habituellement seulement de quelques mots alors que des centaines de catégories sont possibles. Néanmoins, les approches d'apprentissage automatique peuvent être utiles pour une fraction d'enregistrements. Nous incluons des résultats de simulation en utilisant des données sur les professions provenant d'une agence allemande.

FANG HE & DUNCAN MURDOCH (University of Western Ontario)

[Tuesday June 16 / mardi 16 juin, 12:00]

Subscene design in R package rgl

Conception de sous-scène dans le paquet rgl du progiciel R

There are several packages in R (R Core Team, 2014) handling 3D plots, such as rgl, ggplot2, etc. The advantage of rgl is mouse control to rotate or zoom the scene. There were still some limitations in rgl prior to version 0.94. Our changes were inspired by the 2D plot package grid (Murrell, 2006). We have implemented "subscenes" which have properties similar to "viewports" which is a "rectangular region that provides a context for drawing" (Murrell, 2006). These new features could form the basis of interactive editing in the future.

Il y a plusieurs paquets dans le progiciel R (R Core Team, 2014) qui traitent des graphiques en 3D, comme rgl, ggplot2, etc. L'avantage de rgl est le contrôle de la souris pour faire pivoter ou zoomer la scène. Il y avait encore des limites dans rgl avant la version 0.94. Nos changements ont été inspirés du paquet grid de graphiques en 2D (Murrell, 2006). Nous avons mis en place des « sous-scènes » qui ont des propriétés similaires à celles des « fenêtres d'affichage qui forment une « zone rectangulaire qui fournit un contexte pour dessiner » (Murrell, 2006). Ces nouvelles caractéristiques pourraient devenir la base de l'édition interactive dans le futur.

JAMES KIBERD, ADAM MUIR & KEVIN WILSON (Dalhousie University)

[Tuesday June 16 / mardi 16 juin, 12:00]

The Good, the Bad, and the Cite-Able: Study Quality and Statistical Significance as Predictors of Study Citation Rates

Le bon, le mauvais et l'invocable : qualité des études et signification statistique, comme les prédicteurs des taux de citation d'études

The frequency with which an academic article is cited is a marker of its contribution to the literature. Both higher quality studies and those reporting sig-

La fréquence à laquelle un article académique est cité indique son taux de contribution aux ouvrages publiés. Les études de haute qualité et celles qui communiquent des ré-

2C-P: Poster Session Séance d'affichage

nificant results tend to be cited more frequently in the literature compared to lower quality studies or those with null findings. We will evaluate the relative contributions of study quality, reported statistical significance, and agreement with field consensus (based on published meta-analyses) on citation frequency for studies cited in a sample of Cochrane meta-analyses. The impact of these factors on the frequency of citations may indicate potential biases in how authors evaluate the literature.

sultats importants tendent à être cités plus fréquemment dans les ouvrages publiés par rapport aux études de moins bonne qualité et celles qui présentent des résultats non significatifs. Nous analyserons les contributions relatives de la qualité des études, la signification statistique rapportée, et le consensus concernant le domaine (basé sur des méta-analyses publiées) sur la fréquence des citations pour les études mentionnées dans un échantillon de méta-analyses de Cochrane. L'incidence de ces facteurs sur la fréquence des citations peut révéler la façon probablement biaisée dont les auteurs évaluent les ouvrages publiés.

KAREN A KOPCIUK (Cancer Epidemiology and Prevention Research, AHS/University of Calgary), **JI RUAN, FARSHAD FARSHIDFAR & OLIVER F. BATHE** (University of Calgary)

[Tuesday June 16 / mardi 16 juin, 12:00]

Distribution-Based Imputation for Left-Censored Metabolomics Data

Imputation basée sur la distribution pour des données métabolomiques censurées à gauche

Mass spectrometry instruments used to measure metabolomics features have left limits of detection that can result in substantial missing data. Imputing a constant value results in both bias and reduced variability in their distributions while dropping them altogether could distort relationships with other metabolites. Since estimation in projection-based methods such as principal component and partial least squares regression models is based on variation, misleading results are likely. A distribution-based imputation method is proposed to recover the missing data values and compared with imputed constant values using simulated data. Performance is evaluated using the area under the ROC curve.

Les instruments de spectrométrie de masse utilisés pour mesurer les caractéristiques métabolomiques possèdent des limites de détection qui peuvent se traduire par un nombre important de données manquantes. L'imputation d'une valeur constante entraîne un biais et une réduction de la variabilité dans leurs distributions alors que les laisser tomber pourrait fausser les relations avec d'autres métabolites. Puisque l'estimation dans les méthodes basées sur la projection, telles que les composantes principales et les modèles de régression des moindres carrés partiels, est basée sur la variation, cela peut entraîner des résultats trompeurs. Une méthode d'imputation basée sur la distribution est proposée pour récupérer les valeurs des données manquantes et comparée à des valeurs imputées constantes à l'aide des données simulées. La performance est évaluée en utilisant l'aire sous la courbe ROC.

TAKUMA KUROSAWA, ASANAO SHIMOKAWA & ETSUO MIYAOKA (Tokyo University of Science)

[Tuesday June 16 / mardi 16 juin, 12:00]

On Asymptotic Normality of Maximum Partial Likelihood Estimators for Binary Time Series with Transition Models

Normalité asymptotique des estimateurs du maximum de vraisemblance partielle pour des séries chronologiques binaires sur des modèles de transition

In this study, we consider a transition model to deal with binary time series, which have binary observations depended on some past observations and covariates. To estimate the parameters in the model, we can use the maximum partial likelihood estimate method. We prove that the estimators have asymp-

Dans cette étude, nous considérons un modèle de transition pour traiter des séries chronologiques binaires dont les observations binaires dépendent d'observations passées et des covariables. Pour estimer les paramètres du modèle, nous pouvons utiliser la méthode d'estimation du maximum de vraisemblance partielle. Nous montrons que les estima-

2C-P: Poster Session Séance d'affichage

otic normality under some conditions. In addition, we study that asymptotical behavior of the estimators with finite sample size through simulation studies.

teurs sont asymptotiquement normaux sous certaines conditions. En outre, nous étudions ce comportement asymptotique des estimateurs avec taille finie de l'échantillon à l'aide d'études de simulation.

WAN-CHEN LEE (Health Canada)

[Tuesday June 16 / mardi 16 juin, 12:00]

Statistical Strategies for Evaluating the Exposure of Chemical Mixtures during Pregnancy

Stratégies statistiques pour évaluer l'exposition aux à des mélanges chimiques pendant la grossesse

Scientists have discovered exposures to industrial chemicals occurring early in life, either in the womb or during early stages of childhood development, could result in a mild alteration, like how the brain develops resulting in changes in attention span, learning ability, behavioural changes, or other impacts like altering where fat cells are deposited in the body or modifying the development of an organ predisposing it to cancer later in life. We develop statistical criteria to evaluate the exposure of chemical mixtures during pregnancy, which are used in risk assessment of chemicals such as mercury, lead, phthalates and persistent organic pollutants.

Des scientifiques ont découvert que l'exposition précoce, soit dans l'utérus ou pendant les premiers stades de développement de la petite enfance, à des produits chimiques industriels peut altérer le développement. Par exemple, en affectant le cerveau pour modifier le niveau d'attention, la capacité d'apprentissage, ou le comportement; ou encore par d'autres impacts tels que l'altération de l'endroit où les cellules adipeuses s'accumulent ou en modifiant le développement d'organes, menant à une prédisposition au cancer dans le futur. On développe des critères statistiques pour évaluer l'exposition à des mélanges chimiques pendant la grossesse, qui sont ensuite utilisés dans l'évaluation du risque de produits chimiques comme le mercure, le plomb, les phthalates et les polluants organiques persistants.

MICHAEL LI & BENJAMIN BOLKER (McMaster University)

[Tuesday June 16 / mardi 16 juin, 12:00]

Modeling High-Resolution Animal Telemetry Data: Hidden Markov Models and Extensions

Modélisation en haute résolution des données animales par télémétrie : modèles à chaîne de Markov cachée et extensions

Clustering time-series data into discrete groups can improve prediction as well as providing insight into the nature of underlying, unobservable states of the system. However, temporal heterogeneity and autocorrelation (persistence) in group occupancy can obscure such signals. We use latent-state and hidden Markov models (HMMs), two standard clustering techniques, to model high-resolution hourly movement data from Florida panthers. Allowing for temporal heterogeneity in transition probabilities, a straightforward but rarely explored model extension, resolves previous HMM modeling issues and clarifies the behavioural patterns of panthers.

La classification des données de séries chronologiques en groupes discrets peut améliorer la prévision en plus de fournir un aperçu de la nature des états sous-jacents, non observables du système. Cependant, l'hétérogénéité temporelle et l'autocorrélation (persistance) en occupation de groupe peuvent masquer ces signaux. Nous utilisons des modèles à chaîne de Markov d'état latent et cachée (CMC), deux techniques de classification standard, pour modéliser à haute résolution les données de mouvement des panthères de la Floride sur une base horaire. Cela permet une hétérogénéité temporelle des probabilités de transition, une extension du modèle simple, mais rarement explorée, résout les problèmes de modélisation CMC précédents et précise les comportements des panthères.

2C-P: Poster Session Séance d'affichage

JASON LOEPPKY (University of British Columbia)

[Tuesday June 16 / mardi 16 juin, 12:00]

Parameterization of the Gaussian Process for Modelling a Blackbox Function

Paramétrage du processus gaussien pour la modélisation d'une fonction de boîte noire

In this poster we discuss two alternative parameterizations of the Gaussian Process model that are often used in the literature. We begin by discussing the invariance of the two models. We first discuss the lack of invariance and further discuss the lack of a numerical invariance between these two parameterizations. We then discuss the interpretation of parameterizations and show what implications this has in terms of modelling a complex computer code. Additionally we show through a series of examples that one parametrization is typically much better in terms of prediction quality for a set of test data.

Dans cette affiche, nous discutons deux paramétrisations alternatives du modèle de processus gaussien qui sont souvent utilisées dans la littérature. Nous commençons par discuter de l'invariance des deux modèles. Nous discutons d'abord le manque d'invariance et puis discutons l'absence d'une invariance numérique entre ces deux paramétrages. Nous discutons ensuite l'interprétation de paramétrages et montrons leurs conséquences sur la modélisation du code informatique. De plus, nous montrons à l'aide d'une série d'exemples qu'une paramétrisation est habituellement bien meilleure en terme de qualité de la prévision pour un ensemble de données test.

BIN LUO, ALANNA EDGE, CHARMAINE DEAN, KATHLEEN HILL & REG KULPERGER (University of Western Ontario)

[Tuesday June 16 / mardi 16 juin, 12:00]

Spatial Statistical Tools for Genome-Wide Mutation Thunderstorm Detection under a Microarray Probe Sampling System

Outils statistiques spatiaux pour la détection d'orage de mutations pangénomique sous un système d'échantillonnage de sondes de microréseaux

In genetics, the study of mutation showers can help better understand mutagenic mechanisms. A cost-effective method is to use an organism specific genotyping array that is designed to detect mutations at defined sites on probes across the entire genome. Mutations at non-probe sites are unobserved. To establish formal statistical tools for genome-wide mutation detection, several test statistics are proposed and are based on the probe array spatial properties. Power performance of the test statistics are evaluated under Neyman-Scott clustering processes via Monte Carlo simulation. Statistics with good performance are recommended as screening tools for geneticists.

En génétique, l'étude des « pluies » de mutations peut aider à mieux comprendre les mécanismes mutagènes. Une méthode économique est d'utiliser un tableau spécifique de génotypage de l'organisme qui est conçu pour détecter les mutations à des sites définis sur les sondes pour l'ensemble du génome. Des mutations dans les sites sans sondes ne sont pas observées. Pour établir des outils statistiques formels pour la détection de mutations du génome entier, plusieurs statistiques de test sont proposées et sont basées sur les propriétés spatiales de tableau de la sonde. La performance au niveau de la puissance des tests statistiques est évaluée sous les processus de classement de Neyman-Scott à l'aide de simulations de Monte-Carlo. Les statistiques présentant de bonnes performances sont recommandées comme outils de dépistage pour les généticiens.

2C-P: Poster Session Séance d'affichage

ADAM MUIR, JAMES KIBERD & KEVIN WILSON (Dalhousie University)

[Tuesday June 16 / mardi 16 juin, 12:00]

Did a Systematic Review Just Eat Your Citation? The Impact of Reviews on the Citation Frequency of Primary Studies

Est-ce qu'un article de revue vient de dévorer votre citation? L'impact des revues sur la fréquence de citation des études primaires

Systematic reviews occupy a unique position atop the hierarchy of evidence, and are among the most influential contributions to the literature. Research has established that systematic reviews are frequently cited. However, it is as yet unclear whether reviews create citations within a given field by drawing attention to the primary literature, or merely siphon future citations away from the primary studies they cite. In the current investigation, we examine primary studies cited by Cochrane systematic reviews and explore the rates at which the primary studies are cited before and after the publication of the review.

Les articles de revue occupent une place dans le haut de la hiérarchie de la preuve et sont parmi les contributions qui ont le plus d'influence à la littérature. Il a été établi que les articles de revue sont fréquemment cités. Par contre, il n'est pas clair si ces revues créent des citations dans un domaine donné en attirant l'attention sur les publications scientifiques qu'elles citent. Dans cette enquête, nous examinons les publications scientifiques citées par les articles de revues Cochrane et explorons les taux auxquels les publications scientifiques sont citées avant et après la publication de l'article de revue.

DREW NEISH, SANJEENA SUBEDI & ZENY FENG (University of Guelph)

[Tuesday June 16 / mardi 16 juin, 12:00]

Clustering Using Mixtures of Dirichlet-multinomial Regression Models

Partitionnement des données à l'aide de modèles de régression de mélanges Dirichlet multinomiaux

Compositional analysis of the human microbiome is made possible through advanced sequencing techniques, where the output consists of abundance of different bacterial taxa in each microbiome sample. Previously, a Dirichlet-multinomial mixture model has been used for modelling such microbial metagenomic data, where each mixture component represents distinct meta-communities that show similar biota compositions. However, identifying the association of environmental/biological covariates with abundance in different meta-communities remains an important problem. Here, a mixture of Dirichlet-multinomial regression models is proposed and illustrated. These models allow for a probabilistic investigation of the relationship between bacterial abundance and biological/environmental covariates within each inferred meta-community.

Les analyses de composition du microbiome humain sont possibles à l'aide de méthodes de séquençage avancées, où le résultat est une abondance de taxons bactériens différents dans chaque échantillon de microbiome. Précédemment, modèle de mélanges de Dirichlet multinomial était utilisé pour modéliser de telles données métagénomiques de microbes, où chaque composante du mélange représente des méta-communautés distinctes qui présentent des compositions de biotes semblables. Par contre, l'identification de l'association de covariables environnementales ou biologiques avec l'abondance dans différentes méta-communautés demeure un problème sérieux. Ici, un modèle de régression de mélanges Dirichlet multinomiaux est proposé et présenté. Ces modèles permettent une recherche probabiliste de la relation entre l'abondance bactérienne et les covariables environnementales ou biologiques au sein de chaque méta-communauté présumée.

2C-P: Poster Session Séance d'affichage

AURÉLIEN NICOSIA, THIERRY DUCHESNE & LOUIS-PAUL RIVEST (Laval University)

[Tuesday June 16 / mardi 16 juin, 12:00]

A General Directional Random Walk Model: Application to Animal Movement

Un modèle général de marche aléatoire directionnelle : application au mouvement animalier

We propose a general directional random walk model to describe the movement of an animal that takes into account features of the environment. A circular-linear process models the direction and distance between two consecutive localizations of the animal. A hidden process structure enables modeling situations where the animal exhibits various movement behaviors. The main originality of the proposed approach is that many environmental targets can be simultaneously included in the directional model. The model is fitted using the EM algorithm. We illustrate its use by modeling the movement of an animal in the Canadian boreal forest.

Dans cette affiche, nous proposons un modèle de marche aléatoire directionnelle général pour décrire le mouvement d'un animal qui prend en compte les caractéristiques de l'environnement. Un processus circulaire-linéaire modélise la direction et la distance entre deux localisations consécutives de l'animal. Une structure de processus cachée permet des situations où l'animal manifeste des comportements de mouvement différents. L'originalité principale de l'approche proposée est que beaucoup de cibles environnementales peuvent être simultanément incluses dans le modèle directionnel. L'inférence du modèle utilise l'algorithme EM. Nous illustrons son utilisation en modélisant le mouvement d'un animal dans la forêt boréale du Canada.

ANDREW PORTER, O. BRIAN ALLEN & GARY UMPHREY (University of Guelph)

[Tuesday June 16 / mardi 16 juin, 12:00]

Estimating an Experimental Error Variance for Fractional Factorial Designs

Estimation de la variance de l'erreur expérimentale pour des plans factoriels fractionnaires

In analyzing fractional factorial designs it can be uncertain which higher order interactions are inactive and thus can be pooled to estimate the experimental error variance. It is even less clear how pooling mean squares that are contaminated by an active effect may affect bias or the Type I error. We present simulation results on the performance of four methods for constructing the estimated experimental error variance when one of the mean squares may be active. Replacing the largest contribution to the error MS with the expectation of the maximum order statistic from a chi-square distribution demonstrates good bias reduction.

En analysant les plans factoriels fractionnaires il peut être incertain quelles interactions d'ordre supérieur sont inactives et ainsi elles peuvent être groupées afin d'estimer la variance de l'erreur expérimentale. Il est encore moins clair comment grouper les moyennes des carrés qui sont contaminées par un effet actif peut avoir un effet sur le biais ou une erreur de Type I. Nous présentons des résultats de simulation sur la performance de quatre méthodes pour définir la variance de l'erreur expérimentale lorsqu'une des moyennes des carrés est active. En remplaçant la plus grande contribution à l'erreur quadratique moyenne nous espérons que la statistique d'ordre maximal de la loi du khi carré démontrera une baisse satisfaisante du biais.

SHI QIU, LONGHAI LI & CINDY FENG (University of Saskatchewan)

[Tuesday June 16 / mardi 16 juin, 12:00]

Cross-Validatory Model Comparison and Divergent Region Detection using iIS and iWAIC for Disease Mapping

Modèle de comparaison validé croisé et détection de région divergente en employant iIS et iWAIC pour la cartographie d'une maladie

2C-P: Poster Session Séance d'affichage

Two statistical problems arise in using Bayesian hierarchical models for disease mapping. The first is to compare goodness of fit of various models, which can be used to test different hypotheses. The second problem is to identify outlier/divergent regions with unusually high or low residual risk of disease. Leave-one-out cross-validated (LOOCV) model assessment has been used for these two problems. However, actual LOOCV is time-consuming. This paper introduces two methods, namely iIS and iWAIC, for approximating LOOCV with only Markov chain samples simulated from a posterior based on a full data set.

Deux problèmes statistiques se posent dans l'utilisation de modèles hiérarchiques bayésiens pour la cartographie des maladies. Le premier consiste à comparer la qualité d'ajustement des différents modèles qui peuvent être utilisés pour tester différentes hypothèses. Le deuxième problème est d'identifier des valeurs aberrantes/régions divergentes avec risque résiduel inhabituellement élevé ou faible de la maladie. L'évaluation du modèle selon la méthode de validation croisée «leave-one-out» (LOOCV) a été utilisée pour ces deux problèmes. Cependant, la LOOCV réelle prend beaucoup de temps. Cet article présente deux méthodes, à savoir iIS et iWAIC, pour approcher LOOCV avec seulement des échantillons de chaînes de Markov simulés d'une distribution a posteriori basée sur un jeu de données complet.

RHONDA ROSYCHUK & IRAM USMAN (University of Alberta)

[Tuesday June 16 / mardi 16 juin, 12:00]

An Examination of Spatial Scan Statistics Based on Time to Event Data

Étude des statistiques de scanographie spatiale basées sur des données de durée avant événement

The spatial scan statistic (SSS) is used to identify the geographical clusters of higher number of cases of a disease. It can also identify geographic areas with longer time to events using appropriate distributions. Other authors have proposed Exponential and Weibull distributions for the event times. We propose log-Weibull as an alternative distribution for the SSS and compare the three distributions through simulation to investigate Type I censoring. Methods are also illustrated on time to specialist visit (cardiology or internal medicine) data for discharged patients presenting to Emergency Departments for atrial fibrillation and flutter in Alberta during 1999-2011.

La statistique de scanographie spatiale est utilisée pour déterminer les groupements géographiques d'un nombre plus élevé de cas de maladie. Elle permet également de détecter les zones géographiques avec un temps plus long jusqu'à l'événement au moyen de lois appropriées. D'autres auteurs ont proposé des lois exponentielle et de Weibull pour ces durées avant événement. Nous proposons la loi log-Weibull comme autre loi pour les statistiques de scanographie spatiale et nous comparons les trois lois par simulation pour analyser la censure de type I. Ces méthodes sont aussi illustrées avec des données de temps de consultations de spécialistes (cardiologues ou médecins internes) par des malades sortants qui se sont présentés au service des urgences pour une fibrillation atriale et un flutter auriculaire en Alberta de 1999 à 2011.

ASANAO SHIMOKAWA, TAKUMA KUROSAWA & ETSUO MIYAOKA (Tokyo University of Science)

[Tuesday June 16 / mardi 16 juin, 12:00]

Construction of Tree-Structured Prediction Model Based on Interval-Valued Covariates

Création de modèles arborescents de prédiction basés sur des covariables par intervalle

The survival tree, which is constructed using the CART is considered in this study. In our proposed model, covariates can be assumed as interval-valued symbolic variables. The model allows that a concept can be included in several terminal nodes in a tree-structured model. Then, the prediction

Dans cette étude, nous considérons l'arbre de survie construit à l'aide de l'algorithme CART (arbres de classification et de régression). Dans le modèle que nous proposons, les covariables peuvent servir de variables par intervalle symboliques. Ce modèle permet d'inclure un concept dans plusieurs nœuds terminaux dans un modèle arborescent. En-

2C-P: Poster Session Séance d'affichage

of a new concept will be constructed by using all terminal nodes based on observed frequency of the concept included in each terminal node. We present the application results of the proposed approach based on covariates obtained by MRI to patients with brain metastases from breast cancer to show the utility of the model.

suite, un nouveau concept sera prédit avec tous les nœuds terminaux en fonction de la fréquence observée du concept inclus dans chaque nœud terminal. Pour montrer l'utilité de ce modèle, nous présentons les résultats d'application de l'approche proposée en fonction des covariables obtenues par imagerie par résonance magnétique (IRM) sur des patients présentant des métastases cérébrales provenant du cancer du sein.

ANJALI SILVA, SANJEENA DANG, STEVEN ROTHSTEIN, GREGORY DOWNS & YONG-MEI BI (University of Guelph)

[Tuesday June 16 / mardi 16 juin, 12:00]

Comparative Analysis of Clustering Techniques for RNA-seq Data

Analyse comparative des techniques de classification des données de séquençage ARN

RNA sequencing (RNA-seq) is a deep sequencing-based approach for transcriptome profiling. RNA-seq provides counts of transcripts, offering a method to quantify gene expression. Despite the vast availability of RNA-seq data, interpreting these data in their biological context remains a challenge. Using clustering algorithms, a systematic investigation of relationships between genes can be carried out to identify genes sharing similar expression patterns. A comparative study of three clustering techniques is presented using RNA-seq data obtained from a gene expression study looking at the response of maize to nitrogen limitation. Clusters of genes identified by each method are analyzed for biological significance.

Le séquençage de l'ARN (ou RNA-seq) est une approche profonde de la profilation de transcriptomes basée sur le séquençage. Ce séquençage fournit des comptages de transcriptions, offrant ainsi une méthode pour quantifier le niveau d'expression des gènes. Or malgré la vaste disponibilité de données RNA-seq, l'interprétation de ces données dans leur contexte biologique demeure un défi. À l'aide d'algorithmes de classification, il est possible d'effectuer une étude systématique des relations entre gènes pour identifier des gènes qui partagent des profils d'expression similaires. Nous présentons une étude comparative de trois techniques de classification avec des données RNA-seq obtenues d'une étude d'expression génétique sur la réponse du maïs à la restriction de l'azote. Nous analysons les classes de gènes identifiées par chaque méthode afin de déterminer leur importance biologique.

GABRIELLE SIMONEAU (McGill University)

[Tuesday June 16 / mardi 16 juin, 12:00]

An Empirical Comparison of Methods to Meta-Analyze Individual Patient Data of Diagnostic Accuracy

Comparaison empirique de méthodes de méta-analyse des données individuelles sur la précision du diagnostic

Individual patients data (IPD) meta-analysis has many benefits. In the context of diagnostic accuracy studies, pooled sensitivity and specificity are traditionally reported for a given threshold and meta-analyses are conducted via a bivariate approach. With IPD, it is possible to obtain pooled sensitivity and specificity for each possible threshold. One way to analyze these data is to apply the BREM model at every threshold. Another idea is to analyze results for all thresholds simultane-

La méta-analyse des données individuelles (DI) présente de nombreux avantages. Dans le contexte des études sur la précision du diagnostic, la sensibilité et la spécificité communes sont généralement présentées pour un seuil donné et la méta-analyse est effectuée via une approche bivariée. Avec les DI, il est possible d'obtenir la sensibilité et la spécificité communes pour n'importe quel seuil. Une façon d'analyser ces données est d'appliquer le modèle BREM à chaque seuil. Une autre idée consiste à analyser les résultats pour tous les seuils simultanément, ce qui permet de tenir compte

2C-P: Poster Session Séance d'affichage

ously, and thus accounting for the within-study correlation. Our aim is to compare two multivariate approaches to the BREM when IPD are available, empirically.

de la corrélation au sein de l'étude. Notre objectif est de comparer deux approches multivariées au modèle BREM lorsque les DI sont disponibles, et ce de manière empirique.

YUBIN SUNG (University of Guelph)

[Tuesday June 16 / mardi 16 juin, 12:00]

A Two-Step Method for Genetic Association Analysis with Multiple Longitudinal Traits of Samples of Related Subjects

Méthode en deux étapes pour l'analyse d'associations génétiques avec des caractères longitudinaux multiples d'échantillons de sujets apparentés

We propose a two-step procedure to identify pleiotropic effects on multiple longitudinal traits from a family-based data set. The first step analyzes each longitudinal trait via a three-level mixed-effects model. Random effects at the subject-level and at the family-level measure the subject-specific genetic effects and between-subjects intraclass correlations within families, respectively. The second step performs a simultaneous association test between a single nucleotide polymorphism (SNP) and all subject-specific effects for multiple longitudinal traits. This is performed using a quasi-likelihood scoring method, in which the correlations structure among related subjects is adjusted.

Nous proposons une procédure en deux étapes pour identifier les effets pléiotropiques sur des caractères longitudinaux multiples dans un ensemble de données familial. Dans un premier temps, nous analysons chaque caractère longitudinal via un modèle à trois niveaux à effets mixtes. Les effets aléatoires au niveau du sujet et de la famille permettent de mesurer les effets génétiques spécifiques au sujet et les corrélations intraclass entre sujets au sein de la famille, respectivement. Dans un second temps, nous effectuons un test d'association simultané entre un polymorphisme du nucléotide simple (ou SNP) et tous les effets spécifiques au sujet pour de multiples caractères longitudinaux. Nous réalisons ces opérations à l'aide d'une méthode de scoring de la quasi-vraisemblance dans laquelle la structure des corrélations entre sujets apparentés est ajustée.

FODE TOUNKARA & LOUIS-PAUL RIVEST (Université Laval)

[Tuesday June 16 / mardi 16 juin, 12:00]

Archimedean Copulas for Clustered Binary Data

Copule archimédienne pour des données de Bernoulli corrélées

This presentation shows that Archimedean copulas provide several models to accommodate an extra-binomial variation in Bernoulli experiments. These models feature parameters for the marginal probability of success and a copula dependency parameter. Two applications are presented. First, we construct profile likelihood confidence intervals for the intra-cluster correlation. The second is concerned with the estimation of the closed population size from a mark-recapture study. Unit level covariates are recorded on the units that are captured and copulas are used to model a residual heterogeneity that is not accounted for by covariates. A particular copula model can be selected using the AIC.

Cette présentation montre que la famille des copules archimédiennes fournit une large classe de modèles pour la surdispersion dans une expérience de Bernoulli. Ces modèles contiennent des paramètres pour la probabilité marginale de succès et un paramètre pour la dépendance. Nous présentons deux applications. Premièrement, nous construisons un intervalle de confiance de vraisemblance en profil pour la corrélation intra-grappe. La deuxième application concerne l'estimation de la taille d'une population fermée à partir d'une expérience de capture-recapture. Nous utilisons le critère AIC pour la sélection d'un modèle.

2C-P: Poster Session Séance d'affichage

DOUGLAS WHITAKER (University of Florida)

[Tuesday June 16 / mardi 16 juin, 12:00]

Students' Understanding of Histograms and Bar Graphs: Results from the LOCUS Assessments

Compréhension estudiantine des histogrammes et diagrammes à bâtons : résultats des évaluations LOCUS

Histograms and bar graphs are standard data displays, and fluency with their use is a requisite for understanding many important statistical topics. These data displays are frequently included in K-12 mathematics curriculum documents in addition to being taught in college-level introductory statistics courses. Including histograms and bar graphs in standards documents does not necessarily imply that students understand them at the expected level. Using data from the LOCUS assessments (NSF DRL-1118168) administered to over 3500 students ages 12-18 in the United States, this poster illustrates the current level of student understanding and common misconceptions about histograms and bar graphs.

Les histogrammes et les diagrammes à bâtons sont des outils de visualisation de données standard, et la maîtrise de leur utilisation est requise pour comprendre plusieurs sujets statistiques importants. Ces formes de visualisation de données font souvent partie du curriculum mathématique au primaire et au secondaire en plus d'être enseignées dans les cours d'introduction à la statistique à l'université. L'inclusion d'histogrammes et de diagrammes à bâtons dans les documents standard n'implique pas nécessairement que les étudiants comprennent ces concepts au niveau attendu. En se servant des données des évaluations LOCUS (NSF DRL-1118168) administrées à plus de 3500 élèves de 12 à 18 ans aux États-Unis, cette affiche illustre le niveau de compréhension actuel des élèves ainsi que des fausses idées communes à propos des histogrammes et des diagrammes à bâtons.

DOUGLAS WHITAKER (University of Florida)

[Tuesday June 16 / mardi 16 juin, 12:00]

Transition and Collaboration: The Identity of an Advanced Placement Statistics Teacher

Transition et collaboration : l'identité d'un enseignant d'un cours de perfectionnement de placement en statistique

Statistics has a visible and increasing presence in the K-12 curriculum. Despite widespread recognition of statistics as a field independent of mathematics at the tertiary level and beyond, at the K-12 level statistics instruction largely occurs in mathematics classrooms. Without appropriate preparation and professional development, this arrangement can result in an incomplete treatment of statistics. Mathematics teachers charged with teaching statistical content may need to develop a new identity as a statistics teacher to be effective. This poster examines the identity of an in-service secondary level mathematics teacher engaged in collaborative teaching of an Advanced Placement (AP) Statistics course.

La statistique est de plus en plus visible et présente dans les programmes d'études primaires et secondaires. Cependant, bien que cette matière soit largement reconnue comme indépendante des mathématiques au niveau tertiaire et au-delà, au primaire et au secondaire l'instruction en statistique se fait généralement dans le cadre des cours de mathématiques. Or, en l'absence d'une préparation et d'un développement professionnel appropriés, il peut en résulter un traitement insuffisant de la matière. Les enseignants de mathématiques chargés d'enseigner un contenu statistique auront peut-être à se forger une nouvelle identité d'enseignant de statistique pour être efficaces. Cette affiche examine l'identité d'un enseignant de mathématiques en exercice dans le secondaire participant à l'enseignement en collaboration d'un cours de perfectionnement de placement en statistique.

2C-P: Poster Session Séance d'affichage

KEVIN WILSON, JAMES KIBERD & ADAM MUIR (Dalhousie University)

[Tuesday June 16 / mardi 16 juin, 12:00]

Out of Cite, Out of Mind? The Effects of Citation by Review Articles on Frequency of Citation in the Epidemiological Literature

Hors de vue, oublié? Effets de la citation par des articles de synthèse sur la fréquence de citation dans la littérature épidémiologique

The influence of a piece of academic writing is commonly gauged by the frequency with which it is cited in the literature. Systematic reviews are, unsurprisingly, some of the most frequently cited contributions, and occupy a place atop the hierarchy of evidence. In the present study, we evaluate the effect of review papers (general, systematic, and meta-analyses) on the frequency of citations among primary articles in a cohort of epidemiological studies published in 2005. Using Poisson regression with general estimating equations, we will assess whether reviews affect how often researchers cite primary literature.

L'influence d'un texte académique est couramment évaluée selon la fréquence à laquelle il est cité dans la littérature. Bien entendu, les revues systématiques figurent parmi les contributions les plus fréquemment citées et occupent une place privilégiée dans la hiérarchie des preuves. Dans cette étude, nous évaluons l'effet des articles de synthèse (généraux, systématiques et de méta-analyse) sur la fréquence des citations des articles primaires dans une cohorte d'études épidémiologiques publiées en 2005. À l'aide d'une régression de Poisson avec des équations d'estimation générales, nous cherchons à savoir si les articles de synthèse affectent la fréquence de citation de la littérature primaire par les chercheurs.

LINGYUN YE (Canadian Centre for Vaccinology), **BRUCE SMITH** (Dalhousie University), **SHELLY MCNEIL** (Canadian Centre for Vaccinology)

[Tuesday June 16 / mardi 16 juin, 12:00]

The Validity of Test-Negative Case-Control Design

Validité de la méthode cas-témoins avec tests négatifs

We study a test-negative case-control design (TNCC), which is an extension of the traditional case-control design, in which the study population consists of subjects seeking health care services due to acute respiratory illness. Subsequent laboratory testing is then performed to confirm disease outcome. Due to its simplicity, the TNCC has become the "gold standard" for estimating vaccine effectiveness (VE) for influenza and rotavirus. By modeling the case and control series as independent Poisson processes, we show that the TNCC provides a consistent estimate of VE. The rationale, interpretation and several methodological issues of TNCC designs are also discussed.

Nous étudions le devis cas-témoins avec tests négatifs (CTTN), une extension du devis cas-témoins traditionnel, dans laquelle la population à l'étude se compose de sujets accédant aux services de santé pour une maladie respiratoire aiguë. Des tests en laboratoire sont réalisés pour confirmer l'issue de la maladie. En raison de sa simplicité, la méthode CTTN est devenue la référence absolue pour estimer l'efficacité des vaccins (EV) contre la grippe et le rotavirus. En modélisant les séries de cas et de cas-témoins comme des processus de Poisson indépendants, nous montrons que le devis CTTN offre une estimation convergente de l'EV. Nous discutons aussi de la justification, de l'interprétation et de plusieurs problèmes de méthodologie que pose le devis CTTN.

BOYKO ZLATEV (University of Alberta)

[Tuesday June 16 / mardi 16 juin, 12:00]

Studying Selections of Poems by Statistical Methods

Étude de sélections de poèmes par des méthodes statistiques

2C-P: Poster Session Séance d'affichage

A sample of selections of P.B. Shelley's poems is studied by applying various statistical methods. The selections, produced by different editors in the time range from 1840 to 2014, were obtained from published selected editions and anthologies. Multidimensional scaling is applied to both selections and poems, with distances obtained computed by transforming appropriate correlation coefficients. Then cluster and classification analysis is performed on the scaled data. Conclusions about evolution and some non-obvious features of the reception of Shelley's poetry can be made from the study.

Nous étudions un échantillon de sélections de poèmes de P.B. Shelley en lui appliquant diverses méthodes statistiques. Les sélections, publiées par diverses maisons d'édition entre 1840 et 2014, sont tirées de différentes éditions et anthologies. Nous appliquons une mise à l'échelle multidimensionnelle à la fois aux sélections et aux poèmes, les distances étant calculées en transformant les coefficients de corrélation appropriés. Nous effectuons alors une analyse de groupement et de classification sur les données à l'échelle. Nous tirons de cette étude des conclusions sur l'évolution et sur certains aspects non évidents de l'accueil connu par les poèmes de Shelley.

2D-I1: Actuarial Graduate Students Presentations Présentations des étudiants gradués en science actuarielle

Chair/Président: Étienne Marceau (Université Laval)
Organizer/Responsable: Hélène Cossette (Université Laval)

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

Room/Salle: Rowe 1011

Abstracts/Résumés

SABRINA ZHANG & CARY TSAI (Simon Fraser University)

[Tuesday June 16 / mardi 16 juin, 13:30]

The Optimal Payment Reduction Ratios For a Catastrophe Bond.

Les rapports de réduction de paiements optimaux pour une obligation catastrophe

Catastrophe bonds are insurance-linked securities that transfer catastrophe risks from insurance industry to bond holders. If there is a catastrophe, the catastrophe bond is triggered and the future bond payments are reduced. This paper first presents a general pricing formula for a catastrophe bond with coupon payments, adaptable to various assumptions for a catastrophe loss process. Next, it gives formulas for the optimal payment reduction ratios which maximize two measurements of risk reduction, hedge effectiveness rate (HER) and hedge effectiveness (HE), respectively, and examines how the optimal payment reduction ratios help reinsurers or insurers to mitigate extreme losses. Last, it shows how the parameters of the models affect the optimal payment reduction ratios with numerical examples for illustrations.

Les obligations catastrophes sont des titres liés à l'assurance qui transfèrent les risques catastrophes de l'industrie de l'assurance aux détenteurs d'obligations. S'il y a une catastrophe, l'obligation catastrophe est déclenchée et les futurs paiements des obligations sont réduits. Dans cet article, on présente d'abord une formule générale de tarification pour les obligations catastrophes avec paiements de coupon, qui s'adapte à différentes hypothèses pour un processus de pertes catastrophes. Ensuite, on donne des formules pour obtenir les rapports de réduction de paiements optimaux qui maximisent deux mesures de réduction du risque, respectivement le taux d'efficacité de couverture (HER) et l'efficacité de couverture (HE), et on examine de quelle façon les rapports de réduction de paiements optimaux aident les réassureurs ou les assureurs à atténuer les pertes extrêmes. Enfin, on démontre comment les paramètres des modèles affectent les rapports de réduction de paiements optimaux à l'aide d'exemples numériques.

JI RUAN & ALEX DE LEON (University of Calgary)

[Tuesday June 16 / mardi 16 juin, 13:45]

Cluster Analysis of Gene Expression Profiles via Flexible Count Models for RNA-seq Data

Analyse de regroupement de profils d'expression génique via des modèles de dénombrement flexibles pour des données seq-ARN

Clustering RNA-seq data is used to characterize environment-induced (e.g., treatment) differences in gene expression profiles by separating genes into clusters based on their expression patterns.

Le regroupement de données seq-ARN est utilisé pour caractériser les différences dans les profils d'expression géniques causées par l'environnement (par exemple, un traitement) en séparant les gènes dans des grappes en se ba-

2D-I1: Actuarial Graduate Students Presentations Présentations des étudiants gradués en science actuarielle

Wang et al. (2014) recently adopted the bivariate Poisson (BVP) distribution, obtained via the trivariate reduction method, as a model for clustering bivariate RNA-seq data. We discuss the inadequacy of BVP in modelling the correlation between bivariate counts, and its impact on clustering such data. We introduce an alternative Gaussian copula model (GCM) that incorporates a flexible dependence structure for the counts, and report simulation results to compare the performance of GCM and BVP and investigate the impact on clustering of Poisson counts of misspecified dependence structures. We illustrate our methodology on real RNA-seq data.

sant sur leur structure d'expression. Wang et al. (2014) ont récemment adopté la loi bivariable de Poisson (BVP), obtenue via la méthode de réduction à trois variables, comme modèle pour regrouper les données seq-ARN bivariées. Nous discutons de l'insuffisance du BVP dans la modélisation de la corrélation entre les dénombrements bivariés, ainsi que son impact sur le regroupement de ces données. Nous introduisons un modèle alternatif de copule gaussienne (GCM) qui inclut une structure de dépendance flexible pour les dénombrements, et nous présentons des résultats de simulations pour comparer les performances du GCM et du BVP et nous examinons l'impact du regroupement des dénombrements de Poisson sur des structures de dépendance mal spécifiées. Nous illustrons notre méthodologie à l'aide de données seq-ARN réelles.

JEAN-FRANÇOIS BÉGIN (HEC Montréal), **MATHIEU BOUDREAU** (Université du Québec à Montréal), **GENEVIÈVE GAUTHIER** (HEC Montréal)

[Tuesday June 16 / mardi 16 juin, 14:00]

Credit Risk in Corporate Spreads during the Financial Crisis of 2008

Risque de crédit dans les rendements des obligations de sociétés durant la crise financière de 2008

Credit spreads and CDS premiums are investigated before, during and after the financial crisis with a flexible credit risk model. The latter is designed to capture empirical facts: a regime-switching framework adjusts its behaviour to the financial cycles and the negative relationship between recovery rates and default probabilities appears endogenously.

Using a firm-by-firm estimation of 225 companies, notorious empirical questions are revisited, including the famous credit spread puzzle. The proportion of the spread explained by credit risk decreases during the crisis. Liquidity plays a significant role in explaining this gap throughout the financial turmoil and persists thereafter.

On fait l'étude des écarts de rendement des obligations de sociétés et les primes de swaps de défaillance avant, pendant et après la crise financière à l'aide d'un modèle de risque de crédit flexible. Ce dernier permet de reproduire plusieurs caractéristiques empiriques : un changement de régime capte les différents cycles financiers et une relation négative entre les taux de recouvrement et les probabilités de défaut.

En utilisant les données de 225 firmes, quelques questions empiriques notoires sont revisitées, dont le fameux casse-tête des écarts de crédit. La proportion des écarts expliquée par le risque de crédit décroît durant la crise. La liquidité joue un rôle important et explique une partie de cette diminution.

ITRE MTALAI, HÉLÈNE COSSETTE, ETIENNE MARCEAU & DÉRY VEILLEUX (Université Laval)

[Tuesday June 16 / mardi 16 juin, 14:15]

Aggregation Methods for Portfolios of Dependent Risks with Archimedean Copulas

Méthodes d'agrégation des portefeuilles de risques dépendants en utilisant les copules archimédiennes

In this paper, we consider a portfolio of dependent risks represented by a vector of dependent random variables (rvs) whose joint cumulative distribution function (cdf) is defined with an Archimedean cop-

Dans cet exposé, nous considérons un vecteur de variables aléatoires (v.a.) dépendantes dont la fonction de répartition conjointe est définie avec une copule archimédienne. Les copules archimédiennes sont très populaires et leurs ex-

2D-I1: Actuarial Graduate Students Presentations Présentations des étudiants gradués en science actuarielle

ula. Archimedean copulas are very popular and their extensions, nested Archimedean copulas, are well suited for vector of rvs of high dimension. We examine the computation of the cdf of the sum or a variety of functions of those rvs. In particular, we derive the cdf and the TVaR of the sum of those risks using the Frank copula and the Ali-Mikhail-Haq (AMH) copula. The computation of the contribution of each risk under the TVaR-based allocation rule, the covariance allocation rule and the Esscher-based allocation rule is also treated.

tensions, les copules archimédiennes imbriquées, sont bien adaptées pour des vecteurs de v.a. à haute dimension. Nous examinons le calcul de la fonction de répartition de la somme ou une variété de fonctions de ces v.a. En particulier, nous examinons le calcul de la TVaR de la somme de ces risques en supposant la copule de Frank et la copule Ali-Mikhail-Haq (AMH). Le calcul de la contribution de chaque risque selon les règles basées sur la TVaR, la covariance et la transformée d'Esscher est également traité.

MIRABELLE HUYNH & DAVID LANDRIALT (University of Waterloo), **TIANXIANG SHI** (University of Nebraska-Lincoln), **GORDON E WILLMOT** (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 14:30]

On a Risk Model with Claim Investigation

Un modèle de risque avec enquête de réclamation

A queue-based claims investigation mechanism is assumed for the insurer's claims payment process to model its claim adjustment practices. The resulting model may be viewed as a first step in developing models with more realistic claim investigation mechanisms. Claim settlement delays and time dependent payments have been studied by, e.g., Taylor [1979], Cai and Dickson [2002], and Truffin et al. [2011]. However, little has been done on queue-based investigation mechanisms. We demonstrate the impact of a simple claim investigation mechanism on some common ruin-related quantities of interest when claims are assumed to occur according to a Poisson process. An exponentially distributed investigation time is considered. We conclude with numerical examples where we explore questions arising from a risk management standpoint.

Pour que le processus de paiement de réclamations de l'assureur modèle ses pratiques de frais de règlement, un mécanisme d'étude de réclamations à file d'attente est présumé. Le modèle qui en résulte peut être vu comme un premier pas dans le développement de modèles de mécanismes d'enquête de réclamations plus réalistes. Les délais dans le règlement des réclamations et les paiements en fonction du temps ont été étudiés, par exemple par Taylor [1979], Cai et Dickson [2002], et Truffin et al. [2001]. Par contre, peu de travail a été fait sur les mécanismes d'enquête à file d'attente. Nous démontrons l'impact d'un simple mécanisme d'enquête de réclamations sur quelques quantités courantes relatives à la ruine quand les réclamations sont présumées se produire selon un processus de Poisson. Un temps d'enquête distribué de manière exponentielle est considéré. Nous concluons avec des exemples numériques où nous explorons des questions résultant d'un point de vue de gestion du risque.

YISUB KYE & EDWARD FURMAN (York University), **RALUCA VERNIC** (Institute for Mathematical Statistics and Applied Mathematics, Bucharest)

[Tuesday June 16 / mardi 16 juin, 14:45]

On the probability of ruin for non-identically distributed and dependent claims

À propos de la probabilité de ruine pour des réclamations non-identiquement distribuées et dépendantes

In this talk I will consider a continuous risk model with positively dependent and non-identically distributed claim amounts. More specifically, I'll discuss the derivation of the expressions for the ruin

Dans cet exposé, je vais considérer un modèle de risque continu avec des montants de réclamation de loi non-identique et de dépendance positive. Plus précisément, je discuterai de la dérivation des expressions pour la proba-

2D-I1: Actuarial Graduate Students Presentations Présentations des étudiants gradués en science actuarielle

probability at or before a certain claim arrival instant in the situations when claim numbers follow a renewal process with either Erlang or a mixture of exponentials inter-arrival times. I will elucidate the finding with the help of numerical examples.

This is a joint work with Ed Furman of York U, Toronto, Canada and Raluca Vernic of the Institute for Mathematical Statistics and Applied Mathematics, Bucharest, Romania.

bilité de ruine au moment ou avant l'instant d'arrivée d'une réclamation dans des situations où le nombre de réclamations suit un processus de renouvellement avec des temps entre les événements de loi Erlang ou de mélange d'exponentielles. Des exemples numériques seront fournis.

Ceci est un travail conjoint avec Ed Furman de l'université York de Toronto au Canada et Raluca Vernic de l'Institut pour les statistiques mathématiques et les mathématiques appliquées de Bucharest en Roumanie.

2D-I2: Big Data and Neuroimaging Données volumineuses et neuroimagerie

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

Room/Salle: Rowe 1020

Abstracts/Résumés

SEBASTIAN KURTEK (Ohio State University), **IAN H. JERMYN** (Durham University), **QIAN XIE & ANUJ SRIVASTAVA** (Florida State University)

[Tuesday June 16 / mardi 16 juin, 13:30]

Elastic Statistical Shape Models of Anatomical Surfaces using Square-Root Normal Fields

Modèles de forme statistiques élastiques de surfaces anatomiques utilisant des champs normaux racine carrée

We present a comprehensive Riemannian framework for statistical shape analysis of 3D objects, represented by their boundaries (parameterized surfaces). We provide tools for registration (matching), comparison, averaging, principal component analysis and statistical modeling of observed shapes. This framework is based on a square-root normal field (SRNF) representation of surfaces and a related elastic metric. The elastic metric provides a natural interpretation of shape deformations that are being quantified and is invariant to reparameterizations of surfaces. Furthermore, under the SRNF transformation, the complicated elastic metric becomes the standard L2 metric, simplifying parts of the implementation. We showcase the applicability of this framework in shape analysis of anatomical structures in different medical applications including Attention Deficit Hyperactivity Disorder and endometriosis.

Nous présentons un cadre riemannien complet pour l'analyse statistique de la forme d'objets en trois dimensions, représentés par leurs frontières (surfaces paramétrées). Nous donnons des outils pour l'enregistrement (appariement), la comparaison, le calcul de moyenne, l'analyse de composantes principales et la modélisation statistique des formes observées. Ce cadre se base sur une représentation de champ normal racine carrée (CNRC) des surfaces et une métrique élastique connexe. La métrique élastique fournit une interprétation naturelle des déformations de la forme qui sont mesurées et reste invariable suite à la reparamétrisation des surfaces. De plus, lors d'une transformation CNRC, la métrique élastique complexe devient la métrique standard L2, simplifiant ainsi certains éléments de la mise en oeuvre. Nous présentons l'applicabilité de ce cadre dans l'analyse de forme de structures anatomiques pour diverses applications médicales dont le trouble du déficit d'attention avec ou sans hyperactivité et l'endométriose.

LINGLONG KONG (University of Alberta), **HONGTU ZHU** (University of North Carolina at Chapel Hill), **JIANQING FAN** (Princeton University), **XINGCAI ZHOU** (Tongling University), **PARTHA SARATHI MUKHERJEE** (Boise State University), **BAIGUO AN & CHAO HUANG** (University of North Carolina at Chapel Hill)

[Tuesday June 16 / mardi 16 juin, 14:00]

Robust Spatial Varying Coefficient Model in Neuroimaging Data Analysis

Modèle de coefficient spatial variable en analyse de données en neuro-imagerie

Neuroimaging studies aim to analyze imaging data with complex spatial patterns in a large number of locations (called voxels) on a two dimensional (2D) surface or in a 3D volume. We proposed three methods to spatially model the varying association between imaging measures with a set of covari-

Les études en neuro-imagerie ont pour objectif d'analyser les données d'imagerie à motifs spatiaux complexes dans un grand nombre d'emplacements (appelés voxels) sur une surface en deux dimensions (2D) ou dans un volume en trois dimensions (3D). Nous proposons trois méthodes pour faire une modélisation spatiale des différentes associations entre

2D-I2: Big Data and Neuroimaging Données volumineuses et neuroimagerie

ates, namely, spatially varying coefficient model (SVCM), multiscale adaptive composite quantile regression model (MACQRM), and spatially statistical parametric mapping model (SSPM). For each method, we develop a three-stage estimation procedure to simultaneously estimate the effect images and the complex spatial correlation. Theoretically, we establish consistency and asymptotic normality of the adaptive estimates and the asymptotic distribution of the test statistics. Our Monte Carlo simulation and real data analysis have confirmed their excellent performance.

les mesures d'imagerie avec un jeu de covariables, c'est-à-dire un modèle de coefficient à variation spatiale (MCVS), un modèle de régression quantile composée à multi-échelles adaptable (MRQCMA), et un modèle statistique de cartographie paramétrique spatiale (SCPS). Pour chaque méthode, nous développons une procédure d'estimation en trois étapes pour estimer simultanément les images d'effet et la corrélation spatiale complexe. Théoriquement, nous établissons la convergence et la normalité asymptotique des estimations adaptables et la loi asymptotique des statistiques de test. Notre simulation Monte Carlo et une analyse de données réelles ont confirmé leur excellente performance.

TIMOTHY JOHNSON & MING FENG (University of Michigan), **FAROUK NATHOO** (University of Victoria)
[Tuesday June 16 / mardi 16 juin, 14:30]

Bayesian Computation for Log Gaussian Cox Processes with Application to Neuroimaging
Calcul bayésien pour des processus de Cox log-gaussiens avec applications en neuro-imagerie

The Log-Gaussian Cox Process (LGCP) is a commonly used model for the analysis of spatial point pattern data. Inference on the intensity function is intensive due to the doubly stochastic property, i.e., the model is a hierarchical combination of a Poisson process and a Gaussian Process (GP), which leads to an intractable integral over an infinite-dimensional random function. As a result of these challenges a number of computational techniques have been proposed for Bayesian inference. These include Hamiltonian Monte Carlo (HMC), Integrated Nested Laplace Approximation (INLA) and Variational Bayes (VB). In this talk we describe these comparisons in terms of accuracy and computational efficiency using simulation studies as well as through an application in neuroimaging.

Le processus de Cox log-gaussien (PCLG) est un modèle fréquemment utilisé pour l'analyse des données de motifs de points spatiaux. L'inférence sur la fonction d'intensité est intensive en raison de la propriété doublement stochastique, c.-à-d., le modèle est une combinaison hiérarchique d'un processus de Poisson et d'un processus gaussien, ce qui mène à une intégrale insoluble sur une fonction aléatoire de dimension infinie. En raison de ces difficultés, un certain nombre de techniques computationnelles ont été proposées pour l'inférence bayésienne. Ces dernières comprennent le Monte Carlo hamiltonien (MCH), l'approximation de Laplace emboîtée intégrée (ALEI) et le bayésien variationnel. Dans cet exposé, nous les comparons en termes d'efficacité computationnelle et de précision en utilisant des études de simulation ainsi qu'une application en neuro-imagerie.

2D-I3: Recent Developments in Nonparametric Statistics Développements récents en statistique non paramétrique

Chair/Président: Paul Cabilio (Acadia University)

Organizer/Responsable: Ying Zhang (Acadia University)

Room/Salle: McCain 2102

Abstracts/Résumés

MAYER ALVO (University of Ottawa)

[Tuesday June 16 / mardi 16 juin, 13:30]

Bridging the Gap: A Nonparametric Likelihood Function for Ranking Data

Comblant le fossé : une fonction de vraisemblance non paramétrique pour des données de classement

In the parametric setting in statistics, the notion of a likelihood function forms the basis for the development of tests of hypotheses and estimation of parameters. Tests in connection with the analysis of variance stem entirely from considerations of the likelihood function. On the other hand, the empirical likelihood method which is entirely data driven, presents an alternative to the parametric notion. In the present article, we define a likelihood function motivated by characteristics of the ranks of the data. Such a likelihood function can be fruitfully used in several problems in nonparametric statistics involving the use of ranks. This paper points as well to additional areas of exploration.

Dans le cadre paramétrique des statistiques, la notion de fonction de vraisemblance constitue la base pour le développement de tests d'hypothèse et pour l'estimation des paramètres. Les tests reliés à l'analyse de la variance proviennent entièrement de considérations sur la fonction de vraisemblance. D'autre part, la méthode de vraisemblance empirique, qui est entièrement guidée par les données, présente une alternative à la notion paramétrique. Dans cet article, nous définissons une fonction de vraisemblance motivée par les caractéristiques des rangs des données. Une telle fonction de vraisemblance peut être employée de manière utile dans plusieurs problèmes de statistique non paramétrique utilisant les rangs. Cet article montre également d'autres domaines d'exploration.

HEROLD DEHLING (Ruhr-Universitaet Bochum), **OLIVIER DURIEU** (Universite de Tours), **MARCO TUSCHE** (Ruhr-University Bochum), **DALIBOR VOLNY** (Universite de Rouen)

[Tuesday June 16 / mardi 16 juin, 14:00]

New Techniques for Empirical Processes of Dependent Data

Nouvelles techniques pour les processus empiriques de données dépendantes

Empirical processes are a central object of research in probability theory and mathematical statistics. In our talk, we will explain some new techniques for the study of empirical processes when the underlying data are on the border of short-range dependence. We will present empirical process central limit theorems that apply, e.g., when the underlying data arise from Markov chains or from deterministic dynamical systems. As a concrete example, we will give empirical process CLTs for ergodic torus automorphisms and for certain linear processes, indexed by classes of functions.

Les processus empiriques sont un objet central de recherche en théorie des probabilités et en statistique mathématique. Dans notre exposé, nous expliquerons quelques techniques nouvelles pour l'étude des processus empiriques lorsque les données sous-jacentes sont à la frontière de dépendance de courte portée. Nous présenterons un processus empirique des théorèmes limites centraux qui s'applique, par exemple, lorsque les données sous-jacentes proviennent de chaînes de Markov ou de systèmes dynamiques déterministes. Dans un exemple concret, nous donnerons un processus empirique des TLC pour des automorphismes du tore ergodique et pour certains processus linéaires, indexés par des classes de fonctions.

2D-I3: Recent Developments in Nonparametric Statistics Développements récents en statistique non paramétrique

YING ZHANG, PAUL CABILIO & AMBER CORKUM (Acadia University)

[Tuesday June 16 / mardi 16 juin, 14:30]

Robust Confidence Interval Construction with Applications to Time Series Intervention Analysis

Construction d'intervalles de confiance robustes avec applications à l'analyse d'intervention en séries chronologiques

In this talk, we propose a new approach for constructing rank-based robust confidence intervals for the mean/median in time series data. The dependent nature of the data in time series analysis poses difficulties in constructing robust confidence intervals. The key to such construction lies in the appropriate estimation of the variance of the statistics considered. We discuss and compare two variance estimators: the bootstrap variance and the partial sum variance of dependent process. Finally we will demonstrate our approach to an investigation of the effect of a provincial insurance policy change on the consumption of a heartburn medication. This is a joint work with Paul Cabilio and Amber Corkum.

Dans cet exposé, nous proposons une nouvelle approche pour construire des intervalles de confiance robustes basés sur les rangs pour la moyenne/médiane des données de séries chronologiques. La nature dépendante des données dans l'analyse des séries chronologiques pose des difficultés dans la construction d'intervalles de confiance robustes. La clé de cette construction réside dans l'estimation appropriée de la variance des statistiques considérées. Nous discutons et comparons deux estimateurs de la variance : la variance bootstrap et la somme partielle des variances du processus dépendant. Finalement, nous présenterons notre approche pour une enquête sur l'effet d'un changement de la politique d'assurance provinciale sur la consommation d'un médicament de brûlures d'estomac. Ce travail est réalisé conjointement avec Paul Cabilio et Amber Corkum.

2D-I4: Showcasing New Investigating Women in Statistics Nouvelles chercheuses en statistique

Chair/Présidente: Juli Atherton (McGill University)

Organizer/Responsable: Laura Cowen (University of Victoria)

Session sponsored by the Committee on Women in Statistics

Séance parrainée par le Comité des femmes en statistique

Room/Salle: McCain - Scotiabank

Abstracts/Résumés

MIREILLE SCHNITZER (Université de Montréal), **RUSSELL STEELE**, **MICHÈLE BALLY & IAN SHRIER** (McGill University)

[Tuesday June 16 / mardi 16 juin, 13:30]

The Causal Inference Paradigm for Network Meta-analysis with Implications for Feasibility and Practice

Paradigme de l'inférence causale pour la méta-analyse en réseau et ses conséquences sur la faisabilité et la pratique

Network meta-analysis pools the results from randomized trials comparing different treatments. It has been noted that the choice of treatment comparison made at the study-design stage is often dependent on the chosen/available population to be recruited into each study. Therefore, pooling results over studies without adjustment might lead to confounded estimates of the effect of interest and lack of clarity as to the population on which the effects are defined. Drawing from counterfactual theory and the causal inference framework, we 1) define the population of interest in the network meta-analysis, 2) nonparametrically define the target parameter, 3) determine the requirements for identifiability of this parameter, and 4) propose modeling strategies to consistently estimate the target parameter.

La méta-analyse en réseau rassemble les résultats d'essais randomisés qui comparent différents traitements. Il semble que le choix de la comparaison de traitements effectué à l'étape de planification de l'étude dépend souvent de la population sélectionnée ou disponible à être recrutée dans chaque étude. Par conséquent, le regroupement des résultats d'études sans procéder à un ajustement peut mener à des estimations confondues de l'effet observé et à un manque de clarté quant à la population pour laquelle les effets sont définis. En nous inspirant de la théorie contre-factuelle et du cadre d'inférence causale, nous 1) définissons la population d'intérêt dans la méta-analyse en réseau, 2) définissons de façon non paramétrique le paramètre cible, 3) déterminons ses exigences d'identifiabilité, et 4) proposons des stratégies de modélisation pour l'estimer de façon convergente.

YILDIZ YILMAZ (Memorial University)

[Tuesday June 16 / mardi 16 juin, 14:00]

Genetic Association Studies Based on Joint Models of Multiple Phenotypes

Études d'association génétique basées sur des modèles conjoints de phénotypes multiples

Much of the heritable variation in complex traits is still unexplained. This may be due to low power to detect associations between genetic variants and traits under the association models considered. A low minor allele frequency (MAF) of causal ge-

La majorité de la variation héréditaire des traits complexes est toujours inexpliquée. Cela peut s'expliquer par une faible puissance de détection des associations entre les variantes génétiques et les traits à l'aide des modèles d'association utilisés. Cette faible puissance s'explique notamment par une

2D-I4: Showcasing New Investigating Women in Statistics Nouvelles chercheuses en statistique

netic variants is one of the reasons leading to low power. Hence, recent studies have focused on developing gene-based multimarker rare variant association tests to identify causal rare variants (i.e. variants with low MAF). More informative statistical models can be considered by incorporating the dependence structure between multiple phenotypes or multi-level biological measures. We will discuss the gain in power by using such joint models of phenotypes. A whole genome sequence data set will be used to illustrate the modeling approach.

faible fréquence des allèles mineurs des variantes génétiques causales. Les études récentes ont donc mis l'accent sur le développement de tests d'association de variantes rares à marqueurs multiples basés sur des gènes pour identifier les variantes rares causaux (c.-à-d. des variantes présentant une faible fréquence des allèles mineurs). Des modèles statistiques plus informatifs peuvent être pris en considération en incorporant la structure de dépendance entre les phénotypes multiples ou les mesures biologiques à niveaux multiples. Nous discutons du gain de puissance découlant de l'utilisation de ce type de modèles conjoints de phénotypes. Nous illustrons notre approche de modélisation à l'aide d'un jeu de données relatif à la séquence d'un génome complet.

LIANGLIANG WANG (Simon Fraser University), **ALEXANDRE BOUCHARD-CÔTÉ** (University of British Columbia), **ARNAUD DOUCET** (Oxford University)

[Tuesday June 16 / mardi 16 juin, 14:30]

Scalable Monte Carlo Methods for Bayesian Phylogenetic Inference

Méthodes de Monte-Carlo extensibles pour l'inférence bayésienne phylogénétique

Computational issues have increasingly limited large scale Bayesian Phylogenetic inference. We aim at developing methods for Phylogenetic inference that can complement existing Markov Chain Monte Carlo (MCMC) algorithms. Sequential Monte Carlo methods are an alternative to MCMC that have become very popular for time series models. Such methods have been recently developed to address phylogenetic inference problems but currently available techniques are only applicable to a very restricted class of phylogenetic tree models compared to MCMC. I will present scalable Monte Carlo Methods to approximate posterior phylogenetic tree distributions which is applicable to a general class of models.

Les problèmes informatiques limitent de plus en plus l'inférence bayésienne phylogénétique à grande échelle. Nous cherchons à développer des méthodes d'inférence phylogénétique qui peuvent compléter les algorithmes Monte-Carlo par chaînes de Markov (MCMC) existants. Les méthodes séquentielles de Monte-Carlo sont une solution de rechange aux MCMC qui sont devenues très populaires pour les modèles de séries chronologiques. Ces méthodes ont été récemment développées en réponse aux problèmes d'inférence phylogénétique, mais les techniques actuellement disponibles s'appliquent uniquement à une classe très restreinte de modèles d'arbres phylogénétiques, comparativement aux MCMC. Dans cet exposé, je présente des méthodes de Monte-Carlo extensibles pour approximer les distributions a posteriori d'arbres phylogénétiques applicables à une classe générale de modèles.

2D-I5: Survival Modelling with Environmental Applications Modélisation de la survie et applications environnementales

Chair/Président: John Braun (University of British Columbia)

Organizers/Responsables: John Braun (University of British Columbia) and/et Douglas Woolford (Wilfred Laurier University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: Rowe 1028

Abstracts/Résumés

JAMES STAFFORD (University of Toronto)

[Tuesday June 16 / mardi 16 juin, 13:30]

Accommodating Spatial and Temporal Uncertainty in Wildfire Ignitions.

Prise en compte de l'incertitude spatiale et temporelle dans le déclenchement des feux de forêts

Historical forest fire records can be temporally and spatially incomplete. Ignition locations may be viewed as both temporally censored (observed ignition time being a right censoring point) and areally censored within the observed fire perimeters. One approach to consider is a MCMC algorithm for making inference on a log-Gaussian Cox process with the exact times and locations being a latent variable sampled at each iteration. Alternatively, a local-EM algorithm could provide a non-parametric fit to the ignition intensity, where the fitted intensity surface maximizes the expectation of the likelihood function taken over the unknown ignition locations. Both approaches have analogies to disease mapping problems, where cases of a given health outcome are aggregated to administrative areas and reporting periods.

Les registres historiques des feux de forêts peuvent être incomplets dans le temps et l'espace. Les lieux de déclenchement peuvent être considérés comme censurés au niveau temporel (le moment du déclenchement observé étant un point de censure à droite) et géospatial à l'intérieur des périmètres de feux observés. L'une des approches examinées consiste à inférer un processus de Cox log-gaussien à l'aide d'un algorithme MCMC où les temps et les emplacements exacts représentent une variable latente échantillonnée à chaque itération. Par ailleurs, un algorithme EM local peut offrir un ajustement non paramétrique à l'intensité du déclenchement, où la surface ajustée de l'intensité maximise l'espérance de la fonction de vraisemblance par rapport aux lieux de déclenchements inconnus. Les deux méthodes présentent des analogies avec des problèmes de cartographie des maladies, où les cas pour un résultat de santé donné sont regroupés en zones administratives et en périodes de référence.

YI XIONG & JOAN HU (Simon Fraser University), **JOHN BRAUN** (University of British Columbia)

[Tuesday June 16 / mardi 16 juin, 14:00]

Exploring Forest Fire Duration Times

Exploration de la durée des feux de forêts

In attempt to understand how fire duration time is associated with environmental factors together with the fire's location and starting time, we analyze a collection of lightning-caused forest fire records.

Pour tenter de comprendre les liens entre la durée d'un feu et les facteurs environnementaux, l'emplacement du feu et son heure de déclenchement, nous analysons une série de registres de feux de forêts allumés par la foudre.

2D-I5: Survival Modelling with Environmental Applications Modélisation de la survie et applications environnementales

Two classes of semiparametric survival models are considered to avoid unrealistic model constraints and to capture the natural features of the underlying fire process. The analysis integrates the conventional survival analysis procedures with data smoothing techniques. In particular, we present how to deal with practical challenges arising from forest fire data, such as reporting delay, seasonal effect and spatio-temporal clustering. This is a joint work with John Braun (UBC-O) and Joan Hu (SFU).

Nous examinons deux classes de modèles de survie semi-paramétriques pour éviter les contraintes irréalistes d'un modèle et pour saisir les caractéristiques naturelles du processus sous-jacent de l'incendie. L'analyse intègre les procédures d'analyse de survie conventionnelles et des techniques de lissage de données. En particulier, nous présentons une façon de gérer les défis pratiques liés aux données de feux de forêts, comme les retards de déclaration, l'effet saisonnier et le regroupement spatio-temporel. Ces travaux ont été effectués conjointement avec John Braun (UBC-O) et Joan Hu (SFU).

THIERRY DUCHESNE (Université Laval), **JOHN BRAUN** (University of British Columbia), **PETER HALL** (University of Melbourne)

[Tuesday June 16 / mardi 16 juin, 14:30]

Kernel and Deconvolution Methods of Density Estimation for Interval-Censored Data

Méthodes du noyau et de déconvolution pour l'estimation d'une densité avec données censurées par intervalles

In this talk we will compare two nonparametric inference methods for the estimation of a density when the observations are interval-censored. One method is the kernel estimation procedure based on conditional expectations proposed by Braun, Duchesne and Stafford in 2005, while the other method is a new proposal based on deconvolution techniques. We will compare both methods theoretically and numerically under a specific model similar to case-II interval-censoring. The potential applicability of these two methods to forest fire modeling will be discussed.

Dans cet exposé nous comparerons deux méthodes non paramétriques pour l'estimation d'une densité lorsque les observations sont censurées par intervalles. Une méthode est la procédure d'estimation par noyau fondée sur des espérances conditionnelles proposée par Braun, Duchesne et Stafford en 2005, alors que l'autre méthode est une nouvelle proposition fondée sur des techniques de déconvolution. Nous comparerons les deux méthodes théoriquement et numériquement en supposant un certain modèle de censure par intervalle similaire au modèle de censure par intervalle de type II. L'applicabilité potentielle de ces deux méthodes à la modélisation de feux de forêt sera abordée.

2D-I6: Teaching Effective Communication of Statistical Ideas to Non-Statisticians Comment enseigner à bien communiquer les idées statistiques aux non-statisticiens

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

Session sponsored by the Statistics Education Section

Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

NATHAN TABACK (University of Toronto)

[Tuesday June 16 / mardi 16 juin, 13:30]

Teaching Effective Communication Skills to Statistics Students

Enseignement de techniques de communication efficaces pour des étudiants en statistique

Most statistical training is primarily focused on developing students' methodological toolbox, which includes a mix of statistical theory and applications. Unfortunately, communicating statistical concepts is typically something that is left to be 'picked up' later leaving students with an incomplete toolbox. In this talk I will discuss the importance of effective communication skills in the statistics profession, and teaching effective verbal and written communication skills to statistics students. These skills are not only aimed at helping students communicate effectively with both statisticians and non-statisticians, but also help students build their methodological toolbox. Vignettes will be given from a course that I teach at the University of Toronto on statistical consulting.

La plupart des formations en statistique sont principalement axées sur le développement des outils méthodologiques des étudiants, ce qui comprend un mélange de théorie statistique et d'applications. Malheureusement, la transmission des concepts statistiques est généralement laissée dans le but d'être «apprise» plus tard, laissant les étudiants avec une boîte à outils incomplète. Dans cet exposé, je traiterai de l'importance des compétences de communication efficace dans la profession statistique et de l'enseignement des compétences efficaces de communication verbale et écrite pour les étudiants en statistique. Ces compétences ont non seulement pour but d'aider les étudiants à communiquer efficacement avec les statisticiens et les non-statisticiens, mais aussi de les aider à construire leur boîte à outils méthodologiques. Des capsules d'un cours que j'enseigne à l'Université de Toronto sur la consultation statistique seront présentées.

LEHANA THABANE (McMaster University)

[Tuesday June 16 / mardi 16 juin, 14:00]

Effective Strategies to Communicate Statistical Issues to Clinicians in Health Research

Stratégies efficaces pour communiquer les enjeux statistiques aux cliniciens en recherche sur la santé

Teaching statistical concepts to non-statisticians can be rewarding and challenging at times. In this talk, I will share strategies for communicating basic statistical information to non-statistical audience. The coverage will include: how to interpret basic measures of effect in trials such as risk difference, relative risk, odds ratio and number-needed-to-treat; how to frame research questions; how to use tables to summarize analysis plans; and how

L'enseignement des concepts statistiques aux non-statisticiens peut parfois être enrichissant et stimulant. Dans cet exposé, je partagerai des stratégies pour communiquer de l'information statistique de base à un auditoire non statistique. La couverture englobera : comment interpréter des mesures de base de l'effet dans les essais telles que la différence de risque, le risque relatif, le rapport de cotes et le nombre nécessaire à traiter ; comment formuler des questions de recherche ; comment utiliser les tableaux pour résumer des

2D-I6: Teaching Effective Communication of Statistical Ideas to Non-Statisticians **Comment enseigner à bien communiquer les idées statistiques aux non-statisticiens**

to graphics to communicate complex statistical information.

plans d'analyse et comment communiquer l'information statistique complexe à l'aide de graphiques.

DON MCLEISH, DISCUSSANT, University of Waterloo
[Tuesday June 16 / mardi 16 juin, 14:30]

2D-C1: Bayes and Likelihood: Models and Algorithms Bayes et la vraisemblance : modèles et algorithmes

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

Room/Salle: McCain 2170

Abstracts/Résumés

MICHAEL NEWTON (University of Wisconsin, Madison)

[Tuesday June 16 / mardi 16 juin, 13:30]

Graph Partition Modeling

Modélisation de partitionnement des graphes

Information about potential dependencies among variables can often be expressed as an undirected graph -- not the familiar undirected graph that encodes conditional dependencies in a graphical model -- but rather as an encoding of covariate information or *potential* dependence relations. Examples from brain imaging and genomics help to express the definition. The transition from potential to actual dependence occurs through a partition of nodes of the graph, realized in such a way that blocks of the partition are connected subgraphs. I will discuss various models for such graph partitions, product-partition models for data, and Bayesian inference tools.

Les informations portant sur les dépendances potentielles parmi les variables peuvent souvent être exprimées par un graphique non-orienté -- pas le graphe non-orienté familier qui présente les dépendances conditionnelles dans un modèle graphique -- mais plutôt en présentant les informations covariantes ou les relations de dépendance *potential*. Des exemples d'imagerie du cerveau et de génomiques aident à exprimer cette définition. La transition de la dépendance de potentielle à actuelle passe par une partition de noeuds sur le graphe, réalisés de façon à ce que des blocs de partition soient des sous-graphes reliés. Je vais discuter de divers modèles pour de telles partitions graphiques, de modèles de partition de produit pour les données, ainsi que des outils d'inférence bayésienne.

BILJANA JONOSKA STOJKOVA & DAVID CAMPBELL (Simon Fraser University)

[Tuesday June 16 / mardi 16 juin, 13:45]

Simulated Tempering without Normalizing Constants

Atténuation simulée sans constantes normalisantes

When it comes to sampling from multi-modal distributions, it is well known that MCMC methods fail to explore efficiently the posterior surface. Simulated tempering is designed to deal with problems of multi-modality in target distributions by augmenting the posterior space with an auxiliary tempering parameter. In order to ensure the joint sampling of the temperature and model parameters, simulated tempering requires normalizing constants which contributes in large to its impossible implementation through traditional sampling means. The proposed simulated tempering algorithm avoids the need for computation of the normalizing constants.

En ce qui concerne l'échantillonnage de lois plurimodales, il est bien connu que les méthodes MCMC ne parviennent pas à explorer efficacement la surface a posteriori. L'atténuation simulée est conçue pour traiter les problèmes de plurimodalité dans les lois cibles en augmentant l'espace a posteriori avec un paramètre d'atténuation auxiliaire. Afin d'assurer l'échantillonnage conjoint des paramètres de température et de modèle, l'atténuation simulée nécessite des constantes de normalisation, ce qui en gros rend sa mise en œuvre impossible par des moyens d'échantillonnage traditionnels. L'algorithme proposé pour l'atténuation simulée évite la nécessité de calculer les constantes normalisantes.

2D-C1: Bayes and Likelihood: Models and Algorithms Bayes et la vraisemblance : modèles et algorithmes

DANIEL EVAN HOCKEY, SHIRLEY MILLS & SONG CAI (Carleton University)

[Tuesday June 16 / mardi 16 juin, 14:00]

Predicting the Attributes of Nodes in a Dynamic Network

Prévision des attributs des noeuds dans un réseau dynamique

Suppose we have a directed binary network. We assume we know the presence or absence of a link between all pairs of nodes. We divide the nodes in the network into two groups, V_1 and V_2 . In V_1 we know whether or not each node has a certain attribute, whereas in V_2 there is no information about the attribute. We create a method to predict if each node in V_2 has the attribute of interest. This is done by analyzing the correlation of nodes at different lags and by examining the distributions of the attribute based on link direction.

Soit un réseau binaire orienté : nous supposons que nous connaissons la présence ou l'absence d'un lien entre toutes les paires de noeuds. Nous séparons les noeuds du réseau en deux groupes, V_1 et V_2 . Dans V_1 , nous savons si chaque noeud possède ou non un certain attribut, alors que nous ne détenons aucun renseignement sur l'attribut dans V_2 . Nous proposons une méthode pour prévoir si chaque noeud du groupe V_2 possède l'attribut en question. Nous analysons la corrélation entre les noeuds à différents retards et examinons les distributions de l'attribut selon l'orientation du lien.

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

[Tuesday June 16 / mardi 16 juin, 14:15]

A Sequential Monte Carlo Sampler for Deterministically Related Continuous Random Variables

Échantillonneur de Monte-Carlo séquentiel pour des variables aléatoires continues liées de manière déterministe

Computing posterior distributions over variables linked by deterministic constraints is a recurrent problem in Bayesian analysis. Such problems can arise due to censoring, identifiability issues, or other considerations. It is well-known that standard implementations of Monte Carlo methods break down in the presence of these deterministic relationships. Although several Monte Carlo methods for deterministic relationships have been recently developed, few are applicable to deterministic constraints on continuous variables. In this talk, we propose a new Sequential Monte Carlo Sampler strategy for sampling continuous random variables under deterministic constraints.

Le calcul de distributions a posteriori pour des variables liées par des contraintes déterministes constitue un problème récurrent en analyse bayésienne. De tels problèmes peuvent être causés par de la censure, des difficultés d'identifiabilité ou d'autres raisons. Il est bien connu que la mise en œuvre standard de méthodes de Monte-Carlo achoppe en présence de ces relations déterministes. Bien que plusieurs méthodes de Monte-Carlo aient été développées récemment pour tenir compte de relations déterministes, peu sont applicables aux contraintes déterministes sur des variables continues. Dans cet exposé, nous proposons un nouvel échantillonneur de Monte-Carlo séquentiel pour échantillonner des variables aléatoires continues sous des contraintes déterministes.

MICHAEL REGIER (West Virginia University), **ERICA MOODIE** (McGill University)

[Tuesday June 16 / mardi 16 juin, 14:30]

The Orthogonally Partitioned EM Algorithm: Extending the EM Algorithm for Algorithmic Stability for Bias Correction due to Imperfect Data

L'algorithme EM orthogonalement partitionné : étendre l'algorithme EM pour la stabilité algorithmique dans la correction du biais attribué à des données imparfaites

We propose an extension of the EM algorithm that exploits the assumption of unique parameterization, corrects for imperfect data biases, converges

Nous proposons une extension de l'algorithme EM qui exploite l'hypothèse d'une paramétrisation unique, corrige les biais des données imparfaites, converge lorsque la mise en

2D-C1: Bayes and Likelihood: Models and Algorithms

Bayes et la vraisemblance : modèles et algorithmes

when standard implementation of the EM algorithm has a low convergence probability, and reduces a potentially complex algorithm into a sequence of smaller, simpler, self-contained EM algorithms. Both theoretical and finite sample results are considered. Evidence suggests that partitioning can provide better bias reduction in the estimation of model parameters. Breaking down a complex problem into simpler, more accessible steps will permit a broader implementation of the EM algorithm.

œuvre courante de l'algorithme EM a une faible probabilité de convergence et réduit un algorithme potentiellement complexe en une séquence d'algorithmes EM plus petits, plus simples, indépendants. Nous considérons les résultats théoriques et ceux d'échantillons de taille finie. La preuve suggère que le partitionnement peut fournir une meilleure réduction du biais dans l'estimation des paramètres du modèle. Briser un problème complexe en étapes plus simples, plus accessibles permettra une application étendue de l'algorithme EM.

2D-C2: Robust Methods Méthodes robustes

Chair/Présidente: Debbie Dupuis (HEC Montréal)

Room/Salle: McCain 2198

Abstracts/Résumés

PHILIPPE GAGNON (Université de Montréal), **ALAIN DESGAGNÉ** (Université du Québec à Montréal), **MYLÈNE BÉDARD** (Université de Montréal)

[Tuesday June 16 / mardi 16 juin, 13:30]

Robustness to Outliers in a Bayesian Simple Linear Regression Model

Inférence robuste sur les paramètres d'une régression linéaire simple bayésienne

In the field of statistics, and more precisely in a linear regression analysis context, data samples frequently contain outliers. Consequently, statistical inference can be contaminated, leading to results in disagreement with the majority of the observations. The least squares method in regression analysis is generally used but it can lead to inference which is not robust to outliers. This work deals with this problematic in a context of simple linear parametric Bayesian regression model. Indeed, theoretical results ensuring that the posterior inference is robust to outliers are described.

En statistique, et en particulier dans un cadre de régression linéaire, il arrive fréquemment qu'un échantillon de données contienne des valeurs aberrantes. Conséquemment, l'inférence statistique peut être contaminée par leur présence, menant ainsi à des résultats qui sont en désaccord avec la majorité des observations. La méthode de régression des moindres carrés est certes la plus populaire, mais elle peut mener à une inférence non robuste face à ces valeurs aberrantes. Cette présentation traite de régression linéaire simple paramétrique et bayésienne. Des résultats théoriques garantissant la robustesse de l'inférence *a posteriori* face à des valeurs aberrantes sont expliqués.

MARYAM SOHRABI & MAHMOUD ZAREPOUR (University of Ottawa)

[Tuesday June 16 / mardi 16 juin, 13:45]

Bootstrapping the Robust Estimates of the Mean Vector for Multivariate Heavy-Tailed Distributions with Different Indices of Stability

Bootstrap pour les estimations robustes du vecteur moyen pour des distributions de probabilité à ailes lourdes aléatoires avec différents indices de stabilité

We consider a robust estimation of the mean vector for a sequence of independent and identically distributed observations in the domain of attraction of a stable law with different indices of stability $(\alpha_1, \dots, \alpha_p)$ such that $1 < \alpha_i \leq 2, i = 1, \dots, p$. The suggested estimator is asymptotically normal depending on some unknown parameters. We apply an asymptotically valid bootstrap to construct confidence intervals and confidence regions for the

mean vector.

Key words: Stable process; M-estimation; Confidence interval; Bootstrap.

Nous considérons une estimation robuste du vecteur moyen pour une séquence d'observations avec distribution indépendante et identique dans le domaine d'attraction d'une loi stable avec différents indices de stabilité $(\alpha_1, \dots, \alpha_p)$ tels que $1 < \alpha_i \leq 2, i = 1, \dots, p$. L'estimateur suggéré est asymptotiquement normal selon des paramètres inconnus. Nous appliquons une méthode bootstrap asymptotiquement valide afin de bâtir les intervalles de confiance et les régions de confiance pour le vecteur moyen.

Mots clés : Processus stable ; M-estimateur ; intervalle de confiance ; bootstrap

2D-C2: Robust Methods Méthodes robustes

WEI TU & ROHANA KARUNAMUNI (University of Alberta)

[Tuesday June 16 / mardi 16 juin, 14:00]

Robust Efficient Generalized M-Estimation in Regression Models

M-Estimation généralisée robuste et efficace dans des modèles de régression

A new class of robust estimators for the linear regression model and logistic regression model were introduced. They are generalized M-estimators by absorbing a goodness of fit measure into a continuous weight function. Goodness-of-fit measure was computed using the empirical distribution of the residuals of an initial robust estimator in linear regression models, and squared Mahalanobis distances in logistic regression models. A Monte Carlo study showed that the proposed estimators operated at almost full efficiency while maintaining good robustness properties. The asymptotic consistency was proved using empirical process methods.

On a présenté une nouvelle classe d'estimateurs robustes pour le modèle de régression linéaire et le modèle de régression logistique. Ce sont des M-estimateurs généralisés par l'absorption d'une mesure de la qualité de l'ajustement dans une fonction de pondération continue. La mesure de la qualité de l'ajustement a été calculée au moyen de la distribution empirique des résidus d'un premier estimateur robuste dans des modèles de régression linéaire, et au moyen des distances de Mahalanobis dans des modèles de régression logistique. Une étude de Monte-Carlo a montré que les estimateurs proposés étaient presque totalement efficaces lorsqu'on maintenait les bonnes propriétés de robustesse. La convergence asymptotique a été prouvée en utilisant des méthodes de processus empiriques.

OLU AWOSOGA (University of Lethbridge), **JOSEPH MCKEAN & BRADLEY HUITEMA** (Western Michigan University)

[Tuesday June 16 / mardi 16 juin, 14:15]

Meta-type Analysis of Multiple Baseline Time Series Design Intervention Models for Dependent and Independent Series

Méta-analyse de modèles d'interventions en séries chronologiques avec plusieurs situations initiales pour des séries dépendantes et indépendantes

A traditional meta-type analysis for multiple baseline series was developed in this study using robust methodology. The design matrices provided for two-phase (AB) design allow for change in level and change in slope between each phase and the subsequent phase. The robust procedures are similar to the parametric procedures except another norm in place of Euclidean norm is used. A Monte-Carlo study of the methods is provided. Validity of the procedures and power comparisons between the parametric and robust methods was investigated and the results are presented. Diagnostic procedures for the analysis of these data are also developed.

Une méta-analyse traditionnelle pour les séries chronologiques à plusieurs situations initiales a été développée pour cette étude à l'aide de méthodologie robuste. Les matrices d'expérience pour les plans à deux phases (AB) permettent un changement de niveau et de pente entre chaque phase et la phase subséquente. Les procédures robustes sont similaires aux procédures paramétriques, sauf que la norme utilisée n'est pas euclidienne. Une étude de Monte-Carlo de la méthode est présentée. La validité de la procédure est explorée, ainsi que des comparaisons entre la puissance des méthodes paramétriques et robustes. Ces résultats sont présentés et des procédures diagnostiques pour l'analyse de ces données sont aussi développées.

2D-C2: Robust Methods Méthodes robustes

YANG ZHAO (University of Regina)

[Tuesday June 16 / mardi 16 juin, 14:30]

Robust Imputation to Unified Approach for Regression Models with Data Missing by Design

Imputation robuste à l'approche unifiée pour les modèles de régression avec données manquantes selon le modèle

Multiple imputations are widely used in statistical analysis with missing data. However, valid inference based on imputed full data depends on the "correctness" of the imputation models for the missing data especially when the missing percentage is high. In this article we describe a robust imputation to unified approach for regression models with data missing under two-stage studies, where parameter estimation and inference are valid even when the imputation models are not correct. Our simulation studies in various settings indicate that the performance of the proposed method is acceptable for practical implementation.

L'imputation multiple est largement utilisée en analyse statistique des données manquantes. Cependant, l'inférence valable fondée sur les données complètes imputées dépend de la « justesse » des modèles d'imputation pour les données manquantes, en particulier lorsque le pourcentage manquant est élevé. Dans cet article, nous décrivons une imputation robuste à l'approche unifiée pour les modèles de régression avec des données manquantes dans les études à deux degrés, où l'estimation des paramètres et l'inférence sont valides même lorsque les modèles d'imputation sont erronés. Nos études de simulation dans différents contextes montrent que la performance de la méthode proposée est acceptable pour une application pratique.

2D-C3: Theoretical Developments Développements théoriques

Chair/Président: Deli Li (Lakehead University)

Room/Salle: McCain 2184

Abstracts/Résumés

SARA DEVRIES & MOHAMED AMEZZIANE (Central Michigan University), **S. EJAZ AHMED** (Brock University)
[Tuesday June 16 / mardi 16 juin, 13:30]

On Shrinking Nonparametric Regression Models to Linear Regression Models

À propos du rétrécissement de modèles de régression non paramétriques à des modèles de régression linéaire

Several nonparametric estimators are obtained by converting function estimation to a normal means problem and then shrinking the parameters to zero. Shrinkage estimators are also used to incorporate prior information about model parameters in order to improve the performance of sample-based estimators. We combine these two shrinkage approaches to develop semiparametric regression models in the guise of hard-threshold (pre-test), soft-threshold (positive part) and smooth (Stein-James) regression estimators with multiple covariates. The new estimators are assessed by comparing their risk functions to those of their nonparametric competitors. Real data and simulation studies are used to illustrate the theoretical results.

De nombreux estimateurs non paramétriques sont obtenus par la conversion de l'estimation de fonctions à un problème de moyennes normales, suivie du rétrécissement des paramètres à zéro. Les estimateurs de rétrécissement permettent aussi d'intégrer de l'information a priori à propos des paramètres des modèles afin d'améliorer la performance des estimateurs basés sur les échantillons. Nous combinons ces deux approches de rétrécissement pour élaborer des modèles de régression semi-paramétriques sous la forme d'estimateurs de régression lisses (Stein-James), à seuil fort (pré-test) et à seuil faible (portion positive) avec des covariables multiples. Les nouveaux estimateurs sont évalués en comparant leurs fonctions de risque à celles de leurs concurrents non paramétriques. Nous illustrons nos résultats théoriques à l'aide d'études basées sur des données réelles et simulées.

FUQI CHEN (The Fields Institute for Research in Mathematical Sciences), **SÉVÉRIEN NKURUNZIZA** (Université de Windsor / Université de Sherbrooke)

[Tuesday June 16 / mardi 16 juin, 13:45]

On Stein-Rules in Multivariate Regression with Change-Points

Sur les règles de Stein en régression multivariée avec des points de rupture

We present an estimation problem of the matrix of the regression coefficients in context of unknown change-points, when the matrix satisfies an uncertain restriction. Generally, when the restriction holds, the RE dominates the UE. However, the RE may perform poorly when the restriction is violated. We propose a methodology to combine the sample and non-sample information in an optimal way. Namely, we derive a class of estimators which includes shrinkage estimators (SEs), UE and RE. We also generalise the condition for the SEs to domi-

Nous considérons le problème d'estimation d'une matrice des coefficients de régression avec plusieurs points de rupture, lorsque celle-ci satisfait une restriction. Sous la restriction, l'estimateur avec restriction (EAR) domine l'estimateur sans restriction (ESR). Néanmoins, l'EAR performe mal si la restriction est fautive. Nous proposons une méthode qui permet de combiner optimalement l'information échantillonnale et la restriction hypothétique. À propos, nous établissons une classe d'estimateurs qui comprend des estimateurs à rétrécissement (ER), EAR et ESR. Également, nous généralisons la condition pour que les ER dominent l'ESR

2D-C3: Theoretical Developments Développements théoriques

nate the UE and, we extend some results underlying the multidimensional version of Mixingale Central Limit Theorem.

ainsi que certains résultats sous-jacents à la version multidimensionnelle du théorème central limite pour les mixingales.

MIKLOS CSORGO (Carleton University), **YULIYA MARTSYNYUK** (University of Manitoba)

[Tuesday June 16 / mardi 16 juin, 14:00]

Testing for Change in the Mean via Weighted Partial Sums Processes

Vérification du changement dans les processus de sommes partielles moyens et pondérés

Let $\{X_i\}_{i \geq 1}$ be independent and identically distributed random variables that have a positive variance σ and finite absolute moment of order p , for some $p > 2$. Define a sequence of partial sums processes $\{Z_n(t), 0 \leq t \leq 1\}_{n \geq 1}$, where

$$Z_n(t) := \begin{cases} \left(\sum_{i=1}^{\lfloor (n+1)t \rfloor} X_i - \frac{\lfloor (n+1)t \rfloor}{n} \sum_{i=1}^n X_i \right) / (\sigma\sqrt{n}), & 0 \leq t < 1, \\ 0, & t = 1. \end{cases}$$

We prove limit theorems for weighted versions of $\{Z_n(t), 0 \leq t \leq 1\}_{n \geq 1}$ and demonstrate how these theorems can be made use of in nonparametric change-point analysis when studying the problem of testing for at most one change in the mean of independent observations.

Soient $\{X_i\}_{i \geq 1}$ des variables aléatoires indépendantes et identiquement distribuées avec une variance positive de σ et un moment absolu fini d'ordre p , pour $p > 2$. Définissons une séquence de processus de sommes partielles $\{Z_n(t), 0 \leq t \leq 1\}_{n \geq 1}$, où

$$Z_n(t) := \begin{cases} \left(\sum_{i=1}^{\lfloor (n+1)t \rfloor} X_i - \frac{\lfloor (n+1)t \rfloor}{n} \sum_{i=1}^n X_i \right) / (\sigma\sqrt{n}), & 0 \leq t < 1, \\ 0, & t = 1. \end{cases}$$

Nous prouvons les théorèmes limites pour les versions pondérées de $\{Z_n(t), 0 \leq t \leq 1\}_{n \geq 1}$ et montrons comment ces théorèmes peuvent être utilisés dans l'analyse d'identification des changements non paramétriques, pour établir qu'un changement au maximum est intervenu dans la moyenne des observations indépendantes.

CHI DONG & MICHAEL KOURITZIN (University of Alberta)

[Tuesday June 16 / mardi 16 juin, 14:15]

Replication and Approximation of Stochastic Processes in Generalized Spaces

Réplication et approximation de processus stochastiques dans des espaces généralisés

A new technique is developed for replicating stochastic processes living in a quite general topological space. The replica is a quasi-modification of the original process in another compact metric space with similar evolutions. Conditions for the existence of general and stationary solutions are derived for the replica evolution system and then transformed back to the original system under mild conditions. By this approach, the long-term typical behavior of countable approximate solutions is shown to converge weakly to a stationary solution. As an application, weak convergence to a stationary filter is established for the long-term typical behavior of countable approximate filters.

Nous proposons une nouvelle technique de réplication des processus stochastiques définis dans un espace topologique assez général. La réplique est une quasi-modification du processus d'origine dans un autre espace métrique compact présentant des évolutions semblables. Nous obtenons les conditions d'existence de solutions générales et stationnaires pour le système d'évolutions de la réplique et nous les transformons vers le système original sous des hypothèses faibles. Selon cette approche, le comportement typique à long terme des solutions approximatives dénombrables converge faiblement vers une solution stationnaire. Comme application, nous établissons la convergence faible d'un filtre stationnaire pour le comportement typique à long terme des filtres approximatifs dénombrables.

2D-C3: Theoretical Developments Développements théoriques

AMMAR SARHAN (Dalhousie University), **JOSEPH APALOO** (St. Francis Xavier University)

[Tuesday June 16 / mardi 16 juin, 14:30]

The Bivariate Two Parameter Bathtub Hazard Rate Distribution

La loi bivariée avec taux de défaillance en baignoire à deux paramètres

A bivariate two parameter bathtub hazard rate (BTPBT for short) distribution is introduced in this paper. The BTPBT distribution is of Marshall-Olkin type and it is a singular distribution. The marginal distributions of the BTPBT are univariate - two parameter bathtub distributions. Some statistical properties of this distribution are discussed. The BTPBT distribution has four unknown parameters. The maximum likelihood (ML) and Bayes techniques are applied to make inferences for the model parameters. Both techniques cannot provide closed forms. We use numerical techniques for ML and MCMC for the Bayes. Real data are analyzed for illustrative and comparison purposes.

Une loi bivariée avec taux de défaillance en forme de baignoire à deux paramètres (BTPBT) est présentée dans cet article. La loi BTPBT est de type Marshall-Olkin et est une loi singulière. Les lois marginales de la BTPBT sont des lois à une variable avec taux de défaillance en baignoire à deux paramètres. Quelques propriétés statistiques de cette loi sont abordées. La loi BTPBT a quatre paramètres inconnus. L'estimation par maximum de vraisemblance (ML) et des techniques bayésiennes sont utilisées à des fins d'inférence pour les paramètres du modèle. Les deux techniques ne peuvent pas fournir de formules explicites. Nous utilisons des techniques numériques pour ML et MCMC pour l'approche bayésienne. Des données réelles sont analysées pour fins d'illustration et de comparaison.

2E-I1: Extreme Value Theory and Applications Théorie des valeurs extrêmes et applications

Organizer and Chair / Responsable et président: Hao Yu (Western University)

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

FRANCIS ZWIERS (University of Victoria)

[Tuesday June 16 / mardi 16 juin, 15:30]

Detection, Attribution and Extreme Climate and Weather Events

Détection, attribution et événements climatiques et météorologiques extrêmes

Climate research is increasingly concerning itself with questions concerning long-term changes in the frequency and intensity of extreme climate and weather events. While methods appropriate to extremes are not completely settled, results on temperature and precipitation extremes consistently point to a human contribution to observed changes in their characteristics. There is also growing interest in "event attribution", which is the identification of external factors that contributed to the intensity or likelihood of individual events. I will compare and contrast the questions posed by traditional detection and attribution research (what caused observed long-term changes in extremes) and event attribution (what caused the event that has just occurred), and the methods that are used to answer these questions.

La recherche climatique porte de plus en plus sur l'évolution à long terme de la fréquence et de l'intensité des événements climatiques et météorologiques extrêmes. Les méthodes applicables aux extrêmes ne sont pas encore établies une fois pour toute, mais tous les résultats concernant les extrêmes de température et de précipitation indiquent que l'homme contribue aux changements observés dans leurs caractéristiques. Les chercheurs s'intéressent également de plus en plus à « l'attribution des événements », à savoir à identifier les facteurs externes qui contribuent à l'intensité ou à la probabilité d'événements individuels. Je compare et j'oppose les questions posées par la recherche traditionnelle sur la détection et l'attribution (quelles sont les causes des changements à long terme des extrêmes) et par l'attribution des événements (quelles sont les causes de l'événement qui vient de se produire), ainsi que les méthodes utilisées pour y répondre.

DEBBIE DUPUIS (HEC Montréal)

[Tuesday June 16 / mardi 16 juin, 16:00]

The Dynamics of Extreme Daily Temperatures

La dynamique des températures journalières extrêmes

The Southwestern U.S. has experienced some of the most important increases in night-time minimum temperatures over the last 60 years and climate models are projecting more of the same to the end of the century. As climate, geography and population density vary considerably over the area, very diverse extreme temperature levels and dynamics are observed. We study series from 12 lo-

Le sud-ouest des États-Unis a connu d'importantes augmentations des températures minimales nocturnes au cours des 60 dernières années et les modèles climatiques projettent de même jusqu'à l'an 2100. Comme le climat, la géographie et la densité de la population varient considérablement sur l'ensemble du territoire, des niveaux de température extrêmes et des dynamiques très diversifiés sont observés. Nous étudions les données de 12 emplacements sur la période 1950-

2E-I1: Extreme Value Theory and Applications Théorie des valeurs extrêmes et applications

cations over the 1950-2009 period and show how these minima exhibit more complex dynamics than daytime maximum temperatures. We present one model capable of capturing all the features of the data. Changes at the 12 locations vary considerably: many locations have seen large changes while two locations have experienced little or no change.

2009 et montrons comment ces minima présentent des dynamiques plus complexes que leurs maxima homologues. Nous présentons un modèle qui capte toutes les caractéristiques des données. Les changements aux 12 emplacements varient considérablement : de nombreux endroits ont connu de grands changements tandis que deux endroits ont connu peu ou pas de changement.

EDIT GOMBAY (University of Alberta), **FUXIAO LI** (Northwest Polytechnical University, Xi'an, P.R. China), **HAO YU** (University of Western Ontario)

[Tuesday June 16 / mardi 16 juin, 16:30]

Change Detection in Categorical Time Series

Détection des changements dans les séries temporelles catégorielles

We propose methods for detecting change in the coefficients of a multinomial logistic regression model for categorical time series retrospectively, that is off-line. The alternatives to the null hypothesis of a stationary distribution can be either the hypothesis that the null hypothesis is not true, or that there is a temporary change. The statistics are based on the efficient score vector of the partial likelihood function. Its components correspond to the different components of the multinomial logistic regression model's parameter vector. One can test for all or for any subset while treating the rest as nuisance parameters. Our motivating example is a DNA sequence of four categories.

Nous proposons des méthodes permettant de détecter rétrospectivement, soit hors ligne, tout changement dans les coefficients d'un modèle de régression logistique multinomial pour séries temporelles catégorielles. Les hypothèses alternatives à l'hypothèse nulle d'une distribution stationnaire sont soit que l'hypothèse nulle est fautive, soit qu'un changement temporaire est intervenu. Les statistiques sont basées sur le vecteur de score efficace de la fonction de vraisemblance partielle. Ses composantes correspondent à celles du vecteur des paramètres du modèle de régression logistique multinomiale. On peut vérifier l'ensemble des paramètres ou certains seulement, en traitant les autres comme paramètres perturbateurs. Notre exemple motivant est une séquence d'ADN de quatre catégories.

2E-I2: Interviews, Grants, and Family Balance for New Investigators: A Panel Discussion Entrevues, subventions et équilibre familial pour les nouveaux chercheurs (une table ronde)

Organizer and Chair / Responsable et présidente: Mélina Mailhot (Concordia University)

Session sponsored by the New Investigator Committee

Séance parrainée par le Comité des nouveaux chercheurs

Room/Salle: McCain - Scotiabank

Description

In this panel session we will discuss key factors for new researchers. First, we will go through the interview process for early career researchers. Tips on items to emphasize on your CV and research will be provided, from the employer's perspective. There will also be a discussion held around specific items of the NSERC proposal grading, as an early career researcher, such as how to prepare the training plan for graduate students. We will also discuss the stress related to being a successful researcher, teacher, and family balance. In this session, three academics in different stages of their career will share their experience and advice regarding the first years as researchers. Participants in attendance will be invited to ask questions to the panellists and participate in the discussion.

Dans cette table ronde, on va discuter des facteurs clés pour les nouveaux chercheurs. On va d'abord parler du processus d'entrevue pour les chercheurs en début de carrière. Des conseils sur quels éléments doivent être mis en valeur dans votre CV et vos recherches seront donnés selon le point de vue de l'employeur. Il y aura aussi une discussion à propos d'éléments spécifiques de l'évaluation des demandes de subventions au CRSNG pour les nouveaux chercheurs, tel que comment préparer le plan de formation d'étudiants de cycles supérieurs. On discutera également du stress lié au fait d'être un chercheur et enseignant accompli et à l'équilibre familial. Dans cette séance, trois universitaires à différentes étapes de leur carrière vont partager leur expérience et leurs conseils à propos de leurs premières années comme chercheurs. Les participants à la séance seront invités à poser des questions aux panélistes et à se joindre à la discussion.

2E-I3: Recent Developments in Statistical Genomic Modelling and Analysis Développements récents en modélisation et analyse génomiques statistiques

Organizer and Chair / Responsable et présidente: Jinko Graham (Simon Fraser University)

Session sponsored by the Biostatistics Section

Séance parrainée par le Groupe de biostatistique

Room/Salle: Rowe 1020

Abstracts/Résumés

CELIA GREENWOOD, AURÉLIE LABBE, ANTONIO CIAMPI & DAVID STEPHENS (McGill University)

[Tuesday June 16 / mardi 16 juin, 15:30]

Methods for Analysis of DNA Methylation

Méthodes d'analyse pour la méthylation de l'ADN

Methylation of DNA can alter DNA activity, by blocking access to the DNA at methylated sites. For example, DNA methylation determines cell-type-specificity. Recent interest in studying DNA methylation arises from the awareness that methylation patterns can change in response to age, disease or environmental exposures. Since DNA methylation can be measured by high-throughput experimental methods, analysts must cope with large data sets, undesired technical biases, and long-range correlations. The presentation will cover many of these issues, and focus on some proposed methods for studying local patterns in methylation, including dimension reduction methods and functional data analysis.

La méthylation contrôle l'activité de l'ADN en bloquant l'accès aux endroits méthylés dans le génome. Ce phénomène, spécifique à chaque type de cellule, est un des déterminants de la spécificité cellulaire. Les profils de méthylation peuvent varier en fonction de l'âge, de la maladie, et d'autres facteurs environnementaux. Puisque la méthylation est mesurée à l'aide de méthodes expérimentales à haut débit, l'analyste est confronté aux problèmes associés avec les jeux de données de grande taille, ainsi qu'à des biais techniques, et à des corrélations de longue portée. La présentation discutera quelques approches analytiques propres à ce genre de données, se concentrant, en particulier, sur l'étude de formes locales de méthylation, incluant des méthodes de réduction de dimension et d'analyse fonctionnelle.

GABRIELA COHEN FREUE (University of British Columbia)

[Tuesday June 16 / mardi 16 juin, 16:00]

Borrowing Information from Genomics Data to Boost Proteomics Discoveries

Augmentation des découvertes en protéomique grâce à des données de génomique

To date, there is an unmet clinical need to identify molecular indicators (e.g., proteins) of various diseases, including cancer, heart failure, and chronic obstructive pulmonary disease, among others. In this talk, I will focus on the problem of measurement errors in protein quantitations, which may affect the identification of protein biomarkers in a discovery study. As protein levels are regulated in part by gene expression, related genomic data can be integrated to address this problem through the

Il existe à ce jour un besoin clinique non comblé pour identifier les indicateurs moléculaires (p. ex. des protéines) de diverses maladies, y compris le cancer, l'insuffisance cardiaque et la maladie pulmonaire obstructive chronique. Cet exposé porte sur le problème des erreurs de mesures lors de l'analyse quantitative des protéines, ce qui peut affecter l'identification des biomarqueurs protéiques dans le cadre d'une étude. Comme les niveaux de protéines sont partiellement régulés par l'expression génique, des données connexes en génomique peuvent être utilisées pour trai-

2E-I3: Recent Developments in Statistical Genomic Modelling and Analysis Développements récents en modélisation et analyse génomiques statistiques

implementation of instrumental variables estimators. The proposed methodology exploits the plausible mechanisms from existing biological knowledge that relate genes, proteins, and diseases and takes advantage of this knowledge to increase the signal strength of sometimes weak, but biologically relevant --omics signatures.

ter ce problème par la spécification d'estimateurs par variables instrumentales. La méthodologie proposée exploite les mécanismes plausibles tirés de connaissances biologiques existantes qui associent les gènes, les protéines et les maladies, et tire profit de ces connaissances pour augmenter la puissance du signal de signatures protéomiques et génomiques parfois faibles mais pertinentes du point de vue biologique.

LAURENT BRIOLLAIS (University of Toronto)

[Tuesday June 16 / mardi 16 juin, 16:30]

Bayesian Graphical Models for Gene Network Analysis in Large-Scale Problems

Modèles graphiques bayésiens pour l'analyse de réseaux de gènes dans des problèmes à grande échelle

The detection of genetic interactions and inference about gene network topology can be a very valuable approach to understanding the joint basis of complex disease etiology. Network information is, unfortunately, extremely difficult to apprehend and only partially available. Bayesian Graphical models (BGMs) provide a probabilistic framework for making inference and representing our knowledge about these complex structured data. We discuss here the use of BGMs for gene network analysis in high-dimensional problems and how expert prior information and efficient algorithms can help inferring these complex networks. Applications to ongoing genetic and genomic problems in cancer research will be presented.

La détection des interactions génétiques et l'inférence de la topologie de réseaux de gènes peut être une approche très précieuse pour comprendre la base conjointe de l'étiologie de maladies complexes. L'information sur les réseaux de gènes est, malheureusement, très difficile à appréhender et seulement partiellement disponible. Les modèles graphiques bayésiens (MGB) fournissent un cadre probabiliste pour l'inférence et représentent nos connaissances sur ces données structurées complexes. Nous discutons ici l'utilisation des MGB pour l'analyse des réseaux de gènes dans des problèmes à grande dimension et comment l'information experte a priori et des algorithmes efficaces peuvent aider à déduire ces réseaux complexes. Des applications actuelles sur des problèmes génétiques et génomiques dans la recherche sur le cancer seront présentées.

2E-I4: Statistical Approaches for Big Data Approches statistiques des données volumineuses

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

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: Rowe 1028

Abstracts/Résumés

JEFFREY ANDREWS (MacEwan University), **PAUL MCNICHOLAS** (McMaster University), **MATHIEU CHALIFOUR** (MacEwan University)

[Tuesday June 16 / mardi 16 juin, 15:30]

Robust Model-Based Clustering of High-Dimensional Data

Regroupement à l'aide de modèles robustes de données à grandes dimensions

In general, mixtures of factor analyzers provide a sound statistical basis for performing cluster analysis on high-dimensional (particularly with small n) data while avoiding the often dubious assumption of independence among variables. Mixtures of multivariate t -factor analyzers (MM t FA) and modified t -factor analyzers (MOD t) provide a measure of robustness via the assumption of t -distributed latent and manifest variables. The beta version of R software for this family of models will be introduced and illustrated on real data. Issues regarding parameter estimation, initialization, and computation time will be discussed.

En général, les analyseurs de mélanges de facteurs fournissent une base statistique admissible pour effectuer des analyses de groupement sur des données de grandes dimensions (particulièrement avec des petits n) tout en évitant l'hypothèse souvent douteuse d'indépendance entre les variables. Des analyseurs de mélanges de facteurs t multivariés (MM t FA) des analyseurs de facteurs t modifiés (MOD t) fournissent une mesure de robustesse par la supposition de variables manifestes et latentes distribuées selon une loi t . La version bêta du logiciel R pour cette famille de modèles sera présentée et illustrée avec des données réelles. Les enjeux concernant l'estimation, l'initialisation des paramètres et le temps de calcul seront abordés.

HONG GU, MICHAEL BUTLER, SHEN LING, TOBY KENNEY & CHRIS FIELD (Dalhousie University)

[Tuesday June 16 / mardi 16 juin, 16:00]

Machine Learning Diagnoses on Patients Presenting Abdominal Pain

Diagnostics d'apprentissage automatique sur des patients présentant des douleurs abdominales

Currently a massive amount of data has been collected in the electronic database of the emergency department across Canada, which is not yet well utilized. We aim to build an automated diagnosis tool for physicians. It will provide a list of most possible diagnoses for physicians to check for any possibilities overlooked. It can also assess the most informative additional tests, to reduce the time and resources used on unnecessary tests. This is not a standard classification problem since different test results are available for different patients. We pro-

Actuellement, un volume colossal de données, qui n'est pas encore bien utilisé, a été recueilli dans la base de données des services des urgences au Canada. Notre objectif est de créer un outil de diagnostic automatique qui fournira aux médecins une liste des diagnostics les plus probables et qui leur permettra de vérifier toutes les possibilités, sans en omettre. Cet outil peut également analyser les tests supplémentaires les plus informatifs, pour ne pas perdre de temps et de ressources sur des tests inutiles. Ceci n'est pas un problème de classification standard puisque les différents résultats de tests sont disponibles pour différents patients. Nous propo-

2E-I4: Statistical Approaches for Big Data Approches statistiques des données volumineuses

pose a model framework that can combine different models for groups of patients with different variables. The same technique can be used to combine data from different hospitals into one national system.

sons un cadre qui peut combiner différents modèles pour des groupes de patients avec différentes variables. Cette technique peut être utilisée pour combiner des données de différents hôpitaux en un seul système à l'échelle du Canada.

SUNNY WANG & ZHOUQIN HE (St. Francis Xavier University)

[Tuesday June 16 / mardi 16 juin, 16:30]

Business Failure Prediction for Canadian Charitable Organizations Using Data Mining Tools

Prévision de faillite pour les organismes de bienfaisance canadiens par l'utilisation d'outils d'exploration de données

Business failure prediction has been a classic problem. A wide variety of data mining techniques have become the new favorites of researchers for their outstanding prediction abilities. While majority of the researches focused on profitable corporations, few attempted to study the charitable sector due to the lack of available data of this sector. This study attempts to apply data mining tools to perform a business failure prediction task aiming at Canadian charitable organizations. In order to use the panel data containing multiple explanatory variables recorded for thousands of charitable organizations during a three-year period, a novel approach to handle time series attributes using Decision Tree and Random Forest by measuring the Dynamic Time Warping (DTW) dissimilarities between observations is presented.

La prévision de faillite d'entreprise est un problème classique. De nombreuses techniques d'exploration de données sont devenues les nouvelles techniques préférées des chercheurs en raison de leur capacité de prévision exceptionnelle. Bien que la plupart des recherches mettent l'accent sur le domaine des entreprises, peu de chercheurs ont tenté d'étudier le secteur des organismes de bienfaisance en raison du manque de données disponibles pour celui-ci. Cette étude tente d'appliquer des outils d'exploration de données pour effectuer une prévision de faillite visant les organismes de bienfaisance canadiens. Afin d'utiliser les données de panel contenant de multiples variables indépendantes enregistrées pour des milliers d'organismes de bienfaisance sur une période de trois ans, nous présentons une nouvelle approche permettant de traiter des attributs de séries chronologiques au moyen d'arbres décisionnels et de forêts aléatoires par la mesure des dissemblances provenant de la comparaison dynamique entre les observations.

2E-I5: eLearning in Statistics -- Successes and Opportunities Apprentissage électronique en statistique -- succès et opportunités

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

Organizer/Responsable: Bethany White (Western University)

Room/Salle: Rowe 1011

Abstracts/Résumés

IVO DINOV, ALEXANDR KALININ & SYED HUSAIN (University of Michigan), **NICOLAS CHRISTOU** (UCLA)

[Tuesday June 16 / mardi 16 juin, 15:30]

Statistics Online Computational Resource (SOCR) - An Infrastructure for Technology-enhanced Trans-disciplinary Education

Statistics Online Computational Resource (SOCR) - Une infrastructure pour l'éducation transdisciplinaire améliorée par la technologie

Three of the main challenges in modern science education revolve around improving student motivation, engaging students via hands-on activities, and blending IT, conceptual learning and pedagogical techniques into the modern curriculum. SOCR provides web-based open-access to a suite of data, visualization, statistical, analytical, modeling tools (Java-Applets and HTML5-webapps), as well as learning materials and instructional resources. All SOCR resources (<http://SOCR.umich.edu>) are open-source, collaboratively developed, multilingual and embrace trans-disciplinary interactions via Cloud services. In this interactive hands-on presentation we will showcase examples of probability calculators, statistics analysis tools, data modeling, visualization, virtual games, simulations and experiments. Attendees are encouraged to bring their laptops (modern Java/JavaScript-enabled browsers and Internet connection are required) to explore and try out different tools during the presentation.

Trois des principaux défis dans l'enseignement de la science moderne s'articulent autour de l'amélioration de la motivation des élèves, de la participation des élèves par le biais d'activités pratiques et du mélange de l'informatique, de l'apprentissage conceptuel et des techniques pédagogiques dans le programme moderne. Le SOCR fournit un accès ouvert sur le Web à un ensemble de données, de visualisations, d'outils statistiques, analytiques, de modélisation (applets Java et applications Web HTML5), ainsi que des matériels d'apprentissage et des ressources pédagogiques. Toutes les ressources SOCR (<http://SOCR.umich.edu>) sont en code source ouvert, ont été développés en collaboration, sont multilingues et couvrent des interactions transdisciplinaires via les services informatiques en nuages. Dans cette présentation interactive et pratique, nous allons présenter des exemples de calculateurs de probabilité, des outils d'analyses statistiques, de la modélisation de données, de la visualisation, des jeux virtuels, des simulations et des expériences. Les participants sont encouragés à apporter leurs ordinateurs portables (navigateurs Java modernes / JavaScript améliorés et connexion Internet sont nécessaires) pour explorer et essayer différents outils lors de la présentation.

ALISON GIBBS (University of Toronto)

[Tuesday June 16 / mardi 16 juin, 16:00]

A New Blend: An Introductory Statistics Course Integrating a MOOC and Active Learning Sessions, with Seasoning from the Students' Areas of Study

Un nouveau mélange : un cours d'introduction à la statistique intégrant un MOOC et des sessions d'apprentissage actif pimentées selon les domaines d'étude des étudiants

2E-I5: eLearning in Statistics -- Successes and Opportunities Apprentissage électronique en statistique -- succès et opportunités

In an inverted classroom, students' first exposure to new material happens outside of class, freeing class time for active learning. We will report on our experiences using resources originally developed for a MOOC as the foundation for an inverted classroom approach to an introductory statistics course. Traditional lectures were replaced with videos, with frequent feedback embedded. Classroom time was filled with problem solving, case studies, and discussion of conceptual issues. Some course sections had particular disciplinary focuses, allowing students to develop an understanding of the importance of statistical methods within their area. We will describe the motivation and structure of the course, some lessons learned from our experience and from student feedback and surveys, and our future plans.

Dans une salle de classe inversée, le premier contact des élèves avec du nouveau matériel a lieu en dehors de la classe, ce qui libère du temps en classe pour l'apprentissage actif. Nous rendrons compte de nos expériences où l'on utilise les ressources développées à l'origine pour un MOOC comme base d'une approche de classe inversée dans un cours d'introduction à la statistique. Les présentations traditionnelles ont été remplacées par des vidéos intégrant une rétroaction fréquente. Le temps en classe a été comblé par la résolution de problèmes, des études de cas et la discussion de questions conceptuelles. Certaines sections du cours mettaient l'emphase sur des domaines spécifiques, permettant aux étudiants de développer une compréhension de l'importance des méthodes statistiques dans leur domaine. Nous décrivons la motivation et la structure du cours, quelques leçons tirées de notre expérience et de la rétroaction des étudiants et des enquêtes, et nos plans futurs.

BETHANY WHITE (Western University)

[Tuesday June 16 / mardi 16 juin, 16:30]

Going Online in a Big Way: Developing a First-year Statistics Course for the Online Ontario Initiative

Passer en ligne de façon importante : développer un cours de statistique de première année pour l'Initiative en ligne de l'Ontario

Early last year, STAT 1024, an introductory course for non-majors at The University of Western Ontario, was selected for the Ontario Online Initiative by the Ministry of Training, Colleges and Universities (MTCU). Through this initiative, the province is assembling a suite of online courses offered by Ontario universities for students across the province to take and obtain equivalent transfer credit at their home institutions. The STAT 1024 redesign was informed by evidence-based practices and included interactive videos and content, regular self-tests and online discussions. It was offered for the first time to about 120 students last fall. In this presentation, I will discuss the course, the development process and the results of an evaluation that was conducted on the redesign.

Tôt l'an dernier, STAT 1024, un cours d'introduction pour ceux qui ne suivent pas une majeure à l'University Western Ontario, a été sélectionné pour l'Initiative en ligne de l'Ontario par le ministère de la Formation et des Collèges et Universités (MFCU). Grâce à cette initiative, la province développe une série de cours en ligne offerts par les universités de l'Ontario pour les étudiants de toute la province. Les étudiants qui suivent ces cours obtiennent un transfert de crédits équivalent à leur institution d'origine. La refonte du cours STAT 1024 a utilisé des pratiques fondées sur des preuves et a inclus des vidéos et du contenu interactifs, des auto-tests réguliers et des discussions en ligne. Ce cours a été offert pour la première fois à environ 120 étudiants l'automne dernier. Dans cette présentation, je discuterai du cours, du processus de développement et des résultats d'une évaluation menée sur la refonte.

2E-C1: Actuarial Science and Finance 2 Science actuarielle et finance 2

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

Room/Salle: McCain 2198

Abstracts/Résumés

RENÉ FERLAND & FRANÇOIS WATIER (Université du Québec à Montréal)

[Tuesday June 16 / mardi 16 juin, 15:30]

Optimal Weighted Mean-Variance Portfolio Under a No-Bankruptcy Constraint

Portefeuille optimal à variance moyenne pondérée dans un contexte de contrainte de non-faillite

Using the usual Black-Scholes market model, we consider the stochastic portfolio management problem where the investor wishes to attain an expected financial goal at a given terminal date while minimizing a mean-variance based risk measure where more weight (or penalty) is given on downside returns. Under a positive wealth constraint, we show that there exists an explicit optimal strategy given in feedback form.

Dans le contexte du modèle habituel de Black-Scholes, nous examinons le problème stochastique de gestion de portefeuilles pour un investisseur souhaitant atteindre un certain objectif financier à une date limite donnée. La mesure de risque minimisée est du type moyenne-variance pour laquelle une pénalité plus grande est placée sur les rendements négatifs des investissements. Sous une contrainte supplémentaire de richesse positive, nous montrons qu'une stratégie optimale explicite existe et qu'elle peut s'exprimer en terme du niveau courant de la richesse.

FRANÇOIS WATIER (Université du Québec à Montréal), **MOUSSA KOUNTA** (The College of the Bahamas)

[Tuesday June 16 / mardi 16 juin, 15:45]

Regime-Switching Multi-Period Mean-Variance Optimal Portfolio Without Short-Selling

Portefeuille optimal multi-période de type moyenne-variance avec changement de régime sans vente à découvert

We consider a multi-period mean-variance portfolio selection problem where the market parameters' dynamics are subject to a Markov chain. We analytically derive, through dynamic programming, a closed form optimal strategy under a no short-selling constraint (i.e. no stock borrowing).

Nous considérons un problème de sélection de portefeuille de type moyenne-variance où la dynamique des paramètres de marché est régie par une chaîne de Markov. Nous dérivons analytiquement, par la programmation dynamique, une forme fermée pour la stratégie optimale sous une contrainte d'absence de vente à découvert (c'est-à-dire sans emprunt d'actions).

MICHELLE CAREY, JAMES RAMSAY & CHRISTIAN GENEST (McGill University)

[Tuesday June 16 / mardi 16 juin, 16:00]

A Generalized Smoothing Approach to the Calibration, Pricing and Hedging of Financial Options

Une approche de lissage généralisée à l'étalonnage, la tarification et la couverture des options financières

Recently, mathematical motivated financial models have been used to understand the complexity of financial market instruments. The infamous Black-Scholes partial differential equation (PDE) is the current benchmark for option pricing in the deriva-

Récemment, des modèles financiers motivés par les mathématiques ont été utilisés pour comprendre la complexité des instruments du marché financier. L'infâme équation différentielle partielle (EDP) de Black-Scholes est la référence actuelle pour la tarification d'options dans le marché des pro-

2E-C1: Actuarial Science and Finance 2 Science actuarielle et finance 2

tives market. This talk extends the generalized smoothing approach attributable to Ramsay et al. 2007 and Xun et al. 2013 to obtain an estimate of the value of an option, which adheres to the traded price and incorporates the theoretical specifications defined by the PDE, without being completely determined by either competing effect. Our results show a distinct improvement in the calibration, pricing and hedging of financial options.

duits dérivés. Cette présentation étend l'approche de lissage généralisée attribuable à Ramsay et al. 2007 et à Xun et al. 2013 afin d'obtenir une estimation de la valeur d'une option qui adhère au prix négocié et intègre les spécifications théoriques définies par l'EDP, sans qu'un des effets concurrentiels ne la détermine complètement. Nos résultats montrent une nette amélioration de l'étalonnage, la tarification et la couverture des options financières.

JOEL SANGO & PIERRE DUCHESNE (Université de Montréal)

[Tuesday June 16 / mardi 16 juin, 16:15]

On the Diagnostic of Vector Multiplicative Error Models

Sur le diagnostic des modèles vectoriels multiplicatifs à erreurs non négatives

Presented by Cipollini et al. (2006), vector multiplicative error models (VMEM) play an important role in the economic modeling of financial time series. Two main approaches are used to estimate VMEM models: the maximum likelihood method based on copulas and the generalized method of moments (GMM). Up to now, there is no known diagnostic test to validate the specification of these

Présentés par Cipollini et al. (2006), les modèles vectoriels multiplicatifs à erreurs non négatives (VMEM) jouent un rôle très important dans la modélisation économique des données financières. Deux approches principales sont utilisées pour les estimer : celle du maximum de vraisemblance basée sur les copules et la méthode généralisée des moments (GMM). Jusqu'à présent, aucun test n'a encore été proposé pour valider la spécification de ces modèles.

models. In the current work, we propose a specification test based on the generalized method of moment by deriving the asymptotic distribution of residual autocovariance. Key words: GMM, VMEM.

Dans le cadre de ce travail, un test de spécification basé sur la méthode généralisée des moments sera construit en dérivant la loi asymptotique des autocovariances résiduelles.

Mots clés : GMM, VMEM.

MARTIN LYSY (University of Waterloo), **NATESH PILLAI** (Harvard University)

[Tuesday June 16 / mardi 16 juin, 16:30]

Statistical Inference for Stochastic Differential Equations with Memory

Inférence stochastique pour les équations différentielles stochastiques avec mémoire

We propose a framework for statistical inference with discretely-observed stochastic differential equations (SDEs) where the driving noise has "memory". Classical SDE models for inference assume the driving noise to be Brownian motion, thus implying a Markov assumption. We focus on the case when the driving noise is fractional Brownian motion, leading to a broad array of non-Gaussian processes exhibiting long-range dependence. A data augmentation approach is adapted from the classical setting to estimate all model parameters simultaneously. The methodology is illustrated with

Nous proposons un cadre d'inférence statistique pour les équations différentielles stochastiques (EDS) observées de manière discrète où le bruit sous-jacent a une « mémoire ». Les modèles d'inférence EDS classiques partent de l'hypothèse que le bruit sous-jacent est brownien, ce qui implique une hypothèse de Markov. Nous mettons l'accent sur le cas où le bruit sous-jacent est un mouvement brownien fractionnaire, avec un vaste ensemble de processus non gaussiens présentant une dépendance à long terme. Nous adaptons une approche d'augmentation des données du cas classique pour estimer tous les paramètres du modèle simultanément. Nous illustrons notre méthodologie par un modèle de dyna-

2E-C1: Actuarial Science and Finance 2 Science actuarielle et finance 2

a model of financial dynamics and one of microparticle confinement in biological fluids.

mique financière et un autre sur le confinement de microparticules dans les fluides biologiques.

2E-C2: Biostatistics: Methodological Innovation 2 Biostatistique : innovation méthodologique 2

Chair/Présidente: Cindy Feng (University of Saskatchewan)

Room/Salle: McCain 2102

Abstracts/Résumés

SCARLETT KAZIMER (McGill University), **WENYU JIANG** (Queen's University)

[Tuesday June 16 / mardi 16 juin, 15:30]

Sample Size Planning for the Cox Regression Model with Treatment-Biomarker Interaction

Planification de la taille échantillonnale pour le modèle de régression de Cox en présence d'interaction biomarqueur-traitement

Biomarker study has received much attention in clinical trials, and the impact of biomarker on treatment outcomes is typically described by an interaction effect in regression analysis. Using Cox regression model we derive sample size formulas for designing clinical trials that aim to identify the treatment-biomarker interaction effects or varying treatment effects among different biomarker defined patient subsets. The formulas are useful for models with various types of treatment-biomarker interactions, for example, the model with a dichotomized biomarker that captures different treatment effect thresholds on patient subsets. This work is in collaboration with Dr. Wenyu Jiang from Queen's University.

L'étude des biomarqueurs suscite beaucoup d'attention dans les essais cliniques et l'impact du biomarqueur sur le traitement est habituellement décrit comme un effet d'interaction en analyse de régression. À l'aide du modèle de régression de Cox, nous obtenons des formules de taille échantillonnale pour la planification d'essais cliniques qui visent à identifier les effets d'interaction biomarqueur-traitement ou les variations des effets sur le traitement parmi différents sous-groupes de patients définis selon leurs biomarqueurs. Les formules sont utiles pour les modèles présentant divers types d'interactions biomarqueur-traitement, comme le modèle avec un biomarqueur dichotomique qui saisit différents seuils d'effet sur les sous-groupes de patients. Ces travaux sont réalisés en collaboration avec Wenyu Jiang de l'Université Queen's.

MATTHEW STEPHENSON (University of Guelph), **NEIL KLAR** (Western University), **GERARDA DARLINGTON** (University of Guelph), **STEPHANIE DIXON** (Western University)

[Tuesday June 16 / mardi 16 juin, 15:45]

Effect of Accrual on Sample Size Calculation for Cluster Randomized Studies

Effet d'un accroissement sur le calcul de la taille échantillonnale pour des essais randomisés par grappes

In cluster randomized trials, intact social units, or clusters of individuals are randomly assigned to intervention groups. The correlation among responses within clusters is a key challenge of sample size estimation. Sample size estimation methods have been proposed allowing for clustering. A largely unmet challenge for survival outcomes is to allow for accrual at both the cluster and subject level. That is, one may recruit clusters over time while cluster members may be recruited following random assignment. The effect of accrual

Dans des essais randomisés par grappes, les unités sociales intactes ou grappes d'individus sont assignées au hasard à des groupes d'intervention. La corrélation entre les réponses au sein des grappes est un défi clé de l'estimation de la taille des échantillons. Des méthodes d'estimation de la taille échantillonnale permettant une structure en grappes ont été proposées. Un défi toujours largement non résolu concernant les réponses de survie consiste à permettre un accroissement tant au niveau de la grappe que du sujet, c'est-à-dire qu'on peut recruter des grappes au fil du temps, alors que les membres des grappes peuvent être

2E-C2: Biostatistics: Methodological Innovation 2 Biostatistique : innovation méthodologique 2

schemes on sample size is explored by simulation and compared to a newly derived sample size formula.

recrutés suivant une assignation aléatoire. Nous examinons l'effet des schémas d'accroissement sur la taille échantillonnale à l'aide de simulations et le comparons à une formule de taille échantillonnale nouvellement obtenue.

OSVALDO ESPIN-GARCIA (University of Toronto / Lunenfeld-Tanenbaum Research Institute), **LINDA HIRAKI** (The Hospital for Sick Children), **ANGELO J. CANTY** (McMaster University), **ANDREW PATERSON** (The Hospital for Sick Children / University of Toronto), **SHELLEY B. BULL** (Lunenfeld-Tanenbaum Research Institute / University of Toronto)

[Tuesday June 16 / mardi 16 juin, 16:00]

Understanding Joint Models for Longitudinal and Survival Data Analysis of Genetic Association: A Simulation Study
Comprendre les modèles conjoints pour l'analyse des données longitudinales et de survie dans les associations génétiques : une étude de simulation

Cohort study design typically allows mixed model analysis of repeated quantitative trait measures, but inference may be complicated when time trajectories are subject to dropout or censoring. We investigate properties of two alternative formulations of joint models of genotype-trait association, so-called trajectory and shared random effects. Our main focus is the performance of hypothesis testing methods for genotype-time interaction. Our simulations also incorporate available fitting methods for parameter estimation under the two model formulations: two-stage and joint estimation. We evaluate single parameter tests, as well as validity and power of multi-df tests utilizing Wald and likelihood ratio statistics.

Les plans d'études de cohorte permettent généralement l'analyse par modèle mixte des mesures répétées de caractères quantitatifs, mais l'inférence peut être compliquée lorsque les trajectoires temporelles sont sujettes à l'abandon ou à la censure. Nous étudions les propriétés de deux variantes de modèles conjoints de l'association génotype-caractère, trajectoire et effets aléatoires partagés. Nous mettons l'accent sur la performance des méthodes de test d'hypothèses pour l'interaction génotype-temps. Nos simulations intègrent aussi les méthodes d'ajustement disponibles pour l'estimation des paramètres pour les deux variantes de modèles : estimation en deux étapes et conjointe. Nous évaluons les tests à paramètre unique, ainsi que la validité et la puissance des tests à multiples degrés de liberté à l'aide des statistiques de Wald et de rapports des vraisemblances.

WANHUA SU (MacEwan University), **YAN YUAN** (University of Alberta), **MU ZHU** (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 16:15]

Two Threshold-Free Performance Metrics for Medical Tests: Average Positive Predictive Value and Area Under the ROC Curve

Deux mesures de performance sans seuil pour les tests médicaux : valeur prédictive positive moyenne et aire sous la courbe ROC

The area under the receiver operating characteristic curve (AUC) and the average positive predictive value (AP) are two performance metrics for medical tests. Both metrics do not require us to commit to a specific decision threshold. We study both metrics in some depths to elucidate their difference and relationship. More specifically, we explain mathematically why the AP may be more appropriate if the earlier part of the ROC curve is of interest. We propose a framework for making inferences on both

L'aire sous la courbe ROC (ASC) et la valeur prédictive positive moyenne (MP) sont deux indicateurs de performance pour les tests médicaux. Les deux mesures ne nous obligent pas à nous engager à respecter un seuil de décision spécifique. Nous étudions ces deux mesures en profondeur afin d'élucider leur différence et leur relation. Plus précisément, nous expliquons mathématiquement pourquoi la MP peut être plus appropriée si la première partie de la courbe ROC est d'intérêt. Nous proposons un cadre pour faire des inférences à la fois sur l'ASC et la MP. Nous démontrons la per-

2E-C2: Biostatistics: Methodological Innovation 2 Biostatistique : innovation méthodologique 2

the AUC and the AP. The performance of our proposed methods is demonstrated in analysis of real medical tests.

formance des méthodes proposées en analysant des tests médicaux réels.

SIU HUNG CHEUNG (The Chinese University of Hong Kong), **TONGYU LU** (China Jiliang University), **WAI YIN POON** (The Chinese University of Hong Kong)

[Tuesday June 16 / mardi 16 juin, 16:30]

Multiple Comparisons with a Control for Treatments with Skewed Ordinal Responses

Comparaisons multiples au moyen d'un contrôle des traitements avec des réponses ordinales asymétriques

Clinical trials often involve the comparisons of treatments with ordinal responses. The latent normal model has recently been proposed for comparing several treatments with a control when responses are ordered categorical. However, the latent normal model relies on the assumption of that the underlying distributions are symmetric and does not perform well when the data are highly skewed. We propose to use the latent Weibull model and this new approach is superior to the latent normal model.

Des tests cliniques font souvent appel à des comparaisons de traitements avec des réponses ordinales. Le modèle latent normal a récemment été proposé pour comparer plusieurs traitements au moyen d'un contrôle lorsque les réponses sont catégoriques ordonnées. Cependant, le modèle latent normal est fondé sur l'hypothèse que les distributions sous-jacentes sont symétriques et il ne fonctionne pas bien lorsque les données sont très asymétriques. Nous proposons d'utiliser le modèle latent de Weibull. Cette nouvelle approche est meilleure que le modèle latent normal.

ABDELAZIZ ALSUBIE & JIANAN PENG (Acadia University)

[Tuesday June 16 / mardi 16 juin, 16:45]

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

Inférence simultanée pour l'estimation du risque à faible dose avec des données quantiques pour l'analyse comparative

Risk assessment studies where human, animal or ecological data are used to set safe low dose levels of a toxic agent are challenging as study information is limited to high dose levels of the agent. Simultaneous hyperbolic confidence bands for low-dose risk estimation with quantal data have been proposed in the literature. In this talk, a new method using three-segment confidence bands to construct simultaneous upper confidence limits on extra risks and simultaneous lower bounds on the benchmark dose for quantal data is proposed. The proposed method is illustrated with a real data application and simulation studies.

Les études d'évaluation du risque dans lesquelles les données humaines, animales ou écologiques sont utilisées pour établir des niveaux sécuritaires à faible dose d'un agent toxique constituent un bon défi, puisque l'information des études est limitée à des niveaux de doses élevées d'agent. Des bandes de confiance hyperboliques simultanées ont été proposées dans la littérature pour l'estimation du risque à faible dose avec des données quantiques. Cet exposé porte sur une nouvelle méthode applicable aux données quantiques qui utilise des bandes de confiance à trois segments permettant de construire des limites de confiance supérieures simultanées pour les risques supplémentaires et des bornes inférieures simultanées pour la dose de référence. La méthode proposée est illustrée à l'aide de données réelles et simulées.

2E-C3: Data Visualization and Multivariate Methods Visualisation de données et méthodes multivariées

Chair/Président: Richard Lockhart (Simon Fraser University)

Room/Salle: McCain 2170

Abstracts/Résumés

WAYNE OLDFORD & ADRIAN WADDELL (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 15:30]

Loon: An Interactive and Extendable Statistical Data Visualization Toolkit

Loon : une boîte à outils interactive et extensible pour la visualisation de données statistiques

Loon is a data visualization toolkit for interactive exploratory data analysis, using either a direct manipulation interface (via mouse gestures effecting interactive zooming, panning, selection and moving of points) or a command line interface (e.g. from R, Python, or Tcl). Plots are dynamically linkable, contain arbitrary layers (e.g. maps and regression lines), and have customized points (e.g. images, texts, star glyphs). Plot inspectors provide a unified GUI to edit and oversee displays. Mouse, keyboard, and even plot change events can be bound to trigger user defined code (including R) and so create dramatically interactive interfaces for nearly any analysis purpose.

Loon est une boîte à outils de visualisation de données pour l'analyse interactive de données exploratoires qui fait appel à une interface de manipulation directe (par les mouvements de la souris qui activent l'agrandissement, la planification, la sélection et le déplacement interactifs de points) ou à une interface de ligne de commande (p. ex. comme dans R, Python ou Tcl). Les graphes sont connectables dynamiquement et contiennent des couches arbitraires (p. ex. cartes et droites de régression) ainsi que des points personnalisés (p. ex. images, textes, glyphes étoilés). Les inspecteurs des graphes fournissent une interface graphique (GUI) unifiée permettant d'éditer et de coordonner les affichages. La souris, le clavier et même les événements de changement au graphe peuvent être liés pour produire un code défini par l'utilisateur (y compris R) et ainsi créer des interfaces hautement interactives pour presque tout type d'analyse.

JEFFREY PICKA (University of New Brunswick)

[Tuesday June 16 / mardi 16 juin, 15:45]

The Need for Replication in Exploratory Computer Experiments

Le besoin de réplication dans les expériences exploratoires sur ordinateur

In complex systems such as turbulent flow, simulated flows are used to search for spatial statistics which may summarize essential aspects of those flows. If this exploration is done using a designed computer experiment, then the experiments may be greatly misleading unless the computer models are stochastic and the designs include replication for each treatment. The difficulties arising from not using replicates will be explained in the context of the unpredictability of the underlying dynamical system.

Dans des systèmes complexes comme un écoulement turbulent, des courants simulés sont utilisés pour rechercher des statistiques spatiales qui résument les caractéristiques essentielles de ces courants. Si cette analyse est effectuée au moyen d'une expérience informatique, les résultats peuvent être faussés à moins que les modèles informatiques soient stochastiques et que les planifications comprennent des répétitions pour chaque traitement. Les défis découlant de l'absence de répétition sont expliqués dans le contexte de l'imprévisibilité du système dynamique sous-jacent.

2E-C3: Data Visualization and Multivariate Methods Visualisation de données et méthodes multivariées

LIANG JIAXI, SHOJAEDDIN CHENOURI & CHRISTOPHER SMALL (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 16:00]

Performance Analysis for Dimensionality Reduction

Analyse de performance pour la réduction de dimension

While lots of dimension reduction algorithms are developed, the question of how to assess and compare the performances of these different methods is still not well answered. In this talk we will establish a local rank-correlation type of goodness measure to address this issue. The proposed method measures the degree of topology preservation of the low-dimensional configuration. It is robust and easily interpretable. We will also show the possible application of this measure in some related problems, such as neighborhood size selection, and intrinsic dimension estimation.

Bien qu'il y ait plusieurs algorithmes de réduction de dimension développés, la question de comment évaluer et comparer les performances de ces différentes méthodes n'est toujours pas répondue. Dans cette présentation, nous établirons une mesure locale d'adéquation entre le rang et le type de corrélation afin d'adresser cette question. La méthode proposée mesure le degré de préservation topologique de la configuration à basse dimension. Elle est robuste et peut facilement être interprétée. Nous montrerons aussi l'application possible de cette mesure à l'aide de problèmes reliés, tels que la sélection de la taille du voisinage, ainsi que l'estimation de la dimension intrinsèque.

SHOJAEDDIN CHENOURI (University of Waterloo), **BEHDAD MOSTAFAYI & MOHAMMADREZA FARIDROHANI** (Shahid Beheshti University)

[Tuesday June 16 / mardi 16 juin, 16:15]

Sparse Functional Linear Regression and Singular Function Estimation by Reproducing Kernel Hilbert Space Methods

Régression linéaire fonctionnelle simultanément éparses et estimation singulière des fonctions par la reproduction des méthodes d'espace de Kernel Hilbert

In this paper, we study functional linear regression in which both the response and the predictor are sparsely sampled functions of a common variable such as time. To estimate the covariance and the cross-covariance functions we employ a regularization method over a reproducing kernel Hilbert space and use these estimates to obtain an estimate of the regression coefficient function and also functional singular components. We derive the convergence rates of the proposed cross-covariance, the regression coefficient and the singular component function estimators and show that the estimator of the coefficient function β has a minimax optimal rate.

Dans cet article, nous étudions la régression linéaire fonctionnelle dans laquelle la réponse et le prédicteur sont des fonctions échantillonnées sur une grille peu dense, comme une variable commune telle que le temps. Nous employons une méthode de régularisation sur un espace de Hilbert à noyau reproduisant pour estimer les fonctions de covariance et d'intercovariance et pour obtenir une estimation de la fonction du coefficient de régression et des composants singuliers fonctionnels. Nous dérivons les taux de convergence de l'intercovariance proposée, le coefficient de régression et les estimateurs de la fonction de composants singulières. Nous montrons également que l'estimateur de la fonction du coefficient β a un taux optimal minimax.

2E-C3: Data Visualization and Multivariate Methods Visualisation de données et méthodes multivariées

JULES JOSEPH S. DE TIBEIRO (Université de Moncton, Campus de Shippagan)

[Tuesday June 16 / mardi 16 juin, 16:30]

Statistical Validity Tests Based on Usual Probabilistic Models: An Alternative to Correspondence Analysis Based on Hellinger Metric.

Tests de validité statistique basés sur des modèles classiques de probabilité : une alternative à l'analyse des correspondances fondée sur la métrique de Hellinger.

To identify etiologial relationships between the various categories of congenital heart defects, Fraser and Hunter (1975), considered a complex statistical distribution of these deformities in consanguineous subjects, mainly children with the same parents (or those with a single common parent). This distribution is named FH dataset. The question therefore arises, here, as to how and to what extent, to assess these two defects that are separated, but between which there is a certain similarity. Many authors have re-examined the FH dataset. We believe to approach here the FH dataset from the point of view of Correspondence Analysis using Hellinger metric.

En vue de cerner des relations étiologiques entre les diverses catégories de malformations cardiaques congénitales, Fraser et Hunter (1975), ont considéré une distribution statistique complexe de ces malformations chez des sujets consanguins, essentiellement des enfants ayant les mêmes parents. Cette distribution est appelée tableau FH. La question se pose donc d'apprécier dans quelle mesure, même si les deux malformations sont distinctes, il y a entre elles une certaine similitude. Plusieurs auteurs ont examiné ce tableau FH. Nous croyons pouvoir y revenir ici, du point de vue de l'analyse des correspondances à travers la métrique de Hellinger.

2E-C4: Survey Methodology Méthodes d'enquêtes

Chair/Président: Noel Cadigan (Memorial University)

Room/Salle: McCain 2184

Abstracts/Résumés

J. N. K. RAO (Carleton University), **YONGSONG QIN** (Guangxi Normal University), **MALGORZATA WINISZEWSKA** (Carleton University)

[Tuesday June 16 / mardi 16 juin, 15:30]

Confidence Intervals for Mean and Distribution Function Under Imputation for Missing Data

Intervalle de confiance pour la moyenne et la fonction de répartition sous imputation pour les données manquantes

Missing observations are commonly encountered in data from sample surveys due to item non-response. Imputation is used to compensate for non-response. In particular, single or fractional imputation is often used. This paper develops asymptotically valid bootstrap confidence intervals for the mean and the distribution function, using the bootstrap percentile and the bootstrap empirical likelihood methods, under fractional imputation. Imputation is done independently within classes formed on the basis of known auxiliary variables. Performance of the proposed confidence intervals is studied through simulations.

Les observations manquantes sont couramment observées dans les échantillons d'enquêtes en raison de la non-réponse à une question. L'imputation est utilisée pour compenser la non-réponse. En particulier, l'imputation simple ou fractionnaire est souvent employée. Cet article développe des intervalles de confiance bootstrap asymptotiquement valides pour la moyenne et la fonction de répartition en utilisant le centile bootstrap et les méthodes de vraisemblance empirique bootstrap, sous imputation fractionnaire. L'imputation est faite indépendamment à l'intérieur des classes formées sur la base de variables auxiliaires connues. La performance des intervalles de confiance proposés est étudiée à l'aide de simulations.

CAREN HASLER (Université de Neuchâtel), **RADU V. CRAIU** (University of Toronto)

[Tuesday June 16 / mardi 16 juin, 15:45]

Nonparametric Imputation Method for Nonresponse in Surveys

Imputation non paramétrique pour la non-réponse dans les enquêtes

Many imputation methods are based on statistical models that assume that the variable of interest is a noisy observation of a function of the auxiliary variables. Misspecification of this model may lead to severe errors in estimates and to misleading conclusions. A new imputation method for item nonresponse in surveys is proposed based on a nonparametric estimation of the functional dependence between the variable of interest and the auxiliary variables. We consider the use of smoothing spline estimation within an additive model framework to flexibly build an imputation model in the case of multiple auxiliary variables.

La plupart des méthodes d'imputation sont basées sur un modèle qui suppose que la variable d'intérêt se décompose comme la somme d'une fonction des variables auxiliaires et d'un terme d'erreur. Une mauvaise spécification de ce modèle peut engendrer de sévères erreurs dans les estimations et amener à des conclusions erronées. Une nouvelle méthode d'imputation pour la non-réponse partielle dans les enquêtes est proposée. La fonction de dépendance entre la variable d'intérêt et les variables auxiliaires est estimée non paramétriquement au moyen de splines de régression considérées dans un cadre additif, ce qui permet de construire un modèle d'imputation très flexible.

2E-C4: Survey Methodology Méthodes d'enquêtes

KYLE VINCENT (Bank of Canada), **STEVE THOMPSON** (Simon Fraser University)

[Tuesday June 16 / mardi 16 juin, 16:00]

Estimating the Size and Distribution of Networked Populations with Snowball Sampling

Estimer la taille et la distribution de populations en réseau avec l'échantillonnage en boule de neige

A new strategy is introduced for estimating networked population characteristics. Sample selection is based on the one-wave snowball sampling design. A generalized stochastic block model is posited for the population's network topology. Inference is based on a Bayesian data augmentation procedure. This procedure has the advantage over existing methods in that it can be applied to a networked population of unknown size. An application is provided to a study of an empirical population at risk for HIV/AIDS. The results demonstrate that efficient estimates of the size and distribution of the population can be achieved with this novel strategy.

Une nouvelle stratégie est introduite afin d'estimer les caractéristiques de populations en réseau. La sélection des échantillons est basée sur un plan d'échantillonnage en boule de neige. Un modèle de bloc stochastique généralisé est posé pour la topologie du réseau de la population. L'inférence est basée sur la procédure d'augmentation des données bayésienne. Cette procédure a l'avantage sur les autres méthodes existantes de pouvoir être appliquée à une population en réseau de taille inconnue. Une application est fournie pour étudier la population empirique à risque pour le VIH/SIDA. Les résultats démontrent que des estimations efficaces de la taille et de la distribution de la population peuvent être obtenues avec cette nouvelle stratégie.

VALÉRY DONGMO JIONGO (Statistique Canada), **PIERRE NGUIMKEU** (Georgia State University)

[Tuesday June 16 / mardi 16 juin, 16:15]

Bootstrapping Mean Squared Errors of Robust Small Area Estimators

Bootstrap de l'erreur quadratique moyenne d'estimateurs robustes de petits domaines

Robust small-area estimation has received considerable attention in recent years, and the mean squared error has been the main way in which the estimators performance is measured. This paper proposes a new bootstrap procedure for mean squared errors of robust small area estimators. We formally prove the asymptotic validity of the proposed bootstrap method and examine its finite sample performance through Monte Carlo simulations. The results show that our procedure performs reasonably well and outperforms existing ones. We also provide a real data example to illustrate the usefulness of the proposed bootstrap method in practice.

L'estimation robuste de petits domaines a reçu une attention considérable récemment, et l'erreur quadratique moyenne est la façon principale par laquelle la performance des estimateurs est mesurée. Cet article propose une nouvelle procédure de bootstrap pour les erreurs quadratiques moyennes d'estimateurs robustes de petits domaines. Nous démontrons formellement la validité asymptotique de la méthode de bootstrap proposée et nous examinons sa performance à travers des simulations Monte-Carlo. Les résultats démontrent que notre procédure performe relativement bien et est supérieure aux méthodes existantes. Nous présentons aussi un exemple de données réelles afin d'illustrer l'utilité pratique de la méthode de bootstrap proposée.

XICHEN SHE & CHANGBAO WU (University of Waterloo)

[Tuesday June 16 / mardi 16 juin, 16:30]

Analysis of Ordinal Survey Responses with "Don't Know"

Analyse de réponses de sondage ordinales « Je ne sais pas »

Ordinal responses are frequently involved in social and health survey researches to evaluate performance, attitude, severity of diseases, etc... It is

Les réponses ordinales sont souvent utilisées dans les recherches par sondage sociales ou de santé afin d'évaluer la performance, l'attitude, la sévérité des maladies, etc...

2E-C4: Survey Methodology Méthodes d'enquêtes

also a common practice to list "Don't Know" as an option in the responses, especially for questions with sensitive nature. In this talk, we first briefly introduce approaches dealing with regular ordinal data, then explore methods for analyzing ordinal responses with "Don't Know" as part of the response. Consistency and efficiency are compared among alternative estimators and results from a limited simulation study will be discussed.

Il est aussi pratique commune d'inscrire « Ne sais pas » comme option de réponse, particulièrement pour les questions de nature plus sensible. Dans cette présentation, nous présenterons d'abord brièvement les approches pour gérer les données ordinales régulières, puis nous explorons des méthodes pour analyser les réponses ordinales avec « Ne sais pas » comme composante de réponse. La cohérence et l'efficacité sont comparées entre les divers estimateurs et les résultats d'une simulation limités seront discutés.

CHRISTIANE LAPERRIERE (Statistics Canada)

[Tuesday June 16 / mardi 16 juin, 16:45]

Data Collection Using a Diary: The Experience of the Survey of Household Spending

Collecte de données à l'aide d'un carnet : l'expérience de l'Enquête sur les dépenses des ménages

The Survey of Household Spending collects household expenditure data using a personal interview as well as an expenditure diary. While the diary gives us detailed expenditure information, it comes with a set of challenges that are unique to this collection mode. For example, respondents' fatigue increases with the use of diaries, which may lead to under-reporting of certain key expenditures, thus requiring further adjustments. In this talk, we discuss challenges and present the strategies we have implemented so far as well as potential strategies we seek to pursue in the future.

L'Enquête sur les dépenses des ménages recueille des données sur les dépenses des ménages au moyen d'un entretien personnel et d'un carnet de dépenses quotidiennes. Bien que le carnet fournisse des renseignements détaillés sur les dépenses, il comporte son lot de défis uniques à ce mode de collecte. Par exemple, la lassitude des répondants augmentent avec l'usage des carnets, ce qui peut entraîner une sous-déclaration de certaines dépenses clés et nécessiter par conséquent des ajustements supplémentaires. Dans cet exposé, nous abordons ces défis et présentons les stratégies que nous avons mises en œuvre jusqu'ici, ainsi que celles que nous envisageons d'examiner.

3A-A1: CRM-SSC Prize in Statistics Address
Allocution du récipiendaire du Prix CRM-SSC en statistique

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

Room/Salle: McCain - Scotiabank

Abstract/Résumé

MATÍAS SALIBIÁN-BARRERA (University of British Columbia)

[Wednesday June 17 / mercredi 17 juin, 8:45]

Robust Statistics, from Pessimism to Realism

Statistique robuste : du pessimisme au réalisme

Robust Statistics incorporates potential partial model inadequacy into the statistical analysis process, from parameter estimation to inference. Unlike sampling variability, this type of uncertainty does not diminish with larger sample sizes. One of its symptoms is the presence of possible outliers, which can be particularly hard to detect in the increasingly complex data sets available today. This problem has been recognized recently in the literature, where there is growing awareness of the need for robust methods. In this talk I will describe some practical and conceptual challenges in Robust Statistics, discuss some solutions, and present recent advances in the field.

La statistique robuste intègre une insuffisance partielle potentielle du modèle dans le processus d'analyse statistique, à savoir de l'estimation de paramètres à l'inférence. À la différence de la variabilité due à l'échantillonnage, ce type d'incertitude ne diminue pas avec des échantillons de plus grande taille. Il se traduit entre autres par la présence de valeurs aberrantes, qui peuvent être particulièrement difficiles à détecter dans les ensembles de données qui sont aujourd'hui de plus en plus complexes. Ce problème a récemment été reconnu dans les ouvrages publiés où il y a une prise de conscience grandissante de la nécessité d'avoir des méthodes robustes. Dans cet exposé, je décrirai certains défis pratiques et conceptuels en statistique robuste. Je présenterai des solutions et les récentes évolutions dans ce domaine.

3A-A2: SSC Impact Award Address
Allocution du récipiendaire du Prix pour impact de la SSC

Chair/Président: Richard Routledge (Simon Fraser University)
Organizer/Responsable: Thérèse Stukel (Institute for Clinical Evaluative Sciences)

Room/Salle: McCain - Ondaatje

Abstract/Résumé

JOHN PETKAU (University of British Columbia)

[Wednesday June 17 / mercredi 17 juin, 8:45]

Collaborative Research Impacting on Clinical Trials in Multiple Sclerosis

Recherche concertée avec des répercussions sur les essais cliniques portant sur la sclérose en plaques

Much of my research has focused on improving the design, execution and analysis of clinical trials in multiple sclerosis. An underlying theme is developing approaches that make better use of the rich longitudinal data collected in such trials. I will briefly describe some past work before discussing in more detail recent work on longitudinal models for counts of magnetic resonance imaging lesions to detect unusual increases in individual patients and a new approach to trial design based on within-lesion signal intensities that is intended for the assessment of emergent therapies focusing on tissue repair. I will attempt to highlight both the challenges and the rewards that have arisen in the course of this collaborative research.

Une part importante de ma recherche est dédiée à l'amélioration de la conception, de l'exécution et de l'analyse d'essais cliniques portant sur la sclérose en plaques. Un thème sous-jacent vise le développement d'approches qui font meilleur usage des grandes quantités de données longitudinales recueillies dans ces essais. Je décrirai brièvement certains travaux antérieurs avant de discuter plus en détail de projets récents sur des modèles longitudinaux de décomptes de lésions détectées à l'imagerie par résonance magnétique visant à repérer des hausses inhabituelles chez des patients, ainsi qu'une nouvelle méthode de conception d'essais basée sur l'intensité des signaux intralésions et destinée à évaluer de nouvelles thérapies axées sur la réparation des tissus. Je tenterai de souligner les défis et les gains survenus au cours de cette recherche concertée.

3B-I1: Actuarial, Financial and Quantitative Risk Management Gestion du risque actuariel, financier et quantitatif

Chair/Président: Jose Garrido (Concordia University)
Organizer/Responsable: Ruodo Wang (University of Waterloo)

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

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

ADAM METZLER (Wilfrid Laurier University)

[Wednesday June 17 / mercredi 17 juin, 10:20]

State Dependent Correlations in Linear Factor Models

Corrélations dépendant de l'état dans les modèles à facteurs linéaires

For better or worse, linear factor models are used to model correlations in many risk management applications. Shortcomings of the basic linear model are well known, and several variants have been proposed to alleviate specific problems. Unfortunately very few allow for state-dependence in correlations, and (as we show) failure to include this feature can substantially underestimate a wide variety of risk measures. In this talk we present an empirically motivated model that allows for state-dependent correlations in linear factor models and contains both the simple mixture and so-called Random Factor Loading models as special cases. We derive a number of tractable asymptotic approximations and illustrate that state-dependence in correlations effectively precludes moderate default rates, exacerbating both good times and bad.

Que cela nous plaise ou non, de nombreuses applications de gestion des risques emploient des modèles à facteurs linéaires pour modéliser les corrélations. Les faiblesses du modèle linéaire de base étant bien connues, plusieurs variantes ont été mises au point pour remédier à divers problèmes spécifiques. Malheureusement, très rares sont les modèles qui permettent la dépendance de l'état dans les corrélations, or (ainsi que nous le montrons) l'omission de cette fonctionnalité conduit à une nette sous-estimation de nombreuses mesures de risque. Nous présentons ici un modèle d'inspiration empirique qui permet les corrélations dépendant de l'état dans les modèles à facteurs linéaires et qui inclut comme cas spéciaux des modèles de mélange simple et des modèles dits de coefficient de saturation aléatoire. Nous dérivons plusieurs approximations asymptotiques souples et montrons que cette dépendance de l'état élimine en fait les taux de défaillance modérés, exacerbant donc à la fois les hauts et les bas.

JIANDONG REN (Western University), **QIMING HE** (University of Waterloo)

[Wednesday June 17 / mercredi 17 juin, 10:50]

Parameter Estimation of Discrete Multivariate Phase-Type Distributions

Estimation des paramètres des distributions de type phase multivariées en temps discret

This paper considers parameter estimation of a class of discrete multivariate phase-type distributions (DMPH). Such discrete phase-type distributions are based on discrete Markov chains with marked transitions introduced by He and Neuts

Cet article étudie l'estimation des paramètres d'une classe de distributions de type phase multivariées en temps discret (ou DMPH). Ces distributions de type phase en temps discret sont basées sur les chaînes de Markov en temps discret avec transitions marquées, introduites par He et Neuts

3B-I1: Actuarial, Financial and Quantitative Risk Management Gestion du risque actuariel, financier et quantitatif

(1998) and are a generalization of the discrete univariate phase-type distributions. Properties of the DMPHs are presented. An EM algorithm is developed for estimating the parameters for DMPHs. A number of numerical examples are provided to address some interesting parameter selection issues and to show possible applications of DMPHs.

(1998), et sont une généralisation des distributions de type phase univariées en temps discret. Nous présentons les propriétés des DMPH et développons un algorithme EM permettant d'estimer les paramètres des DMPH. Nous donnons ensuite des exemples numériques pour répondre à des problèmes de sélection de paramètres intéressants et montrer les applications potentielles des DMPH.

CHENGGUO WENG (University of Waterloo)
[Wednesday June 17 / mercredi 17 juin, 11:20]
Optimal Reinsurance Design
Optimisation de la réassurance

Reinsurance can be an effective risk management technique for an insurer. An appropriate use of reinsurance reduces the adverse risk exposure of an insurer and improves the overall viability of the underlying business. The use of reinsurance, on the other hand, incurs additional cost to the insurer in the form of reinsurance premium. This implies that an insurer is faced with a classical tradeoff between risk spreading and risk retaining. An optimal reinsurance design aims at striking a balance between these two conflicting objectives. In this talk, I will introduce some of my recent works on this topic.

La réassurance est une technique de gestion des risques intéressante pour les assureurs. L'une de ses applications permet à l'assureur de réduire son exposition aux risques négatifs et d'améliorer la viabilité globale de son activité. Cependant, la réassurance donne lieu à des dépenses additionnelles sous la forme des primes de réassurance. L'assureur doit donc trouver un compromis entre répartition et conservation du risque. L'optimisation de la réassurance vise à établir un équilibre entre ces deux objectifs contradictoires. Je présente ici mes derniers travaux à ce sujet.

**3B-I2: DASF 90 - A Celebration of the Contributions of Don Fraser, OC
DASF 90 -- Une célébration des contributions de Don Fraser, OC**

Organizer and Chair / Responsable et présidente: Nancy Reid (University of Toronto)

Room/Salle: McCain - Scotiabank

Abstracts/Résumés

STEPHEN FIENBERG (Carnegie Mellon University)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Fraserian Statistical Inference

Inférence statistique fraserienne

In statistics we often read of Bayesian Inference or Fisherian Inference, sometimes known as Fiducial Inference. But in Canada, over the past 60 years, we have observed the evolution of a related statistical approach rooted in the longstanding work of D.A.S. Fraser. In this talk we discuss some aspects of Fraserian Inference and its influence, including on the present author.

En statistique, il est souvent question d'inférence bayésienne et d'inférence fisherienne, aussi connue sous le nom d'inférence fiduciaire. Au Canada cependant, nous avons observé au cours des 60 dernières années l'évolution d'une approche statistique connexe ancrée dans les travaux de longue date de D.A.S. Fraser. Cet exposé porte sur certains aspects de l'inférence fraserienne et son influence, y compris sur le présent auteur.

ANA-MARIA STAICU, YICHI ZHANG & ARNAB MAITY (North Carolina State University)

[Wednesday June 17 / mercredi 17 juin, 10:40]

Testing for Additivity in Nonparametric Regression

Tests d'additivité en régression non paramétrique

In this talk I will discuss a novel approach for testing for additivity in nonparametric regression. We represent the model using a linear mixed model framework and equivalently re-write the original testing problem as testing for a subset of zero variance components. We propose two testing procedures: the restricted likelihood ratio test and the generalized F test. We develop the finite sample null distribution of the restricted likelihood ratio testing and generalized F test using the spectral decomposition of the restricted likelihood ratio and the residual sum of squares, respectively. Numerical investigation shows that the proposed testing procedures overperform the available methods for both fixed and random designs in terms of size and power.

Cet exposé porte sur une nouvelle approche de tests d'additivité en régression non paramétrique. Nous présentons le modèle dans le cadre d'un modèle linéaire mixte et ré-écrivons de façon équivalente le problème de test original comme le test pour un sous-ensemble de composantes à variance nulle. Nous proposons deux procédures de tests : le test du rapport de vraisemblance restreint et le test F généralisé. Nous développons la distribution nulle sous échantillon fini pour les deux tests à l'aide de la décomposition spectrale du rapport de vraisemblance restreint et de la somme des carrés des résidus, respectivement. Une étude numérique montre que les procédures de tests proposées surpassent les méthodes existantes tant pour les plans fixes qu'aléatoires en termes de taille et de puissance.

GRACE YI (University of Waterloo), **YANYUAN MA** (University of South Carolina), **DONNA SPIEGELMAN** (Harvard School of Public Health), **RAYMOND CARROLL** (Texas A&M University)

[Wednesday June 17 / mercredi 17 juin, 11:00]

Functional and Structural Methods with Mixed Measurement Error and Misclassification in Covariates

Méthodes fonctionnelles et structurelles tenant compte de l'erreur de mesure et de classification des covariables

3B-I2: DASF 90 - A Celebration of the Contributions of Don Fraser, OC DASF 90 -- Une célébration des contributions de Don Fraser, OC

Covariate measurement imprecisions or errors arise frequently in many areas. It is well known that ignoring such errors can substantially degrade the quality of inference or even yield erroneous results. Although in practice both covariates subject to measurement error and covariates subject to misclassification can occur, research attention in the literature has mainly focused on addressing either one of these problems separately. In this paper, we develop estimation and inference methods that accommodate both characteristics simultaneously. Specifically, we consider measurement error and misclassification in generalized linear models under the scenario that an external validation study is available, and develop several functional and structural methods.

DAVID BELLHOUSE (Western University)

[Wednesday June 17 / mercredi 17 juin, 11:20]

Pre-Don in Toronto

Époque pré-Don à Toronto

Prior to Don Fraser's arrival as a University of Toronto student in the 1940s, there were three general eras in the development of the field of statistics: the Victorian statistics movement, the development of mathematical statistics by Karl Pearson beginning in the late nineteenth century and the Fisherian revolution beginning in the 1920s. Within each of these eras I examine the impact of statistical developments on the faculty and curriculum at the University of Toronto, culminating with a description of statistics in the Department of Mathematics when Don entered the University of Toronto.

Les erreurs ou les imprécisions de mesure des covariables se produisent fréquemment dans de nombreux domaines. Il est bien connu qu'ignorer ces erreurs peut sensiblement réduire la qualité de l'inférence et même mener à des résultats erronés. Même si des covariables peuvent en pratique être sujettes simultanément à une erreur de mesure et à une mauvaise classification, les recherches dans la littérature sont essentiellement axées sur le traitement d'un seul de ces problèmes à la fois. Dans cet article, nous élaborons des méthodes d'estimation et d'inférence qui tiennent compte de ces deux caractéristiques en même temps. Plus précisément, nous examinons l'erreur de mesure et de classification dans des modèles linéaires généralisés dans le cadre d'un scénario où une étude de validation externe existe, et nous développons plusieurs méthodes fonctionnelles et structurales.

Avant l'arrivée de Don Fraser comme étudiant à l'Université de Toronto dans les années 1940, il existait trois ères générales dans le développement de la statistique : le mouvement statistique victorien, le développement de la statistique mathématique par Karl Pearson qui a commencé vers la fin du XIXe siècle et la révolution fisherienne qui a commencé dans les années 1920. Pour chacune de ces ères, j'examine l'impact des progrès statistiques sur la faculté et le curriculum à l'Université de Toronto, pour terminer avec une description de la statistique au département de mathématiques quand Don a fait son entrée à l'Université de Toronto.

3B-I3: Statistical Methods for Genetic Dissection of Complex Traits Méthodes statistiques pour la dissection génétique de caractères complexes

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

Room/Salle: Rowe 1011

Abstracts/Résumés

JERRY LAWLESS (University of Waterloo), **ANDRIY DERKACH** & **LEI SUN** (University of Toronto)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Outcome-Dependent Sampling in Two-Phase Genetic Association Studies

Échantillonnage dépendant des résultats dans des études d'association génétique biphasée

In many studies individuals in a large cohort are differentially sampled and genotyped on the basis of previously observed outcomes or covariates. Case-control studies, where individuals who have a particular condition are sampled with higher probability than those who do not, are a familiar example. Another is where persons with type 1 diabetes are differentially sampled according to their rates of progression of retinopathy or nephropathy. Estimation based on two-phase, outcome-dependent samples has been thoroughly studied, but testing has received less attention. We will present and illustrate methods that unify a wide range of phenotypic outcome models and types of sampling. This is joint work with Andriy Derkach and Lei Sun.

Dans de nombreuses études, les individus d'une cohorte importante sont sélectionnés différemment et génotypés sur la base de résultats ou de covariables observés antérieurement. Les études cas-témoin, où les individus qui ont une condition particulière sont sélectionnés avec une probabilité supérieure que ceux qui n'en ont pas, est un exemple très connu. Un autre exemple est celui où les personnes avec un diabète de type 1 sont sélectionnées différemment selon le rythme de progression de leur rétinopathie ou de leur néphropathie. L'estimation basée sur des échantillons biphasés dépendants des résultats a été minutieusement étudiée, mais l'expérimentation a reçue moins d'attention. Nous présenterons et illustrerons des méthodes qui unifient un large éventail de modèles phénotypiques et des méthodes d'échantillonnage. Ceci est un travail commun avec Andriy Derkach et Lei Sun.

LISA STRUG & **ZEYNEP BASKURT** (University of Toronto)

[Wednesday June 17 / mercredi 17 juin, 10:50]

Evidential Analysis of Genetic Association in General Pedigrees Using Robust Composite Profile Likelihoods

Analyse probante d'association génétique dans des pedigrees quelconques utilisant le profil robuste des vraisemblances composites

We propose composite likelihood (CL) ratios for genetic association analysis of a binary response in pedigrees of varying size and complexity, under the Evidential statistical paradigm (EP). We show that profile CL ratios can be made robust, satisfying the EP requirements of a valid evidence function. Using 7000 SNPs in 68 multi-generational pedigrees from a linkage region for reading disability, we use CL ratios to assess relative support for genetic association. Using a robust working indepen-

Nous proposons des rapports de vraisemblance composite (VC) pour l'analyse de l'association génétique d'une réponse binaire dans des pedigrees de taille et de complexité variables, par le paradigme de preuves statistique (PP). Nous démontrons que les rapports de profil VC peuvent être robustes, satisfaisant ainsi les exigences PP d'une fonction de preuves valide.

Nous servant de 7000 SNP dans 68 pedigrees multi-générationnels d'une région lien pour troubles de lecture, nous utilisons les rapports VC pour évaluer le support relatif

3B-I3: Statistical Methods for Genetic Dissection of Complex Traits

Méthodes statistiques pour la dissection génétique de caractères complexes

dence model, we compare our findings to an analysis in a mixed model framework.

CL ratios, properly robustified, can be used generally in the EP, and we provide an easy to imple-

ment approach to assess association with binary response data and a complex correlation structure.

pour l'association génétique. En utilisant un modèle fonctionnel d'indépendance, nous comparons nos conclusions à

une analyse dans le cadre d'un modèle mixte.

Les rapports VC, rendus convenablement robustes, peuvent généralement être utilisés dans le PP, et nous fournissons une approche facile à mettre en oeuvre pour évaluer une association avec des données de réponses binaires et une structure de corrélation complexe.

ZENY FENG (University of Guelph)

[Wednesday June 17 / mercredi 17 juin, 11:20]

A Genetic Association Test to Identify Pleiotropic Effect with Multivariate Longitudinal Traits

Un test d'association génétique pour identifier l'effet pléiotropique sur des caractères longitudinaux multivariés

In genetic epidemiology, multiple longitudinal traits are often analyzed and exploring the pleiotropic effect is important, where pleiotropy refers to the genetic effect of a single locus on multiple traits. However, performing genome-wide association analysis with more than one longitudinal trait simultaneously is challenging. In this paper, we propose a 2-step strategy to efficiently perform genetic association testing on multiple longitudinal traits simultaneously. Empirical results show that our method controls the type I error rate at nominal levels. Our method is more powerful for detecting the pleiotropic effects than the union of individual tests on each trait. We also apply our method to analyze the Framingham Heart Disease data provided by the 16th Genetic Analysis Workshop (GAW16).

En épidémiologie génétique, de multiples caractères longitudinaux sont souvent analysés et il est important d'étudier l'effet pléiotropique. La pléiotropie réfère à l'effet d'un seul gène sur plusieurs caractères. Par contre, l'analyse pangénomique d'association sur plus qu'un caractère longitudinal constitue un défi. Dans cet article, on propose une stratégie en deux étapes pour tester efficacement l'association génétique sur plusieurs caractères simultanément. Les résultats empiriques démontrent que notre méthode contrôle le taux d'erreur de type I au niveau nominal. La méthode proposée est plus puissante pour détecter les effets pléiotropiques que l'union des tests individuels pour chaque caractère. On applique aussi notre méthode aux données de l'étude Framingham sur les maladies du cœur, fournies par le 16th Genetic Analysis Workshop (GAW16).

3B-I4: Total Survey Error Erreur d'enquête totale

Organizer and Chair / Responsable et présidente: Karla Fox (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section
Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: Mona Campbell 1108

Abstracts/Résumés

MELANIE ABEYSUNDERA, MARIE-HÉLÈNE MIVILLE, KARLA FOX, JEANNINE CLAVEAU, WISNER JOCELYN, FÉLIX LABRECQUE-SYNNOTT & REMAINING MEMBERS OF THE MODE EFFECT WORKING GROUP
(Statistics Canada / Statistique Canada)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Using Total Survey Error Frameworks to Study Mode Effect

L'utilisation des cadres de l'erreur d'enquête totale pour étudier l'effet du mode

An increasing number of surveys at Statistics Canada have introduced electronic questionnaires as the primary mode of data collection. This change has raised questions regarding the potential differences in response outcomes due to collection mode. Definitions of mode and mode effect are proposed. The relationship between mode and the components of total survey error (TSE) is studied using a proposed extension of the Groves-Lyberg (2010) framework which includes an additional branch for collection. Relationships between the components of TSE and the stochastic mechanisms associated with the sampling process are examined. An example of how the framework was used to evaluate the effect of a change in sampling frame in the General Social Survey will be presented.

À Statistique Canada, des questionnaires électroniques sont devenus le mode principal de collecte de données dans un nombre croissant d'enquêtes. Ce changement a soulevé des questions concernant les différences potentielles dans les résultats de participation en raison du mode de collecte. Des définitions de mode et des effets de mode sont proposés. Nous étudions la relation entre le mode et les composants de l'erreur d'enquête totale en proposant d'étendre le cadre de l'erreur d'enquête totale de Groves-Lyberg (2010) qui comprend une branche supplémentaire de collecte. Nous examinons la relation entre les composants de l'erreur d'enquête totale et les mécanismes stochastiques associés au processus d'échantillonnage. Nous présenterons un exemple sur la façon dont le cadre a été utilisé pour évaluer l'effet d'un changement dans la base d'échantillonnage dans l'enquête sociale générale.

JENNY THOMPSON, BRODERICK OLIVER & JENNIFER BECK (U.S. Census Bureau)

[Wednesday June 17 / mercredi 17 juin, 10:50]

Using Response Rates and Other Quality Metrics to Assess the Effects of the Mixed Collection Modes for Business Surveys: A Case Study

Utilisation des taux de réponse et d'autres mesures de qualité pour évaluer les effets des modes de collecte mixtes pour les enquêtes auprès des entreprises : une étude de cas

In the last decade, offering multiple modes of data collection has become increasingly popular. However, the benefits of offering multiple modes should

Au cours de la dernière décennie, l'offre de multiples modes de collecte de données est devenue de plus en plus populaire. Toutefois, les avantages des modes de collecte mul-

3B-14: Total Survey Error Erreur d'enquête totale

not come at the cost of data quality. Using historic data from two federal business surveys, we investigate data quality as a function of mode of data collection using various quality measures, including the unit response rate (the unweighted proportion of responding reporting units) and the quantity response rate (the weighted proportion of an estimate obtained from reported data). For these analyses, we associate data quality as the percentage of retained reported data after processing. The results suggest mode-based differences in data quality. We discuss the implication of these results for multi-mode data collection.

tiples ne doivent pas se faire au détriment de la qualité des données. En utilisant des données historiques à partir de deux enquêtes fédérales auprès des entreprises, nous étudions la qualité des données en fonction du mode de collecte des données à l'aide de diverses mesures de qualité, incluant le taux de réponse des unités (la proportion non pondérée des unités répondantes rapportées) et le taux de réponse de l'ensemble (la proportion pondérée estimée à partir des données rapportées). Pour ces analyses, nous associons la qualité des données comme le pourcentage des données rapportées retenues après le traitement. Les résultats suggèrent des différences dans la qualité des données selon le mode de collecte. Nous discutons de l'implication de ces résultats pour la collecte de données multi-mode.

BRADY WEST (University of Michigan), **MICHAEL ELLIOTT** (University of Michigan-Ann Arbor)

[Wednesday June 17 / mercredi 17 juin, 11:20]

New Methodologies for the Study and Decomposition of Interviewer Effects in Surveys

Nouvelles méthodologies pour l'étude et la décomposition des effets des intervieweurs dans les enquêtes

Methodological studies of the effects that human interviewers can have on the quality of survey data have long been limited by two critical assumptions: that interviewers in a given survey are assigned random subsets of the larger overall sample that is being studied (also known as interpenetrated assignment), and interviewer effects arise entirely from measurement difficulties, rather than selection effects due to differential sample assignments or nonresponse. In this presentation, we will introduce two new ideas for overcoming a lack of interpenetrated assignment when estimating interviewer effects, and discuss an approach using multilevel modeling and multiple imputation to decompose interviewer effects. Selected methods will be illustrated using data from the 2012 Behavioral Risk Factor Surveillance System (BRFSS).

Des études méthodologiques des effets que les intervieweurs humains peuvent avoir sur la qualité des données d'enquête ont longtemps été limitées par deux hypothèses importantes : que dans une étude donnée, on attribue aux intervieweurs des sous-ensembles aléatoires d'un plus grand échantillon global qui est étudié (aussi appelé attribution imbriquée), et que les effets des intervieweurs proviennent entièrement des difficultés de mesure, plutôt que des effets de sélection en raison des attributions d'échantillons différentiels ou de non-participation. Dans cette présentation, nous vous présenterons deux nouvelles idées pour pallier un manque d'attribution imbriquée lorsqu'on estime les effets des intervieweurs. Nous discuterons d'une approche basée sur une modélisation à plusieurs niveaux et une imputation multiple pour décomposer les effets des intervieweurs. Les méthodes sélectionnées seront illustrées par des données de 2012 du système américain de surveillance des facteurs de risque pour la santé (Behavioral Risk Factor Surveillance System -BRFSS).

3B-15: Using TMB to Quickly and Robustly Solve Problems from Marine Ecology Comment utiliser TMB pour résoudre des problèmes en écologie marine de manière rapide et robuste

Organizer and Chair / Responsable et présidente: Joanna Mills Flemming (Dalhousie University)

Room/Salle: Rowe 1028

Abstracts/Résumés

CHRISTOFFER ALBERTSEN & KASPER KRISTENSEN (Technical University of Denmark)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Maximum Likelihood Estimation Using Template Model Builder

Estimation du maximum de vraisemblance au moyen du modélisateur Template Model Builder

Maximum Likelihood Estimation consists of three steps: Describing the probability of the data, solving the score equations to find an estimate, and using the Fisher information to describe the variance of the estimator. This is not always easy. In hierarchical models the joint distribution of observed and unobserved data is easy to describe, but the marginal distribution of observed data is often not. Even in fully observed models finding the score functions can be difficult and time consuming. These difficulties only multiply when we try to find the (observed) Fisher information. The R package Template Model Builder (TMB) combines automatic differentiation with the Laplace approximation for general model building in a Maximum Likelihood framework.

L'estimation du maximum de vraisemblance consiste en trois étapes : la description de la probabilité des données, la résolution des équations de score pour trouver une estimation, et l'utilisation de l'information de Fisher pour décrire la variance de l'estimateur. Ce n'est pas toujours facile. Dans des modèles hiérarchiques, la loi conjointe des données observées ou non observées est facile à décrire, mais la loi marginale de données observées ne l'est pas souvent. Même dans les modèles complètement observés, trouver des fonctions de score peut être difficile et prendre du temps. Ces difficultés se multiplient lorsque nous essayons de trouver l'information (observée) de Fisher. Le paquet R Template Model Builder combine la dérivation automatique à l'approximation de Laplace pour créer un modèle général dans un contexte de maximum de vraisemblance.

MARIE AUGER-METHE (Dalhousie University), **IAN D. JONSEN** (Macquarie University), **CHRISTOFFER ALBERTSEN** (Technical University of Denmark), **GLENN T. CROSSIN** (Dalhousie University), **ANDREW E. DE-ROCHER** (University of Alberta), **KATHARINE STUDHOLME & JOANNA MILLS FLEMMING** (Dalhousie University)

[Wednesday June 17 / mercredi 17 juin, 10:50]

Tackling the Challenges of Fitting Movement Models to Marine Data

Relever les défis de l'ajustement des modèles de déplacement des données maritimes

Ecologists often investigate animal behavior by applying statistical models to movement data. Using such an approach with marine data is challenging. For example, accurate positioning systems are not well suited to the marine environment and models must account for the effect of measurement error. In addition, the ocean is not static and models should disentangle ocean drift from the volun-

Les écologistes ont souvent étudié le comportement animal en appliquant des modèles statistiques aux données de déplacement. L'utilisation d'une telle approche avec des données maritimes constitue un défi. Par exemple, les systèmes de positionnement de précision ne sont pas bien adaptés au milieu marin et les modèles doivent tenir compte des effets des erreurs de mesure. De plus, l'océan n'est pas statique et les modèles devraient séparer les dérives océa-

3B-I5: Using TMB to Quickly and Robustly Solve Problems from Marine Ecology Comment utiliser TMB pour résoudre des problèmes en écologie marine de manière rapide et robuste

tary movement of animals. Accounting for these aspects is essential when modeling the behavior of marine animals, but results in complex models. Fitting these models to large datasets often requires colossal amounts of computational power and may give inaccurate results. We show how using Template Model Builder reduces computational time and increases the robustness of both state and parameter estimates.

niques des déplacements volontaires des animaux. La prise en compte de ces aspects est essentielle lorsqu'on modélise le comportement des animaux marins, mais cela engendre des modèles complexes. L'ajustement de ces modèles à un grand ensemble de données nécessite une puissance de calcul colossale et risque de donner des résultats imprécis. Nous montrons comment l'utilisation du modélisateur Template Model Builder permet de réduire le temps de calcul et d'augmenter la robustesse des estimations d'états et de paramètres.

KIM WHORISKEY (Dalhousie University), **CHRISTOFFER ALBERTSEN** (Technical University of Denmark), **DAVID YURKOWSKI** (University of Windsor), **ANDERS NIELSEN** (Technical University of Denmark), **JOANNA MILLS FLEMMING** (Dalhousie University)

[Wednesday June 17 / mercredi 17 juin, 11:20]

Fast Fitting of Robust State-Space Models to Animal Movement Data via Template Model Builder

Ajustement rapide des modèles à espace d'états robustes pour des données de déplacement d'animaux par le modélisateur Template Model Builder

Reliable location estimates are essential to the understanding of marine animal movement. State-space models (SSMs) can estimate locations as well as parameters describing stochastic animal movement processes from error prone data, like that collected by the Argos satellite-based system. Here we recommend using the novel R-package:

Template Model Builder (TMB) for the fast fitting of robust SSMs to Argos data. We demonstrate that TMB produces similar results to Markov Chain Monte Carlo (MCMC) and Kalman filter methods commonly used for SSM inference while also allowing the estimation of additional parameters, all without requiring a substantial increase in computational time.

Des estimations de localisation fiables sont essentielles pour comprendre les déplacements des animaux marins. Les modèles à espace d'états peuvent estimer les emplacements ainsi que les paramètres décrivant les processus stochastiques des déplacements des animaux à partir de données sujettes à l'erreur, comme celles recueillies par le système satellitaire Argos.

Nous recommandons d'utiliser le nouveau paquet R Template Model Builder (TMB), à savoir le modélisateur permettant d'ajuster rapidement des modèles robustes à espace d'état aux données d'Argos. Nous démontrons que TMB produit des résultats similaires à ceux des méthodes de Monte-Carlo par chaînes de Markov (MCMC) et du filtre de Kalman souvent utilisées pour l'inférence des modèles à espace d'états. Nous mettons en évidence que TMB permet également d'estimer des paramètres supplémentaires sans trop augmenter le temps de calcul.

3B-C1: Big Data and Networks Données volumineuses et réseaux

Chair/Président: Thierry Duchesne (Université Laval)

Room/Salle: McCain 2198

Abstracts/Résumés

YILEI WU, YINGLI QIN, YILEI WU & MU ZHU (University of Waterloo)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Linear and Quadratic Discriminant Analyses for High-Dimensional Data

Analyses discriminantes linéaire et quadratique pour des données de grande dimension

High-dimensional classification is an important and challenging statistical problem. We consider both linear and quadratic discriminant rules which do not require sparsity assumptions --- either on the covariance matrices of each class (or their inverses), or on the standardized between-class distance. Under moderate conditions on the eigenvalues of population covariance matrices, our rules enjoy good asymptotic properties. Computationally, they are easy to implement and do not require large-scale mathematical programming. Numerically, they perform well in finite dimensions and with finite sample sizes. We also present real-data analyses of several classical micro-array data sets.

Le classement de grande dimension est un problème statistique important et difficile. Nous considérons les règles discriminantes linéaire et quadratique qui ne nécessitent pas d'hypothèses de parcimonie --- soit sur les matrices de covariance de chaque classe (ou leur inverse) ou sur la distance normalisée entre les classes. Sous des conditions modérées des valeurs propres des matrices de covariance de la population, nos règles bénéficient de bonnes propriétés asymptotiques. Informatiquement, elles sont faciles à mettre en œuvre et ne nécessitent pas de programmation mathématique à grande échelle. Numériquement, elles se comportent bien en dimension finie et avec des tailles d'échantillons finies. Nous présentons également des analyses de données réelles sur plusieurs jeux de données classiques de micro-réseaux.

YINGLI QIN (University of Waterloo), **WEIMING LI** (Beijing University of Posts and Telecommunications)

[Wednesday June 17 / mercredi 17 juin, 10:35]

Testing the Order of a Population Spectral Distribution for High-Dimensional Data

Vérification de l'ordre d'une distribution spectrale de la population pour des données de grande dimension

Large covariance matrices play a fundamental role in high-dimensional statistics. Investigating the behavior of their eigenvalues can reveal informative structures of large covariance matrices. In this paper, we propose to test the number of distinct population eigenvalues, i.e. the order of the Population Spectral Distribution (PSD). The proposed statistic is based upon a series of bias-reduced estimators of PSD moments. We develop the limiting distributions of our test statistic and the moment estimators. We also prove the $(n; p)$ strong consistency of these estimators, which is clearly demonstrated in our simulation study.

Les grandes matrices de covariance jouent un rôle fondamental dans la statistique de grande dimension. L'étude du comportement de leurs valeurs propres peut révéler les structures informatives de ces grandes matrices. Dans cette présentation, nous cherchons à vérifier le nombre de valeurs propres distinctes de la population, c.-à-d. l'ordre de la distribution spectrale de la population (DSP). La statistique que nous proposons est fondée sur une série d'estimateurs à biais réduit des moments de la DSP. Nous étudions les distributions limites de notre statistique de test et les estimateurs des moments. Nous prouvons également la convergence forte $(n; p)$ de ces estimateurs, qui ressort clairement dans notre étude de simulation.

3B-C1: Big Data and Networks Données volumineuses et réseaux

JABED TOMAL (University of Toronto Scarborough), **WILLIAM WELCH & RUBEN H. ZAMAR** (University of British Columbia)

[Wednesday June 17 / mercredi 17 juin, 11:05]

Ranking Homologous Proteins using a Fast Ensemble of Phalanxes Based on Logistic Regression Model

Classement de protéines homologues à l'aide d'un ensemble rapide de phalanges fondé sur un modèle de régression logistique

We have developed a predictive ensemble to rank rare-class homologous proteins in unseen test data by grouping predictor variables into subsets (we call phalanxes) using training data. The ranking in the test data is obtained by averaging the rare-class probabilities across random forests each build on a phalanx of variables. This method is further improved using a logistic regression model, and is applied in protein homology data downloaded from the KDD Cup website. This improved ensemble is fast, and the results are comparable to the winning procedures in the 2004 KDD Cup competition.

Nous avons mis au point un ensemble prédictif pour classer des protéines homologues de classe rare dans les données de test non observées en regroupant les variables prédictives en sous-ensembles (que nous appelons phalanges) à l'aide de données de formation. Le classement des données de test s'obtient en calculant la moyenne des probabilités des classes rares sur des forêts d'arbres décisionnels construits sur une phalange de variables. Nous améliorons cette méthode à l'aide d'un modèle de régression logistique, avant de l'appliquer à des données sur l'homologie des protéines tirées du site Web de la Coupe KDD. L'ensemble amélioré est rapide et les résultats sont comparables aux procédures qui ont remporté le concours de la Coupe KDD 2004.

AVIK HALDER (Queen's University and b3Intelligence), **GLEN TAKAHARA** (Queen's University)

[Wednesday June 17 / mercredi 17 juin, 11:20]

Social Media Analytics using Multistate Intensity Models

Analytique des médias sociaux à l'aide de modèles d'intensité multiétats

Social media analytics companies create social media performance score for their clients. One use for these scores to determine the posting characteristics (e.g., type and frequency) that contribute to higher scores. Based on monthly scores for some car manufacturing companies, we show analysis results based on a recent parametric multistate intensity model (R package msm, Jackson, 2014) as well as a recent nonparametric Bayesian multistate intensity model due to Kim, James and Weissbach (2011), using some novel computation techniques. We also demonstrate how to use a predictive model to improve a company's performance in social media.

Les entreprises spécialisées en analytique des médias sociaux créent des scores de performance relatifs aux médias sociaux pour leurs clients. Ces scores permettent notamment de déterminer les caractéristiques des publications (p. ex. le type et la fréquence) qui favorisent des scores plus élevés. À partir de scores mensuels de certains fabricants automobiles, nous présentons des résultats d'analyse obtenus à l'aide de nouvelles techniques informatiques et basés sur un récent modèle paramétrique multiétats d'intensité (progiciel R msm, Jackson, 2014), ainsi que sur le récent modèle bayésien non paramétrique multiétats d'intensité de Kim, James et Weissbach (2011). Nous montrons aussi comment utiliser un modèle prédictif pour améliorer la performance d'une entreprise sur les réseaux sociaux.

3B-C2: Biostatistics in Action 2 Biostatistique en action 2

Chair/Présidente: Geneviève Lefebvre (Université du Québec à Montréal)

Room/Salle: McCain 2102

Abstracts/Résumés

JOEL DUBIN & JOON LEE (University of Waterloo)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Challenges in Predicting Outcomes for Patients in Intensive Care Units (ICU's)

Défis lors de la prédiction de résultats pour les patients dans les unités de soins intensifs (USI)

A primary goal for ICU patients is treating them to achieve positive patient outcomes (e.g., hospital discharge alive, extended survival). A major analytical issue is the preponderance of information available at ICU entry (e.g., age, sex, co-morbidities, prescriptions, vital signs), and longitudinally (e.g., vital sign changes, dynamic renal function, in-ICU treatment). One approach to patient treatment is identifying similar past patients to view their treatment and eventual outcomes. We discuss ways to identify similar patients and weight them in analyses. We present simulated (and, with time, real) ICU data results, with the goal of accurately predicting outcomes for ICU patients.

Le premier objectif pour les patients en USI est de les traiter afin d'obtenir des réponses positives pour le patient (p. ex. congé de l'hôpital vivant, survie prolongée). Un problème analytique majeur est la prépondérance d'information disponible à l'arrivée à l'USI (p. ex. âge, sexe, comorbidités, prescriptions, signes vitaux) et longitudinalement (p. ex. changement de signes vitaux, fonction rénale dynamique, traitement aux SI). Une approche au traitement de patients est d'identifier des patients antérieurs similaires afin de voir leur traitement et leurs réponses potentielles. Nous discutons de manières d'identifier des patients semblables et d'en tenir compte dans les analyses. Nous présentons des données d'USI simulées (et éventuellement, réelles), avec l'objectif de pouvoir prédire avec succès le pronostic pour les patients aux USI.

BRIAN YOUNHO HONG (University of Ottawa), **ELLEN MCDONALD, GARY FOSTER, MAXIME RHÉAUME, NADER KHALIDI, MAGGIE LARCHE & LEHANA THABANE** (McMaster University)

[Wednesday June 17 / mercredi 17 juin, 10:35]

Exploring the Potential Preventative Effect of Immunomodulators on Digital Ulcer Formation for Scleroderma Patients: A Retrospective Cohort Study

Analyse de l'effet préventif potentiel des immunomodulateurs sur la formation d'ulcères digitaux chez les patients atteints de sclérodémie : étude de cohorte rétrospective

As part of the skin involvement in scleroderma, digital ulcers present a major cause of pain and functional impairment. Using the prospective data accumulated by the Canadian Scleroderma Research Group, this study aims to investigate whether immunomodulators can prevent the development of digital ulcers in patients with scleroderma. Analyses included multivariable logistic regression and generalized estimating equations and sensitivity analyses included propensity score analysis and

Au chapitre de l'atteinte cutanée causée par la sclérodémie, les ulcères digitaux constituent une cause importante de douleur et de déficience fonctionnelle. À partir des données prospectives recueillies par le Groupe de recherche canadien sur la sclérodémie, cette étude vise à déterminer si les immunomodulateurs peuvent prévenir le développement d'ulcères digitaux chez les patients atteints de sclérodémie. Les analyses effectuées comprennent une régression logistique multivariée et des équations d'estimation généralisées, et les analyses de sensibilité réalisées incluent l'analyse des

3B-C2: Biostatistics in Action 2

Biostatistique en action 2

Markov Chain Monte Carlo multiple imputation (MI). Our findings show that immunomodulators had no significant effect on preventing digital ulcers. Results of propensity score and MI sensitivity analyses did not change the findings.

scores de propension et l'imputation multiple par chaînes de Markov Monte-Carlo. Nos résultats suggèrent que les immunomodulateurs n'ont pas d'effet significatif sur la prévention des ulcères digitaux. Les résultats des analyses de sensibilité sur les scores de propension et l'imputation multiple n'ont pas modifié ces résultats.

BEI JIANG (Columbia University and New York University), **EVA PETKOVA** (New York University), **THADDEUS TARPEY** (Wright State University), **TODD OGDEN** (Columbia University)

[Wednesday June 17 / mercredi 17 juin, 10:50]

Modeling Placebo Response using EEG Data

Modélisation de la réponse placebo à l'aide de données EEG

There is growing evidence that individual differences among depression patients on Electrophysiology (EEG), fMRI and other brain imaging measurements may be predictive of potential treatment response. In this talk we discuss approaches to identifying potential placebo responders, i.e., a subgroup who benefits sufficiently from inactive drug treatments, using EEG measurements as a functional predictor. We propose to formulate this task as a statistical learning problem with support vector machine (SVM) classifier. Our approach will be evaluated through simulations and will be applied to data from a large clinical trial for patients with major depressive disorders.

Il y a de plus en plus de preuves que les différences individuelles entre les patients en dépression pour l'électro-physiologie (EEG), l'IRMf, et autres mesures d'imagerie cérébrale pourraient potentiellement prédire la réponse au traitement. Dans cette présentation nous discuterons d'approches afin d'identifier des répondants placebo potentiels, c'est à dire, un sous-groupe qui bénéficie suffisamment de traitements avec drogues inactives, en utilisant des mesures d'EEG comme indicateur. Nous proposons de formuler cette tâche comme un problème d'apprentissage statistique avec une classification de machine à vecteurs de support (MVS). Notre approche sera évaluée à travers des simulations et sera appliquée à des données provenant d'un important essai clinique avec des patients souffrant de troubles dépressifs majeurs.

NARGES NAZERI RAD (Lunenfeld-Tanenbaum Research Institute), **YUN-HEE CHOI** (University of Western Ontario), **KAREN A KOPCIUK** (University of Calgary), **JERRY LAWLESS** (University of Waterloo), **LAURENT BRIOLLAIS** (Lunenfeld-Tanenbaum Research Institute)

[Wednesday June 17 / mercredi 17 juin, 11:05]

Multistate Models for Screening Evaluation in Family-based Designs

Modèles multi-états pour l'évaluation du dépistage dans des plans de type familiaux

The effectiveness of existing screening procedures to reduce cancer risks in family-based designs has never been rigorously assessed. We consider the interest of multistate models (MSMs) to evaluate the impact of screening on polyps and colorectal cancer incidence in high-genetic risk families. The states in the MSMs are defined in terms of individuals' colonoscopy results, whose values are measured intermittently at colonoscopy schedules depending on the individuals' screening programs. We discuss how the estimation of transition intensities, covariate effects, and prevalence probabilities

L'efficacité des procédures de dépistage existantes pour réduire les risques de cancer dans des plans de type familiaux n'a jamais été analysée rigoureusement. Nous nous intéressons aux modèles multi-états pour analyser les répercussions du dépistage sur l'incidence des polypes et du cancer du côlon dans des familles à risques génétiques élevés. Les états dans les modèles multi-états sont définis selon les résultats de coloscopies d'individus, dont les valeurs sont mesurées de façon intermittente à des échéanciers de coloscopie en fonction des programmes de dépistages des individus. Nous discutons de la façon dont l'estimation des intensités de transition, les effets des covariables et les probabilités

3B-C2: Biostatistics in Action 2

Biostatistique en action 2

in MSMs can help assessing screening efficiency and examine the performance of the estimates under different screening scenarios.

de prévalence dans des modèles multi-états peuvent aider à analyser l'efficacité du dépistage. Nous examinons également la performance des estimateurs dans divers scénarios de dépistage.

YINGYING XU & JOEL DUBIN (University of Waterloo), **JOHN DANZIGER** (Harvard Medical School), **JOON LEE** (University of Waterloo)

[Wednesday June 17 / mercredi 17 juin, 11:20]

Predicting Recovery from Acute Kidney Injury Among Critically Ill Patients

Prédiction du rétablissement d'une insuffisance rénale aiguë chez les patients gravement malades

Acute kidney injury (AKI) in critically ill patients is associated with increased mortality rate in the intensive care unit (ICU). Accurate prediction of recovery from AKI provides opportunity for targeted treatment planning. Using patient data extracted from a public ICU database (MIMIC-II), we aim to identify markers that predict renal recovery after the occurrence of AKI in ICU. We assess model performance and perform internal model validation on this dataset. Also, we perform external model validation on a more recent cohort of ICU patients. We emphasize the methodology used, including nuances particular to developing, evaluating, and validating our prediction model.

L'insuffisance rénale aiguë chez les patients gravement malades est associée à un taux de mortalité élevé dans les services de soins intensifs. Une prédiction précise de guérison d'une insuffisance rénale aiguë offre une occasion de planifier des traitements ciblés. En utilisant des données de patients provenant d'une base de données (MIMIC-II) des services de soins intensifs, nous visons à identifier des indicateurs qui prédisent le rétablissement rénal après l'apparition d'insuffisance rénale aiguë dans des services de soins intensifs. Nous analysons le rendement du modèle et effectuons une validation interne de celui-ci sur ce jeu de données. De plus, nous effectuons une validation externe du modèle sur un groupe plus récent de patients aux services de soins intensifs. Nous mettons l'accent sur la méthodologie utilisée, y compris sur les nuances particulières au développement, à l'évaluation et à la validation de notre modèle de prédiction.

CHELSEA UGGENTI (Wilfrid Laurier University), **COLIN MCFAYDEN** (Ontario Ministry of Natural Resources), **DOUGLAS WOOLFORD** (Wilfrid Laurier University)

[Wednesday June 17 / mercredi 17 juin, 11:35]

Exploring the Impact of Restricted Fire Zones on the Risk of People-caused Forest Fires in Ontario

Explorer l'impact de zones de feu restreintes sur le risque de feux de forêts causés par les gens en Ontario

In Ontario, wildland fires are ignited by lightning and by people. People-caused forest fires occur due to industrial, recreational, and residential causes. During periods of extreme fire weather, fire bans are implemented and restricted fire zones put in place. The objective of a fire ban is to mitigate the risk of certain types of people-caused wildland fires. We analyze the relationship between fire bans and people-caused fire occurrence by performing a case-control study. To reduce the impact of other potential confounding variables, a matched design is used. The effectiveness of historical fire bans is quantified by comparing odds ratios.

En Ontario, les feux de forêt sont allumés par la foudre et les gens. Les feux de forêts allumés par les humains sont reliés à des causes industrielles, récréatives et résidentielles. Lors de périodes de météo extrême pour les feux, des interdictions de feu sont mises en place et des zones de feu restreintes sont établies. L'objectif d'une interdiction de feu est d'atténuer le risque de certains types de feux causés par les humains. Nous analysons la relation entre les interdictions de feu et l'occurrence de feux causés par les humains en réalisant une étude de cas-témoins. Afin de réduire l'impact d'autres variables confondantes potentielles, un plan avec appariement est utilisé. L'efficacité des interdictions de feu historiques est quantifiée en comparant les rapports de cotes.

3B-C3: Biostatistics: Causal Inference and Measurement Error Biostatistique : inférence causale et erreur de mesure

Chair/Président: Mohammad Ehsanul Karim (University of British Columbia)

Room/Salle: McCain 2116

Abstracts/Résumés

STEVE FERREIRA, MIREILLE SCHNITZER, AMÉLIE FORGET & LUCIE BLAIS (Université de Montréal)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Causal Inference Methodology to Evaluate the Time-Varying Effect of Asthma Medication on Pregnancy Duration
Méthodes en inférence causale afin d'évaluer l'effet variant dans le temps de certains médicaments contre l'asthme sur le temps de gestation

Uncontrolled asthma during pregnancy could have potentially adverse effects on the fetus' health. To assess the comparative impact of maternal use of different asthma controlling therapies on pregnancy duration we used a cohort of pregnant women from the linkage of provincial administrative databases (RAMQ and MED-ECHO). Since asthma control is a time-varying confounder it must be adjusted for using a causal analysis. We will address this issue by defining the causal parameter of interest through Marginal Structural Cox Models estimated using Inverse Probability of Treatment Weighting and Targeted Minimum Loss-based Estimation. I will present early results from this study.

L'asthme non-contrôlé pendant la grossesse peut avoir des effets néfastes sur la santé du fœtus. Afin d'évaluer l'effet comparatif de différentes thérapies de contrôle d'asthme maternel sur la durée gestationnelle, nous utilisons une cohorte de grossesses issues du couplage de bases de données administratives provinciales. Le contrôle de l'asthme étant un facteur de confusion variant dans le temps, doit être ajusté utilisant une analyse causale. Nous définirons le paramètre causal d'intérêt par des modèles de Cox structurels marginaux estimés par pondération par inverse de probabilité de traitement et par estimation par maximum de vraisemblance ciblée. Je présenterai des résultats préliminaires.

PABLO GONZALEZ GINESTET & ROBERT PLATT (McGill University)

[Wednesday June 17 / mercredi 17 juin, 10:35]

Bayesian Adjustment for Confounding in Bayesian Propensity Score Estimation
Ajustement bayésien pour la confusion dans l'estimation bayésienne du score de propension

I propose a Bayesian approach for Propensity Score (PS) variable selection and estimating the average causal effect as a weighted average over different PS. The approach is a two-stage modeling and is based on specifying three models: (1) the prognostic score model; (2) the PS model, and (3) the outcome model. The key to my approach is the incorporation in the second stage of an informative prior distribution on the parameter that controls the inclusion of each covariate in the PS, which is obtained in the first stage. Throughout, I use Reversible Jump MCMC algorithm for each stage.

Je propose une approches bayésienne pour la sélection de variable par le score de propension (SP) et l'estimation de l'effet causal moyen sous forme de moyenne pondérée sur différents SP. Cette approche constitue une modélisation en deux étapes et se base sur la définition de trois modèles : (1) le modèle de score pronostique ; (2) le modèle du SP, et (3) le modèle du résultat. La clé de cette approche est l'utilisation à la deuxième étape d'une loi a priori informative sur le paramètre qui contrôle l'inclusion de chaque covariable dans le SP, qui est obtenu à la première étape. L'algorithme MCMC à sauts réversibles est utilisé à chaque étape.

3B-C3: Biostatistics: Causal Inference and Measurement Error Biostatistique : inférence causale et erreur de mesure

YEYING ZHU (University of Waterloo), **DEBASHIS GHOSH** (Colorado School of Public Health), **DONNA COFFMAN** (Pennsylvania State University)

[Wednesday June 17 / mercredi 17 juin, 10:50]

Variable Selection in the Potential Outcomes Framework

Sélection de variables dans le cadre de résultats potentiels

A recent topic of much interest in causal inference is model selection. In this talk, we describe a framework in which to consider penalized regression approaches to variable selection for causal effects. Analogies and links with the literature on machine learning methods, missing data and imputation are drawn. A difference LASSO algorithm is defined, along with its multiple imputation analogues. The procedures are illustrated using a well-known right heart catheterization dataset.

La sélection du modèle se veut un sujet récent auquel beaucoup d'intérêt est porté en analyse causale. Dans cet exposé, nous décrivons un cadre dans lequel les approches de régression pénalisée sont considérées pour la sélection des variables des effets de causalité. Nous établissons des analogies et des liens avec la littérature sur les méthodes d'apprentissage automatique, les données manquantes et l'imputation. Nous présentons un algorithme LASSO de différence et ses parallèles au niveau de l'imputation multiple. Les procédures sont illustrées à l'aide d'un jeu de données bien connu sur le cathétérisme cardiaque de la cavité droite.

MICHELLE XIA (Northern Illinois University), **PAUL GUSTAFSON** (University of British Columbia)

[Wednesday June 17 / mercredi 17 juin, 11:05]

Bayesian Inference for Unidirectional Misclassification in Regression Models

Inférence bayésienne pour la classification erronée unidirectionnelle dans les modèles de régression

We consider unidirectional misclassification, meaning that the direction of error is known. Well-known examples of unidirectional misclassification include misrepresentation in self-reported measures. We investigate the identifiability of Bayesian regression models when either the binary covariate or response is subject to unidirectional misclassification. We consider whether the knowledge on the direction of error suffices, so that adjustment for misclassification can be undertaken without any source of information on the magnitude of the error. While measurement error models are generally nonidentified, for the case of unidirectional misclassification, we can obtain model identifiability (partial identification) when the response variable is non-binary (binary).

Nous considérons la classification erronée unidirectionnelle, ce qui signifie que la direction de l'erreur est connue. Des exemples bien connus de classification erronée unidirectionnelle comprennent de fausses déclarations de mesures autodéclarées. Nous étudions l'identifiabilité des modèles bayésiens de régression lorsque la covariable binaire ou la réponse est soumise à une classification erronée unidirectionnelle. Nous considérons que le fait de connaître la direction de l'erreur suffit de sorte que l'ajustement pour la classification erronée peut être entrepris sans aucune source d'information sur l'ampleur de l'erreur. Les modèles d'erreur de mesure sont généralement non identifiés, mais dans le cas d'erreurs de classification unidirectionnelles, nous pouvons obtenir l'identifiabilité du modèle (identification partielle) lorsque la variable réponse n'est pas binaire (binaire).

KHOKAN SIKDAR & HUDE QUAN (University of Calgary)

[Wednesday June 17 / mercredi 17 juin, 11:20]

Estimating the Prevalence of Hypertension from Administrative Data in the Absence of a Gold Standard

Estimer la prévalence de l'hypertension à partir de données administratives en l'absence d'un étalon de référence

3B-C3: Biostatistics: Causal Inference and Measurement Error Biostatistique : inférence causale et erreur de mesure

The study proposes a case ascertainment algorithm for surveillance of hypertension in children and youth in Alberta. First, multiple case ascertainment algorithms were used and compared to estimate the prevalence of hypertension across the province, assessing for regional differences and the effects of demographic factors. Second, a Bayesian latent class regression model was developed to assess the performance of the algorithm, when there is no perfect reference for a gold standard. The real data included all patients with hypertension aged 20 years or below in years 1994/95 to 2009/10; Population-based administrative data were used to identify patients diagnosed with hypertension.

Cette étude propose un algorithme de détermination des cas pour la surveillance de l'hypertension chez les enfants et les jeunes en Alberta. D'abord, des algorithmes de détermination de cas multiples ont été utilisés et comparés afin d'estimer la prévalence de l'hypertension dans la province, afin d'évaluer les différences régionales et les effets de facteurs démographiques. Ensuite, un modèle bayésien de régression de classe latente a été développé afin d'évaluer la performance de l'algorithme lorsqu'il n'y a pas de valeur parfaite comme étalon de référence. Les données réelles comprenaient celles de patients avec l'hypertension âgés de 20 ans ou moins entre les années 1994/95 et 2009/10 ; des données administratives basées sur des populations ont été utilisées pour identifier des patients diagnostiqués avec l'hypertension.

ROJIAR HADDADIAN, YULIYA MARTSYNYUK & LIQUN WANG (University of Manitoba)

[Wednesday June 17 / mercredi 17 juin, 11:35]

Simulation-based Estimation in Generalized Linear Models with a Binary Response Variable and Mismeasured Covariates

Estimation basée sur la simulation dans les modèles linéaires généralisés avec une variable réponse binaire et des covariables mesurées incorrectement

The method of moments estimation provides feasible alternatives to the likelihood approach when the likelihood function involves multiple integrals which do not have closed forms. We present a simulation-based method-of-moments approach for constructing estimators for unknown parameters of generalized linear models with a binary response variable and mismeasured covariates. We prove consistency and asymptotically normality of the proposed estimators under some regularity conditions on the distribution of the unobserved covariates and error components. We also derive a model specification test. Our approach is illustrated through simulation studies and data from Aids Clinical Trial Group (ACTG175).

L'estimation de la méthode des moments offre des substituts acceptables à l'approche de la vraisemblance lorsque la fonction de vraisemblance inclut de multiples intégrales sans forme fermée. Nous présentons une approche de méthode des moments fondée sur la simulation qui permet de construire des estimateurs pour les paramètres inconnus des modèles linéaires généralisés avec une variable réponse binaire et des covariables mesurées incorrectement. Nous prouvons la cohérence et la normalité asymptotique des estimateurs proposés, sous certaines conditions de régularité, sur la distribution des covariables non observées et des composantes d'erreur. Nous dérivons aussi un test de spécification du modèle. Nous illustrons notre approche par des études en simulation et les données du Aids Clinical Trial Group ACTG175.

3B-C4: Time Series and Longitudinal Analysis Séries chronologiques et analyse longitudinale

Chair/Présidente: Eleanor Pullenayegum (University of Toronto & Hospital for Sick Children)

Room/Salle: McCain 2162

Abstracts/Résumés

JOSHUA POHLKAMP-HARTT & DAVID L RIEGERT (Queen's University)

[Wednesday June 17 / mercredi 17 juin, 10:20]

Tackling Data Synthesis Using a Multitaper Spectrum Estimation Technique

S'attaquer à la synthèse de données à l'aide d'une technique d'estimation spectrale à fenêtres de pondération multiples

Whether you are interested in interpolation or prediction of a time series, synthesizing data can be a difficult task. This is especially true under suboptimal conditions. Our method of modeling the periodic elements of a process using Multitaper Spectrum Estimation is no different. Using re-sampling and cross validation to improve the performance of our model, we demonstrate an effective synthesis method.

Qu'il soit question d'interpolation ou de prévision de séries chronologiques, la synthèse de données peut s'avérer difficile, particulièrement dans des conditions sous-optimales. Notre méthode de modélisation des éléments périodiques d'un processus par l'estimation spectrale à fenêtres de pondération multiples ne fait pas exception. Nous proposons une méthode de synthèse efficace à l'aide d'un rééchantillonnage et d'une validation croisée pour améliorer la performance de notre modèle.

NADARAJAH THARSHANNA, ASOKAN MULAYATH VARIYATH & J CONCEPCIÓN LOREDO-OSTI (Memorial University of NL)

[Wednesday June 17 / mercredi 17 juin, 10:35]

Variable Selection for Longitudinal Data Analysis

Sélection de variables pour l'analyse de données longitudinales

High-dimensional longitudinal data with large number of covariates, have become increasingly common in many biological applications. The joint likelihood function for longitudinal data is challenging, particularly for correlated discrete outcome data. In such a situation, we propose penalized empirical likelihood based on generalized estimating equations (GEE). The proposed approach only requires specifying the first two marginal moments and a correlation structure. Simulation studies show that when model assumptions are true, its performance is comparable to that of the existing methods and when the model is misspecified, our method has clear advantages over the existing methods.

Les données longitudinales de grande dimension présentant un grand nombre de covariables sont de plus en plus fréquentes en biologie. Or il est difficile de calculer la fonction de vraisemblance conjointe pour les données longitudinales, notamment pour les données discrètes corrélées sur les résultats. Dans cette situation, nous proposons une vraisemblance empirique pénalisée fondée sur des équations d'estimation généralisées (GEE). Dans cette méthode, on peut se contenter de spécifier les deux premiers moments marginaux et une structure de corrélation. Des études de simulation montrent que lorsque les hypothèses du modèle sont exactes, la performance de celui-ci est comparable à celle des méthodes existantes et que même en cas d'erreur de spécification du modèle, notre méthode présente des avantages évidents par rapport aux méthodes existantes.

3B-C4: Time Series and Longitudinal Analysis Séries chronologiques et analyse longitudinale

SHAHEDUL KHAN (University of Saskatchewan)

[Wednesday June 17 / mercredi 17 juin, 10:50]

Modeling a Mixture of Linear and Change-point Trajectories for Longitudinal Time-Series Data

Modélisation d'un mélange de trajectoires linéaires et à point de changement pour des données longitudinales de séries chronologiques

Longitudinal change-point data arise in many applications. Examples include transition of core body temperature following the hypothermia therapy and prostate-specific antigen levels following treatment. The trend change occurs due to a shock (e.g., treatment) to the system. Thus, an individual exhibiting a linear trend could be an indication of insignificant effects of the shock. One of the goals of this type of study is to investigate whether the shock is significantly associated in changing the trajectory trend. We develop bent-cable methodology accounting for trajectories exhibiting either a linear trend or a trend change characterized by gradual or abrupt transition.

Les données longitudinales à point de changement existent dans de nombreuses applications, comme lors de la transition de la température corporelle à la suite d'un traitement contre l'hypothermie ou la mesure du taux d'antigène prostatique spécifique après un traitement. L'inversion de la tendance se produit à la suite d'un choc (p. ex. un traitement) subi par le système. Ainsi, une personne présentant une tendance linéaire pourrait indiquer des effets non significatifs du choc. L'un des objectifs de ce type d'étude consiste à déterminer si le choc a un effet significatif sur la modification de la trajectoire. Nous développons une méthodologie à corde pliée qui tient compte des trajectoires présentant une tendance linéaire ou une modification de la tendance caractérisée par une transition graduelle ou abrupte.

KEXIN JI & JOEL DUBIN (University of Waterloo)

[Wednesday June 17 / mercredi 17 juin, 11:05]

A Bivariate Semiparametric Stochastic Mixed Model

Un modèle stochastique mixte semiparamétrique bivarié

The analysis of bivariate cyclic longitudinal data remains an open problem. In this talk, we propose a bivariate semiparametric stochastic mixed model for such analysis, where each univariate model is described by its own fixed effects, a periodic smooth nonparametric function for the underlying time effect, a random effect accounting for between-subject variance, and a Gaussian process driving the within-subject variance. We will briefly describe the likelihood estimation/prediction procedure of all model parameters/effects. The proposed method is applied to a simulation study and, as time permits, an analysis from a study of longitudinal levels of female hormone data.

L'analyse des données longitudinales cycliques bivariées reste un problème ouvert. Dans cet exposé, nous proposons un modèle stochastique mixte semiparamétrique bivarié pour une telle analyse, où chaque modèle univarié est décrit par ses propres effets fixes, une fonction non paramétrique lisse périodique de l'effet du temps sous-jacent, un effet aléatoire représentant la variance entre les sujets et un processus gaussien gouvernant la variance intra-sujet. Nous décrirons brièvement la procédure d'estimation/prévision de vraisemblance de tous les paramètres du modèle/effets. La méthode proposée est appliquée à une étude de simulation et, si le temps le permet, à une analyse menée à partir d'une étude des niveaux longitudinaux de données sur les hormones féminines.

WEI QIAN & J. N. K. RAO (Carleton University), **LAURA DUMITRESCU** (National School for Statistics and Information Analysis)

[Wednesday June 17 / mercredi 17 juin, 11:20]

Quadratic Inference Functions for Longitudinal Survey Data

Fonctions d'inférence quadratiques pour les données d'enquête longitudinales

3B-C4: Time Series and Longitudinal Analysis Séries chronologiques et analyse longitudinale

This paper considers marginal models for longitudinal survey data, via the quadratic inference functions (QIF) method (Qu et al. 2000). We propose weighted QIF method, and sampling weights are used to account for survey design feature. We adopt a two-phase framework under which the inference subjects to two sources of randomness: model and sampling design. We study large sample properties of the weighted QIF estimator: consistency and normality. We also use the weighted QIF to construct likelihood ratio type test and examine the limiting distribution of the test statistic. Simulation results are presented.

Key Words: Joint Randomization, Longitudinal Survey Data

Cet article analyse des modèles marginaux pour des données d'enquête longitudinales au moyen de la méthode de fonctions quadratiques d'inférence (Qu et coll. 2000). Nous proposons une méthode de fonctions d'inférence quadratiques pondérées et utilisons des poids d'échantillonnage pour tenir compte des caractéristiques du plan d'échantillonnage. Nous adoptons un cadre de travail à deux degrés dans lequel l'inférence est soumise à deux sources de hasard : le plan de modélisation et le plan d'échantillonnage. Nous étudions deux propriétés de l'estimateur des fonctions d'inférence quadratiques pondérées dans le contexte de grands échantillons : la convergence et la normalité. Nous utilisons également les fonctions d'inférence quadratiques pondérées

pour créer un test de type rapport de vraisemblance et pour examiner la loi limite des statistiques tests. Nous présentons les résultats des simulations. Mots clés : Répartition aléatoire conjointe, données d'enquête longitudinale

YUKUN ZHANG & TOLULOPE SAJOBI (University of Calgary), **LISA LIX** (University of Manitoba)

[Wednesday June 17 / mercredi 17 juin, 11:35]

On Covariance Misspecification in Repeated Measures Discriminant Analysis

À propos de la mauvaise spécification de la covariance dans l'analyse discriminante de mesures répétées

Repeated measures discriminant analysis (RMDA) models have been developed for predicting group membership in multivariate repeated measures data. But these procedures are sensitive to departures from the multivariate normality assumption. This study investigates the effects of covariance structure misspecification on the predictive performance of the RMDA procedures based on Kronecker product covariance structures in non-normal repeated measures data using Monte Carlo techniques. The RMDA procedures exhibited inflated predictive accuracies when the covariance structures are misspecified under normally distributed data but becomes attenuated in non-normal repeated measures data. Some relevant guidelines for choosing among these proposed procedures are discussed.

Les modèles d'analyse discriminante pour les mesures répétées (ADMR) ont été développés pour prédire l'appartenance à un groupe pour les données multivariées de mesures répétées. Mais ces procédures sont sensibles aux écarts par rapport à l'hypothèse de normalité multivariée. Cette étude examine les effets de la structure de mauvaise spécification de la covariance sur la performance prédictive des procédures ADMR en se basant sur les structures de covariance de produits de Kronecker sur des données non-normales de mesures répétées à l'aide des techniques de Monte-Carlo. Les procédures ADMR ont exposé des exactitudes prédictives gonflées lorsque les structures de covariance sont mal spécifiées sous l'hypothèse de données distribuées normalement, mais elles s'atténuent pour des données non-normales de mesures répétées. Nous discutons certaines lignes directrices pertinentes pour choisir parmi ces procédures proposées.

3C-A1: Isobel Loutit Invited Address Allocution invitée Isobel Loutit

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

Session sponsored by the Business and Industrial Statistics Section

Séance parrainée par le Groupe de statistique industrielle et de gestion

Room/Salle: Rowe 1028

Abstract/Résumé

PAOLO GIUDICI (University of Pavia)

[Wednesday June 17 / mercredi 17 juin, 13:30]

Twitter Data Analysis in Business and Industry

Analyse des données de Twitter dans le commerce et l'industrie

Twitter data may be very useful to forecast business performance and, in particular, to predict intangible assets, such as company reputation. While Twitter data are becoming widespread, methods aimed at assessing their reliability are not yet available. Furthermore, there is little research on using Twitter data to measure the reputation of a company. To overcome the first problem, we propose an effective statistical method that embeds the well known h-index in the context of operational risk measurement. We then provide a methodology aimed at categorising and classifying twitter data for the measurement of the financial reputation of a company. Our proposal will be tested on a large database of tweets that refers to a group of industrial and financial companies.

Les données de Twitter peuvent être très utiles pour prédire la performance d'une entreprise et, en particulier, pour prédire des actifs intangibles tels que la réputation d'une entreprise. Bien que les données de Twitter deviennent de plus en plus répandues, les méthodes visant à évaluer leur fiabilité ne sont pas encore disponibles. De plus, peu de recherche a été menée sur l'utilisation des données de Twitter pour mesurer la réputation d'une entreprise. Pour surmonter le premier problème, nous proposons une méthode statistique efficace qui intègre l'indice h bien connu dans le contexte de la mesure du risque opérationnel. Nous fournissons ensuite une méthodologie visant à catégoriser et à classifier les données de Twitter pour mesurer la réputation financière d'une entreprise. Notre proposition sera testée sur une grande base de données de gazouillis d'un groupe d'entreprises industrielles et financières.

3C-A2: Survey Methods Section Presidential Invited Address
Allocution de l'invité du président du Groupe des méthodes d'enquête

Organizer and Chair / Responsable et président: Wesley Yung (Statistics Canada / Statistique Canada)

Session sponsored by the Survey Methods Section
Séance parrainée par le Groupe de méthodes d'enquête

Room/Salle: Rowe 1011

Abstract/Résumé

MARY THOMPSON (University of Waterloo)

[Wednesday June 17 / mercredi 17 juin, 13:30]

Big Data, Official Statistics and Survey Science

Données volumineuses, statistiques officielles et science des bases de données

Following a brief discussion of each of the terms in the title, the talk will focus on the aim of timely and accurate descriptions of survey or census populations in the context of new data sources, and the evolving relationship between information and insight. Three issues will be discussed in some detail: (i) describing emerging markets such as the market for Vaporized Nicotine Products; (ii) the promise and perils of population registries as frames; (iii) using network structures in population sampling.

À la suite d'une brève discussion sur chaque terme du titre, cet exposé mettra l'accent sur l'objectif des descriptions à jour et précises de populations d'enquête ou de recensement dans le contexte de nouvelles sources de données, et sur la relation en évolution entre les renseignements et les connaissances. Trois sujets seront abordés en détail : (i) la description des marchés émergents comme ceux des produits de nicotine vaporisée ; (ii) les promesses et les dangers des registres de population comme cadres ; (iii) l'utilisation de structures en réseau dans l'échantillonnage de la population.

3C-A3: CJS Award Address
Allocution du récipiendaire du Prix de la RCS

Organizer and Chair / Responsable et président: Richard Lockhart (Simon Fraser University)

Room/Salle: McCain - Ondaatje

Abstract/Résumé

DOUGLAS SCHAUBEL (University of Michigan), **HUI ZHANG** (U.S. Food and Drug Administration), **JOHN D. KALBFLEISCH & XU SHU** (University of Michigan)

[Wednesday June 17 / mercredi 17 juin, 13:30]

Semiparametric Methods for Survival Analysis of Case-Control Data Subject to Dependent Censoring

Méthodes semi-paramétriques pour les analyses de survie de données cas-témoins sujettes à la censure dépendante

Case-control sampling can be an efficient and cost-saving study design. It is well-established that Cox regression can be applied to case-control data. However, available estimation techniques assume that failure times are independently censored. Since independent censoring is often violated in observational studies, we propose methods for the analysis of survival data obtained through case-control sampling, but subject to dependent censoring. The proposed methods are based on weighted estimating equations, with separate inverse weights used to account for the case-control sampling and dependent censoring. Asymptotic properties are developed, with finite sample properties examined through simulation. The methods are illustrated through an analysis of pre-transplant mortality among end-stage liver disease patients obtained from a national organ failure registry.

L'échantillonnage cas-témoin peut être une méthodologie efficace et économique. Il est bien établi que la régression de Cox peut s'appliquer à des données cas-témoins. Cependant, les techniques d'estimation existantes supposent que les temps de défaillance sont censurés de façon indépendante. Puisque la censure indépendante est souvent bafouée dans le cadre d'études observationnelles, nous proposons des méthodes d'analyse de données de survie obtenues par un échantillonnage cas-témoin, mais qui sont sujettes à la censure dépendante. Les méthodes proposées sont fondées sur des équations d'estimation pondérées dont des poids inverses séparés permettent de tenir compte de l'échantillonnage cas-témoin et de la censure dépendante. Des propriétés asymptotiques sont développées avec des propriétés d'échantillons de taille finie étudiées par simulation. Les méthodes sont illustrées au moyen d'une analyse sur le taux de mortalité pré-greffe chez les patients atteints d'une maladie hépatique en phase terminale provenant d'un registre national de défaillances d'organes.

3C-A4: Pierre Robillard Award Address
Allocution du récipiendaire du Prix Pierre-Robillard

Chair/Président: Alexandre Leblanc (University of Manitoba)
Organizer/Responsable: Éric Marchand (Université de Sherbrooke)

Room/Salle: McCain - Scotiabank

Abstract/Résumé

YING YAN (University of North Carolina)

[Wednesday June 17 / mercredi 17 juin, 13:30]

Statistical Methods on Survival Data with Measurement Error

Méthodes statistiques sur des données de survie avec erreur de mesure

Covariate measurement error has attracted extensive interests in survival analysis. In this talk, we systematically investigate measurement error problems under additive hazards models. New insights of measurement error effects are revealed, as opposed to well-documented results for proportional hazards models. In particular, we explore asymptotic bias of ignoring measurement error in the analysis. To correct for the induced bias, we develop a class of functional correction methods for error effects. The validity of the proposed methods is carefully examined, and we investigate issues of model checking and model misspecification. Theoretical results are rigorously established, and are complemented with numerical assessments and a real data analysis.

L'erreur de mesure des covariables a suscité un grand intérêt en l'analyse de survie. Dans cet exposé, nous examinons systématiquement les problèmes d'erreur de mesure dans le contexte de modèles de risques additifs. De nouvelles perspectives sur les effets liés aux erreurs de mesure sont révélées comparativement à des résultats bien documentés de modèles de risques proportionnels. Nous explorons notamment le biais asymptotique causé par la non-considération de l'erreur de mesure dans l'analyse. Pour corriger ce biais, nous avons développé une classe de méthodes de correction fonctionnelles pour les effets d'erreur. Nous étudions attentivement la validité des méthodes proposées ainsi que les problèmes de vérification et d'erreur de spécification de modèles. Les résultats théoriques sont rigoureusement établis et sont complétés par des évaluations numériques et une analyse de données réelles.

3D-I1: Applications of Mixed-Effect Models in the Health Sciences Applications des modèles à effets mixtes en sciences de la santé

Chair/Président: Patrick Brown (University of Toronto)
Organizer/Responsable: Mahmoud Torabi (University of Manitoba)

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

Room/Salle: Rowe 1011

Abstracts/Résumés

SUNIL RAO (University of Miami), **JIMING JIANG** (University of California, Davis), **JIE FAN** (University of Miami), **THUAN NGUYEN** (Oregon Health and Science University)

[Wednesday June 17 / mercredi 17 juin, 15:00]

Classified Mixed Model Prediction

Prévision classifiée à l'aide d'un modèle mixte

Mixed model prediction (MMP) has a fairly long history starting with Henderson's early work in animal breeding (Henderson 1948). Nowadays, new and challenging problems have emerged, to which methods of MMP are potentially applicable, but not without further methodological and computational developments. Often problems occur when interest is at subject level (e.g., personalized medicine), or (small) sub-population level (e.g., community, center). The challenges have to do with better prediction of a mixed effect, or a future observation, by identifying the class that a new subject belongs to. We propose a new method, called classified mixed model prediction (CMMP), to solve this problem. We develop a theory for CMMP and investigate its empirical performance through simulation studies and a real-data application.

L'origine de la prévision à l'aide de modèles mixtes remonte aux premiers travaux d'Henderson sur la reproduction animale (Henderson, 1948). De nos jours, de nouveaux défis ont émergé pour lesquels les méthodes de prévision à l'aide de modèles mixtes peuvent s'appliquer, mais qui nécessitent des progrès méthodologiques et informatiques. Les problèmes surgissent souvent lorsque l'intérêt se situe au niveau du sujet (p. ex. en médecine personnalisée) ou de la sous-population (p. ex. la collectivité, un centre). Les défis concernent l'amélioration de la prévision d'un effet mixte ou d'une observation future par l'identification de la classe à laquelle le sujet appartient. Nous proposons une nouvelle méthode appelée prévision classifiée à l'aide d'un modèle mixte pour résoudre ce problème. Nous élaborons une théorie pour cette méthode et examinons sa performance empirique au moyen d'études de simulations et d'une application sur des données réelles.

CINDY FENG (University of Saskatchewan)

[Wednesday June 17 / mercredi 17 juin, 15:30]

Zero Modified Models for Modeling Length of Hospital Stay for patients with Ischaemic Heart Disease

Modèles modifiés à zéro pour la modélisation de la durée des séjours hospitaliers des patients atteints de cardiopathie ischémique

Length of stay (LOS) in hospital is often used as an indicator of hospital efficiency and a proxy of resource consumption, which may be characterized as zero-inflated if there is an over-abundance of ze-

La durée des séjours hospitaliers est souvent utilisée comme indicateur de l'efficacité des hôpitaux et comme approximation de l'utilisation des ressources qui peut être caractérisée par une surreprésentation de zéros, s'il y en a plus que

3D-I1: Applications of Mixed-Effect Models in the Health Sciences Applications des modèles à effets mixtes en sciences de la santé

roes, or zero-deflated if there are fewer zeroes than expected under a standard count model. Such data may also have a highly right-skewed distribution for the nonzero values. We developed and compared a series of zero modified models with various configurations of fixed and random effects, as well as allowing analysis of nonlinear effects of time, spatially structured variation and unstructured heterogeneity. Modeling and inference use the fully Bayesian approach via Markov Chain Monte Carlo (MCMC) simulation techniques.

prévu, ou une sous-représentation, s'il y en a moins que prévu, selon un modèle de décompte standard. Les valeurs non nulles parmi ces données peuvent être fortement asymétriques vers la droite. Nous élaborons une série de modèles modifiés à zéro et les comparons à différentes configurations d'effets fixes ou aléatoires, en plus d'effectuer l'analyse des effets non linéaires du temps, de la variation structurée dans l'espace et de l'hétérogénéité non structurée. La modélisation et l'inférence se fondent entièrement sur l'approche bayésienne selon des techniques de simulation par chaînes de Markov Monte-Carlo.

MAHMOUD TORABI (University of Manitoba)

[Wednesday June 17 / mercredi 17 juin, 16:00]

Multivariate Mixture Spatial Generalized Linear Mixed Models

Modèles de mélange spatiaux multivariés de la classe des modèles linéaires mixtes généralisés

Disease mapping has been widely studied with considering only one disease in the estimated models. Simultaneous modeling of related diseases can also be a valuable tool both from the epidemiological and from the statistical point of view. In particular, when we have several measurements recorded at each spatial location, we need to consider multivariate models in order to handle the dependence among the multivariate components as well as the spatial dependence between locations. In many circumstances, it is a very strong assumption to have the same underlying distribution for all regions of population study. In this talk, we consider the multivariate mixture spatial models for areal data for Normal and non-Normal responses in the class of generalized linear mixed models.

La cartographie des maladies a fait l'objet de nombreuses études dont les modèles estimés tiennent compte d'une seule maladie. La modélisation simultanée de maladies apparentées peut aussi constituer un outil précieux des points de vue épidémiologique et statistique. En présence de plusieurs mesures à chaque emplacement, il faut notamment considérer les modèles multivariés pour gérer la dépendance parmi les composantes multivariées et la dépendance spatiale entre les emplacements. L'obtention de la même distribution sous-jacente pour toutes les régions de l'étude populationnelle constitue souvent une hypothèse très forte. Dans cet exposé, nous étudions les modèles de mélange spatiaux multivariés de la classe des modèles linéaires mixtes généralisés pour des données géospatiales de réponses normales et non normales.

3D-I2: Limit Theorems With Applications Théorèmes limites et applications

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

Session sponsored by the Probability Section

Séance parrainée par le Groupe de probabilité

Room/Salle: McCain 2198

Abstracts/Résumés

MIKLOS CSORGO (Carleton University), **ENDRE CSAKI** (Alfred Renyi Institute of Mathematics), **ANTONIA FOLDES** (College of Staten Island, CUNY), **PAL REVESZ** (Technische Universitat Wien)

[Wednesday June 17 / mercredi 17 juin, 15:00]

Two-dimensional Anisotropic Random Walks: Fixed versus Random Column Configurations for Transport Phenomena

Marches aléatoires anisotropes à deux dimensions: configurations fixes et aléatoires des colonnes pour les phénomènes de transport

Anisotropic random walks on the square lattice of the plane could serve as models for the study of transport phenomena of statistical physics. Two-dimensional anisotropic random walks with asymptotic density conditions as in Heyde (J. Statist. Physics 27 (1982), 721-730, and Stochastic Processes. Springer, New York, pp. 125-130) yield fixed column configurations, and nearest-neighbour random walks in a random environment on the square lattice of the plane as in den Hollander (J. Statist. Physics 75 (1994), 891-918) result in random column configurations. We will survey related results in a historical context and, in both cases, also conclude simultaneous weak Donsker and strong Strassen type invariance principles in terms of appropriately constructed anisotropic Brownian motions on the plane.

Les marches aléatoires anisotropes sur le réseau carré d'un plan peuvent servir de modèles pour l'étude des phénomènes de transport en physique statistique. Les marches aléatoires anisotropes à deux dimensions comportant des conditions de densité asymptotique telles que présentées par Heyde (J. Statist. Physics 27 (1982), p. 721-730 et Stochastic Processes. Springer, New York, p.125-130) produisent des configurations de colonnes fixes, alors que les marches aléatoires des plus proches voisins dans un environnement aléatoire sur le réseau carré d'un plan telles que présentées par den Hollander (J. Statist. Physics 75 (1994), p. 891-918) produisent des configurations de colonnes aléatoires. Nous étudions les résultats connexes dans un contexte historique et concluons dans les deux cas des principes simultanés d'invariance faible de type Donsker et forte de type Strassen en termes de mouvements browniens anisotropes construits de façon appropriée sur le plan.

DELI LI (Lakehead University), **HAN-YING LIANG** (Tongji University, China)

[Wednesday June 17 / mercredi 17 juin, 15:30]

A Comparison Theorem for the Law of Large Numbers

Théorème de comparaison pour la loi des grands nombres

In this work we establish what we call a comparison theorem for the law of large numbers for i.i.d. \mathbf{B} -valued random variables. To establish this comparison theorem for the law of large numbers, we invoke two tools: 1) a comparison theorem for sums

Dans ce travail, nous établissons ce que nous appelons un théorème de comparaison pour la loi des grands nombres pour les variables aléatoires i.i.d. de valeur \mathbf{B} . Pour y arriver, nous utilisons deux outils : 1) un théorème de comparaison pour la somme des variables aléatoires indépen-

3D-I2: Limit Theorems With Applications Théorèmes limites et applications

of independent \mathbf{B} -valued random variables and, 2) a symmetrization procedure for the law of large numbers for sums of independent \mathbf{B} -valued random variables. A few consequences of our main results are provided.

dantes de valeur \mathbf{B} et 2) une procédure de symétrisation de la loi des grands nombres pour les sommes des variables aléatoires indépendantes de valeur \mathbf{B} . Nous présentons certaines conséquences de nos résultats principaux.

PRISCILLA GREENWOOD & LAWRENCE WARD (University of British Columbia)

[Wednesday June 17 / mercredi 17 juin, 16:00]

Stochastic Neural Dynamics

Modèle stochastique de dynamique neuronale

Probability problems abound in the study of stochastic models of neurons and networks of neurons. For example, the inter-spike interval distribution of most stochastic single neuron models has been studied only by simulation. Similarly, the behavior of stochastic networks of neurons is little understood, results coming primarily from simulation. This talk will introduce probabilists and statisticians to several inviting research problems in the dynamic area of neuroscience.

Les problèmes de probabilité abondent dans l'étude des modèles stochastiques de neurones et de réseaux de neurones. Dans la plupart des modèles stochastiques à neurone unique, par exemple, la distribution des intervalles entre les impulsions est étudiée seulement par simulation. De plus, le comportement des réseaux stochastiques de neurones est encore mal compris et les résultats proviennent essentiellement de simulations. Cet exposé présente aux probabilistes et aux statisticiens des problèmes de recherche stimulants dans le domaine dynamique de la neuroscience.

3D-I3: Maritime Ecology Écologie marine

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

Room/Salle: McCain - Scotiabank

Abstracts/Résumés

CHRIS FIELD, HONG GU & XIMING XU (Dalhousie University)

[Wednesday June 17 / mercredi 17 juin, 15:00]

Modelling Microbial Multinomial Time or Space Series

Modélisation de séries multinomiales microbiennes chronologiques ou spatiales

This talk is motivated by issues arising from some microbial oceanic data which typically consists of counts of OTU's (operational taxonomic units) from ocean samples varying either in space or time. We can think of the data as a multinomial series where the number of cells is large (typically 1000 or more). I will describe how I approach the problem of finding an appropriate model and analysis. The proposed analysis consists of two parts, a dimension reduction via clustering of the OTU's followed by a modelling of the dynamics of the clustered OTU's over time and/or space. I will illustrate with data from Bedford Basin and also from a European experiment of the effects of carbon loading on the OTU's.

Cet exposé est motivé par des questions qui découlent de certaines données océaniques microbiennes qui sont normalement composées de nombres d'unités taxonomiques opérationnelles provenant d'échantillons océaniques variant soit dans l'espace, soit dans le temps. Nous pouvons penser aux données comme des séries multinomiales dans lesquelles le nombre de cellules est élevé (typiquement 1 000 ou plus). Je décrirai la façon dont j'approche le problème pour trouver un modèle et une analyse appropriés. L'analyse proposée consiste en deux parties : une réduction de la dimension par le regroupement des unités taxonomiques opérationnelles et la modélisation de la dynamique de regroupement des unités taxonomiques opérationnelles dans le temps et/ou l'espace. J'illustrerai l'analyse par des données provenant du Bassin de Bedford et d'une expérience européenne des effets des charges de carbone sur les unités taxonomiques opérationnelles.

MICHAEL DOWD (Dalhousie University), **PAUL MATTERN** (University of California), **JON BRIGGS** (University of Auckland), **KATJA FENNEL** (Dalhousie University), **RENATE MEYER** (University of Auckland)

[Wednesday June 17 / mercredi 17 juin, 15:30]

Data Assimilation for Ocean Biology

Assimilation des données pour la biologie océanique

New observing technologies and improvements in numerical ocean models are furthering our scientific understanding of ocean biology, and making marine environmental prediction systems possible. A major challenge is identifying and implementing advanced statistical approaches that can successfully combine information from the nonlinear dynamic biological models with the non-Gaussian observations. Here, I explore the general statistical data assimilation problem, and specific appli-

De nouvelles technologies et améliorations d'observation dans les modèles océaniques numériques permettent à nos scientifiques d'approfondir leur compréhension de la biologie océanique, et au système de prévision environnementale marine d'exister. Un gros défi consiste à déterminer et à mettre en œuvre des approches statistiques avancées qui peuvent combiner avec succès l'information issue de modèles biologiques dynamiques non linéaires à des observations non gaussiennes. Dans cet article, j'explore le problème d'assimilation des données statistiques générales, et

3D-I3: Maritime Ecology Écologie marine

cations to the lower trophic level marine planktonic ecosystem (ocean biogeochemistry). State space models provide a unifying framework for current efforts, and use sampling based approaches for state and parameter estimation. It is argued that fully Bayesian approaches are the target for comprehensive statistical data assimilation and, in this context, challenges and potential new directions are discussed.

les applications spécifiques à un niveau trophique inférieur d'écosystème du plancton marin (biogéochimie océanique). Les modèles d'espace d'états fournissent un cadre de travail unificateur pour les efforts actuels, et utilisent des approches basées sur l'échantillonnage pour l'estimation des états et des paramètres. Certains affirment que les approches complètement bayésiennes sont préférables pour l'assimilation de données statistiques générale, et dans ce contexte, nous discutons des défis et des orientations potentielles futures.

KEITH THOMPSON & YIHAO YIN (Dalhousie University), **EMMANUEL DEVRED** (Université Laval), **FLORENCE BIROL** (Laboratoire d'Etudes en Géophysique et Océanographie Spatiales)

[Wednesday June 17 / mercredi 17 juin, 16:00]

Satellite Sensing of the Marine Environment and the Issue of Aliasing

La détection par satellite du milieu marin et le problème de repliement de spectre

Satellite-borne sensors have revolutionized observation of the marine environment and they now provide snapshots of a wide range of physical properties on a routine basis. A major problem is the satellite's orbital period often exceeds the dominant time scales of variability, leading to aliasing. I will illustrate the problem using sea level measured every 9.9 days by altimeters moving along a fixed track crossing from the deep ocean to the coast. I will describe the standard techniques for de-aliasing (i.e., using output from tide and surge models) and show how the results can be improved using hourly sea level observations from coastal tide gauges. Models are fit in the time domain (using lasso based regression) and frequency domain.

Les détecteurs satellitaires ont révolutionné l'observation du milieu marin : désormais, ils fournissent systématiquement des instantanés qui ont de nombreuses propriétés physiques. La période orbitale du satellite qui dépasse souvent les échelles de temps de variabilité dominantes constituent un gros problème, ce qui amène à un repliement de spectre. J'illustrerai ce problème avec le niveau de la mer mesuré tous les 9,9 jours par des altimètres qui bougent le long d'une trajectoire déterminée qui va des eaux profondes à la côte. Je décrirai les techniques courantes pour éliminer le repliement de spectre (avec les résultats provenant des modèles de marée et de houle). Je montrerai comment on peut améliorer les résultats en utilisant des observations horaires du niveau de l'eau issues des marégraphes le long de la côte. Des modèles sont adaptés dans les domaines temporel (au moyen de la régression LASSO) et fréquentiel.

3D-I4: R Packages for Teaching Utiliser R pour l'enseignement

Chair/Présidente: Bethany White (Western University)
Organizer/Responsable: Duncan Murdoch (Western University)

Session sponsored by the Statistics Education Section
Séance parrainée par le Groupe d'éducation en statistique

Room/Salle: McCain - Ondaatje

Abstracts/Résumés

JOHN FOX (McMaster University)

[Wednesday June 17 / mercredi 17 juin, 15:00]

Teaching Social Statistics Courses with R and the R Commander

Enseignement de cours de statistique sociale avec R et R Commander

In statistics courses for social-science students -- and in other application-area disciplines that use statistical methods -- teaching statistical computing isn't an end in itself. Statistical software should support the fundamental goals of the course, which differ by the level of the course. Use of statistical software should therefore differ as well. I focus on two typical courses: an introductory statistics course, and a second course that emphasises applied regression analysis. Both courses use R software, the first via the R Commander graphical user interface and the second via the RStudio interactive programming environment. I discuss how these software choices support course goals, and demonstrate the use of the R Commander for a typical homework exercise from the first course.

Dans les cours de statistique pour les étudiants en science sociale -- et dans d'autres disciplines-domaines d'application utilisant les méthodes statistiques -- l'enseignement de l'informatique statistique n'est pas une fin en soi. Le logiciel statistique devrait soutenir les objectifs fondamentaux du cours, qui diffèrent selon le niveau du cours. L'utilisation de logiciels statistiques devrait donc également être différente. Je me concentre sur deux cours typiques : un cours d'introduction à la statistique et un deuxième cours qui met l'accent sur l'analyse de régression appliquée. Les deux cours utilisent le logiciel R, le premier par l'entremise de l'interface graphique R Commander et le second par l'entremise de l'environnement de programmation interactif RStudio. Je discute de la manière dont le choix de ces logiciels soutient les objectifs du cours et démontre l'utilisation de R Commander dans un exercice typique d'un devoir du premier cours.

DUNCAN MURDOCH (Western University)

[Wednesday June 17 / mercredi 17 juin, 15:35]

R in Teaching and Testing

R dans l'enseignement et les examens

In this presentation I will describe two R packages: rgl, used for 3D demonstrations of mathematical and statistical concepts, and Sweavetest, used for preparing and analyzing examinations.

Dans cette présentation, je vais décrire deux paquets R : rgl, utilisé pour démonstrations en trois dimensions de concepts mathématiques et statistiques, et Sweavetest, utilisé pour la préparation et l'analyse des examens.

3D-I5: Statistical Application and Modeling for Functional Data Analysis Application et modélisation statistiques pour l'analyse de données fonctionnelles

Organizer and Chair / Responsable et président: Haocheng Li (University of Calgary)

Room/Salle: Rowe 1028

Abstracts/Résumés

JAMES RAMSAY (McGill University)

[Wednesday June 17 / mercredi 17 juin, 15:00]

Data Smoothing with Tuneable Penalties

Lissage de données avec pénalités ajustables

If the linear differential operator D^2 is replaced by a more general differential operator L that annihilates most of the variation in the observed curves, then the resulting smooth has less bias and greatly reduced mean squared error. This talk will show how we can use the data to estimate such an operator L . The differential equation estimated in this way represents the dynamics of the processes being estimated. It also emerges that there are exciting new ways of representing the data by the use of an orthogonal basis system defined by the estimated operator. These results will be illustrated through the analysis of some handwriting data.

Si l'opérateur différentiel linéaire D^2 est remplacé par un opérateur différentiel plus général L qui supprime la majorité de la variation dans les courbes observées, alors le lissé résultant présente moins de biais et une erreur quadratique moyenne nettement réduite. Nous montrons comment utiliser les données pour estimer un tel opérateur L . L'équation différentielle ainsi estimée représente la dynamique des processus estimés. Il apparaît également qu'il y a de nouvelles façons intéressantes de représenter les données en utilisant un système de bases orthogonales défini par l'opérateur estimé. Nous illustrons ces résultats par l'analyse de données manuscrites.

MARK KOUDSTAAL & FANG YAO (University of Toronto)

[Wednesday June 17 / mercredi 17 juin, 15:30]

Simultaneous White Noise Models and Optimal Recovery of Functional Data

Modèles de bruit de fond simultanés et récupération optimale de données fonctionnelles

We consider the white noise representation of functional data taken as i.i.d. realizations of a Gaussian process. We begin with establishing an asymptotic equivalence in Le Cam's sense between an experiment which simultaneously describes these realizations and a collection of white noise models. In this context, we project onto an arbitrary basis and apply a novel variant of Stein-type estimation for optimal recovery of the realized trajectories. A key inequality is derived showing that the corresponding risks, conditioned on the underlying curves, are optimal and can be made arbitrarily close to those that an oracle with knowledge of the process would attain. Empirical performance is illustrated through simulated and real data examples.

Nous étudions la représentation du bruit de fond des données fonctionnelles considérées comme des réalisations i.i.d d'un processus gaussien. Nous commençons par établir une équivalence asymptotique au sens de Le Cam entre une expérience qui décrit ces réalisations simultanément et une collection de modèles de bruit de fond. Dans ce contexte, nous opérons une projection sur une base arbitraire et appliquons une nouvelle variante de l'estimation de type Stein pour une récupération optimale des trajectoires réalisées. Nous dérivons une inégalité clé qui montre que les risques correspondants, conditionnés aux courbes sous-jacentes, sont optimaux et qu'ils peuvent être arbitrairement rapprochés de ceux d'un oracle ayant connaissance du processus. Nous illustrons la performance empirique de notre méthode avec des exemples de données réelles et de simulation.

3D-I5: Statistical Application and Modeling for Functional Data Analysis

Application et modélisation statistiques pour l'analyse de données fonctionnelles

JASON NIELSEN & NIKOLINA BEG (Carleton University)

[Wednesday June 17 / mercredi 17 juin, 16:00]

Model-Based Penalty Smoothing for Panel Data Under a Markov Assumption

Lissage avec pénalité basée sur un modèle pour données de panel dans une hypothèse de Markov

A flexible semiparametric model for analyzing multi-state panel count data is presented. Panel data here refers to observing the state of individuals or objects under observation at a discrete set of time points. It is assumed that the data is being generated by a continuous-time, discrete state space Markov process. The transition intensities are represented with model-based penalty smoothing. This provides a class of models ranging between a fully nonparametric and parametric assumption for the transition intensities. Estimation and inference of the proposed approach is outlined and the finite sample properties of the methodology investigated. Finally, the method is illustrated on some health data.

Nous présentons un modèle semi-paramétrique souple permettant l'analyse de données de comptage de panel multi-états. On entend ici par données de panel l'observation de l'état d'individus ou d'objets à des moments discrets. On présume que les données sont générées par un processus de Markov d'espace d'états discrets en temps continu. Les intensités des transitions sont représentées par un lissage avec pénalité basée modèle. Cela produit une classe de modèles allant de l'hypothèse entièrement non paramétrique à l'hypothèse paramétrique pour les intensités de transitions. Nous étudions l'estimation et l'inférence dans l'approche proposée et les propriétés de population finie de la méthodologie. Enfin, nous illustrons la méthode avec des données de santé.

Author List • Liste des auteurs

- Abeysundera, Melanie, 57, 220 (3B-I4)
 Acar, Elif, 39, 135 (2B-I2)
 Achath, Sudhakar, 26, 87 (1B-C4)
 Aeberhard, William H., 25, 83 (1B-C3)
 Aghababazadeh, Farnoosh Abbas, 40, 146 (2B-C2)
 Ahmad, Fatimah Al, 35, 123 (1E-C2)
 Ahmed, S. Ejaz, 49, 188 (2D-C3)
 Albert-Green, Alisha, 31, 102 (1D-C2)
 Albertsen, Christoffer, 57, 222, 223 (3B-I5)
 Allen, O. Brian, 44, 161 (2C-P)
 Alsubie, Abdelaziz, 52, 205 (2E-C2)
 Alvo, Mayer, 40 (2B-C2), 46 (2D-I3), 146 (2B-C2), 174 (2D-I3)
 Amezziane, Mohamed, 26 (1B-C3), 49 (2D-C3), 83 (1B-C3), 188 (2D-C3)
 An, Baiguo, 46, 172 (2D-I2)
 Anand, Sonia, 41, 149 (2B-C3)
 Andrews, Jeffrey, 50, 196 (2E-I4)
 Ané, Cécile, 33, 115 (1E-I4)
 Apaloo, Joseph, 49, 189 (2D-C3)
 Arguin, Louis-Pierre, 34, 118 (1E-I5)
 Asgharian, Masoud, 40, 143 (2B-C1)
 Auger-Methe, Marie, 57, 222 (3B-I5)
 Augustyniak, Maciej, 23, 70 (1B-I3)
 Awosoga, Olu, 43 (2C-P), 49 (2D-C2), 154 (2C-P), 186 (2D-C2)
 Aziz, Imoudden, 41, 146 (2B-C2)
- Bahamyirou, Asma, 27 (1C-D1), 32, 107 (1D-C4)
 Baillargeon, Sophie, 38, 135 (2B-I2)
 Bally, Michèle, 47, 176 (2D-I4)
 Bandyopadhyay, Rohosen, 29, 94 (1D-I4)
 Baral, Janarthan, 27 (1C-D1)
 Barnett, Ian, 24, 74 (1B-I5)
 Baskurt, Zeynep, 56, 218 (3B-I3)
 Bathe, Oliver F., 43, 157 (2C-P)
 Baudouin, Antoine, 27 (1C-D1)
 Beck, Jennifer, 57, 220 (3B-I4)
- Bédard, Mylène, 48, 185 (2D-C2)
 Beg, Nikolina, 63, 246 (3D-I5)
 Bégin, Jean-François, 45, 169 (2D-I1)
 Belalia, Mohamed, 32, 109 (1D-C4)
 Belius, David, 34, 118 (1E-I5)
 Bellhouse, David, 56, 217 (3B-I2)
 Benoit, Hugues, 34, 119 (1E-C1)
 Bergeron, Pierre-Jérôme, 40, 143 (2B-C1)
 Beyene, Joseph, 41, 149 (2B-C3)
 Bi, Yong-Mei, 44, 163 (2C-P)
 Bickel, David R., 40, 146 (2B-C2)
 Bickis, Mikelis, 32, 108 (1D-C4)
 Bingham, Derek, 25 (1B-C2), 36 (1E-C4), 79 (1B-C2), 129 (1E-C4)
 Birol, Florence, 63, 244 (3D-I3)
 Blais, Lucie, 59, 229 (3B-C3)
 Blanchin, Myriam, 29, 92 (1D-I3)
 Bolker, Benjamin, 44, 158 (2C-P)
 Bond, Marjorie, 20, 23, 68 (1B-I2)
 Bordet, Charles, 36, 131 (1E-C4)
 Bouchard-Côté, Alexandre, 24 (1B-C1), 33 (1E-I4), 47 (2D-I4), 48 (2D-C1), 76 (1B-C1), 115 (1E-I4), 177 (2D-I4), 183 (2D-C1)
 Boudreault, Mathieu, 23 (1B-I3), 45 (2D-I1), 70 (1B-I3), 169 (2D-I1)
 Bouezmarni, Taoufik, 32, 107, 109 (1D-C4)
 Braun, John, 31 (1D-C2), 47 (2D-I5), 102 (1D-C2), 178, 179 (2D-I5)
 Briggs, Jon, 40 (2B-I5), 63 (3D-I3), 141 (2B-I5), 243 (3D-I3)
 Briollais, Laurent, 50 (2E-I3), 58 (3B-C2), 195 (2E-I3), 227 (3B-C2)
 Brobbey, Anita, 36, 130 (1E-C4)
 Brown, Lawrence D., 28, 88 (1D-I1)
 Brown, Patrick, 39, 137 (2B-I3)
 Bull, Shelley B., 52, 204 (2E-C2)
 Butler, Michael, 51, 196 (2E-I4)

- Cabilio, Paul, 46, 174 (2D-I3)
 Cadigan, Noel, 34 (1E-C1), 36 (1E-C3), 119 (1E-C1),
 127 (1E-C3)
 Cai, Song, 48, 182 (2D-C1)
 Cai, Yun, 24, 77 (1B-C1)
 Calderhead, Ben, 40, 141 (2B-I5)
 Campbell, David, 40 (2B-I5), 48 (2D-C1), 141 (2B-I5),
 182 (2D-C1)
 Cantoni, Eva, 25, 83 (1B-C3)
 Canty, Angelo J., 52, 204 (2E-C2)
 Carey, Michelle, 51, 200 (2E-C1)
 Carrier, Félix, 27 (1C-D1)
 Carriere, Keumhee Chough, 40, 144 (2B-C1)
 Carroll, Raymond, 35 (1E-C1), 56 (3B-I2), 121 (1E-C1),
 216 (3B-I2)
 Carson, Stuart, 35, 126 (1E-C3)
 Casquilho-Resende, Camila, 25, 80 (1B-C2)
 Chai, Hao, 32, 110 (1E-I1)
 Chakraborty, Adrijo, 29, 94 (1D-I4)
 Chalifour, Mathieu, 50, 196 (2E-I4)
 Chaudhuri, Sanjay, 29, 94 (1D-I4)
 Che, Menglu, 27 (1C-D1)
 Chen, Bingshu, 40, 144 (2B-C1)
 Chen, Bo, 24 (1B-C1), 27 (1C-D1), 77 (1B-C1)
 Chen, Entao, 43, 154 (2C-P)
 Chen, Fuqi, 49, 188 (2D-C3)
 Chen, Hao (Nelson), 43, 155 (2C-P)
 Chen, Jiahua, 38, 132 (2A-A)
 Chen, Min, 39, 139 (2B-I4)
 Chen, Susan, 33, 113 (1E-I3)
 Chenouri, Shojaeddin, 53, 206, 207 (2E-C3)
 Cheung, Siu Hung, 52, 205 (2E-C2)
 Chipman, Hugh, 35, 123 (1E-C2)
 Chkrebti, Oxana, 40, 141 (2B-I5)
 Choi, Yun-Hee, 58, 227 (3B-C2)
 Christou, Nicolas, 51, 198 (2E-I5)
 Chu, Jingjia, 43, 155 (2C-P)
 Chuang, Junior, 28 (1C-D2), 30, 101 (1D-C2)
 Chuangpishit, Hoda, 22, 66 (1B-I1)
 Ciampi, Antonio, 50, 194 (2E-I3)
 Cigsar, Candemir, 35, 124 (1E-C2)
 Claveau, Jeannine, 57, 220 (3B-I4)
 Coffman, Donna, 59, 229 (3B-C3)
 Cook, Richard, 31 (1D-C3), 33 (1E-I2), 40 (2B-C1), 104,
 105 (1D-C3), 112 (1E-I2), 143 (2B-C1)
 Cooper-Barfoot, Tricia, 28, 90 (1D-I2)
 Corkum, Amber, 46, 174 (2D-I3)
 Cossette, H el ene, 30 (1D-C1), 45 (2D-I1), 98 (1D-C1),
 169 (2D-I1)
 Cotton, Cecilia, 31, 104 (1D-C3)
 Cox, Danielle, 23, 68 (1B-I2)
 Cox, David, 29, 96 (1D-I5)
 Craiu, Radu V., 53, 209 (2E-C4)
 Crawford, Forrest, 34, 115 (1E-I4)
 Cribben, Ivor, 24, 76 (1B-C1)
 Crossin, Glenn T., 57, 222 (3B-I5)
 Csaki, Endre, 62, 241 (3D-I2)
 Csorgo, Miklos, 49 (2D-C3), 62 (3D-I2), 189 (2D-C3),
 241 (3D-I2)
 Dai, David, 28 (1C-D2)
 Dai, Wei Hao (David), 30, 31, 101, 102 (1D-C2)
 Daignault, Katherine, 28 (1C-D2)
 Dang, Sanjeena, 44, 163 (2C-P)
 Danziger, John, 58, 228 (3B-C2)
 Darlington, Gerarda, 52, 203 (2E-C2)
 Datta, Gauri, 29, 94 (1D-I4)
 Dawe, Earl, 36, 127 (1E-C3)
 Dean, Charmaine, 24 (1B-I5), 31 (1D-C2), 35 (1E-C1),
 44 (2C-P), 75 (1B-I5), 102 (1D-C2), 121 (1E-
 C1), 159 (2C-P)
 Deardon, Rob, 39, 138 (2B-I3)
 Dehling, Herold, 46, 174 (2D-I3)
 de Leon, Alex, 45, 168 (2D-I1)
 Deng, Bo, 40, 144 (2B-C1)
 Derkach, Andriy, 56, 218 (3B-I3)
 Derocher, Andrew E., 57, 222 (3B-I5)
 Desgagn e, Alain, 48, 185 (2D-C2)
 de Tibeiro, Jules Joseph S., 53, 207 (2E-C3)
 Devred, Emmanuel, 63, 244 (3D-I3)
 DeVries, Sara, 49, 188 (2D-C3)
 Dewanji, Anup, 31, 106 (1D-C3)
 Diao, Liqun, 24 (1B-I4), 31 (1D-C3), 72 (1B-I4), 104 (1D-
 C3)
 Dinov, Ivo, 51, 198 (2E-I5)
 Dinsdale, Daniel, 36, 127 (1E-C3)
 Dixon, Stephanie, 52, 203 (2E-C2)
 Do, Kim-Anh, 41, 151 (2B-C3)
 Dong, Chi, 49, 189 (2D-C3)
 Doucet, Arnaud, 47, 177 (2D-I4)
 Dowd, Michael, 40 (2B-I5), 63 (3D-I3), 141 (2B-I5), 243
 (3D-I3)
 Downs, Gregory, 44, 163 (2C-P)
 Duan, Xingde, 25, 82 (1B-C3)
 Duanmu, Haosui, 26, 86 (1B-C4)

- Dubin, Joel, 58 (3B-C2), 59 (3B-C4), 226, 228 (3B-C2), 233 (3B-C4)
- Duchesne, Pierre, 51, 201 (2E-C1)
- Duchesne, Thierry, 44 (2C-P), 47 (2D-I5), 160 (2C-P), 179 (2D-I5)
- Dumitrescu, Laura, 59, 233 (3B-C4)
- Dunham, Bruce, 33, 113 (1E-I3)
- Dupuis, Debbie, 50, 191 (2E-I1)
- Durieu, Olivier, 46, 174 (2D-I3)
- Edge, Alanna, 44, 159 (2C-P)
- Ejaz, Ahmed, 32, 111 (1E-I1)
- Elliott, Michael, 57, 221 (3B-I4)
- Entezari, Reihaneh, 27 (1C-D1)
- Espin-Garcia, Osvaldo, 52, 204 (2E-C2)
- Fan, Jianqing, 46, 172 (2D-I2)
- Fan, Jie, 62, 239 (3D-I1)
- Fan, Wenqi, 28 (1C-D2)
- Fan, Zhao Zhi, 31, 105 (1D-C3)
- Fang, Yizhou, 36, 129 (1E-C4)
- FaridRohani, MohammadReza, 53, 207 (2E-C3)
- Farrell, Patrick, 31 (1D-C2), 31 (1D-C3), 103 (1D-C2), 106 (1D-C3)
- Farshidfar, Farshad, 43, 157 (2C-P)
- Feng, Cindy, 44 (2C-P), 62 (3D-I1), 161 (2C-P), 239 (3D-I1)
- Feng, Ming, 46, 173 (2D-I2)
- Feng, Zeny, 44 (2C-P), 56 (3B-I3), 160 (2C-P), 219 (3B-I3)
- Fennel, Katja, 40 (2B-I5), 63 (3D-I3), 141 (2B-I5), 243 (3D-I3)
- Ferland, René, 51, 200 (2E-C1)
- Ferland-Raymond, Bastien, 25, 80 (1B-C2)
- Ferreira, Steve, 28 (1C-D2), 59, 229 (3B-C3)
- Ferreira, Tadeu, 27 (1C-D1)
- Field, Chris, 51 (2E-I4), 62 (3D-I3), 196 (2E-I4), 243 (3D-I3)
- Fienberg, Stephen, 56, 216 (3B-I2)
- Fisher, Jon, 36, 127 (1E-C3)
- Flemming, Joanna Mills, 35, 126 (1E-C3)
- Flowers, Gwenn E., 25, 79 (1B-C2)
- Foldes, Antonia, 62, 241 (3D-I2)
- Forget, Amélie, 59, 229 (3B-C3)
- Foster, Gary, 36 (1E-C4), 58 (3B-C2), 130 (1E-C4), 226 (3B-C2)
- Fox, John, 63, 245 (3D-I4)
- Fox, Karla, 57, 220 (3B-I4)
- Fraser, Don, 56 (3B-I2)
- Freue, Gabriela Cohen, 50, 194 (2E-I3)
- Fuller, Wayne A., 39, 140 (2B-I4)
- Furman, Edward, 46, 170 (2D-I1)
- Gagnon, Philippe, 27 (1C-D1), 48, 185 (2D-C2)
- Gao, Bowen, 27 (1C-D1)
- Gao, Jane, 22, 66 (1B-I1)
- Garrido, Jose, 50 (2E-I2)
- Gauthier, Geneviève, 45, 169 (2D-I1)
- Gavanji, Parisa, 40, 144 (2B-C1)
- Genest, Christian, 38 (2B-I2), 51 (2E-C1), 135 (2B-I2), 200 (2E-C1)
- Ghahramani, Melody, 26, 86 (1B-C4)
- Ghandehari, Mahya, 22, 66 (1B-I1)
- Ghosh, Debashis, 59, 229 (3B-C3)
- Ghosh, Malay, 29, 94 (1D-I4)
- Gibbs, Alison, 51, 198 (2E-I5)
- Ginestet, Pablo Gonzalez, 27 (1C-D1), 59, 229 (3B-C3)
- Girard, Claude, 25, 79 (1B-C2)
- Girolami, Mark, 40, 141 (2B-I5)
- Giudici, Paolo, 61, 235 (3C-A1)
- Gombay, Edit, 50, 192 (2E-I1)
- Graham, Jinko, 24, 78 (1B-C1)
- Gravel, Christopher A., 31, 106 (1D-C3)
- Greenlaw, Keelin, 24, 78 (1B-C1)
- Greenwood, Celia, 50, 194 (2E-I3)
- Greenwood, Priscilla, 62, 242 (3D-I2)
- Grosskopf, Michael, 36, 129 (1E-C4)
- Group, The BeAMS Study, 30, 101 (1D-C2)
- Gu, Hong, 24 (1B-C1), 41 (2B-C3), 51 (2E-I4), 62 (3D-I3), 77 (1B-C1), 150 (2B-C3), 196 (2E-I4), 243 (3D-I3)
- Guerzhoy, Michael, 36, 129 (1E-C4)
- Guilleux, Alice, 29, 92 (1D-I3)
- Guindani, Michele, 41, 151 (2B-C3)
- Gustafson, Paul, 30 (1D-C2), 59 (3B-C3), 101 (1D-C2), 230 (3B-C3)
- Gweon, Hyukjun, 43, 155 (2C-P)
- Haddadian, Rojjar, 59, 231 (3B-C3)
- Halder, Avik, 58, 225 (3B-C1)
- Hall, Peter, 47, 179 (2D-I5)
- Hamel, Martin, 25, 79 (1B-C2)
- Hamid, Jemila, 41, 149 (2B-C3)
- Hannig, Jan, 30, 96 (1D-I5)
- Harari, Ofir, 25, 79 (1B-C2)
- Hardouin, Jean-Benoit, 29, 92 (1D-I3)

- Harper, Adam, 34, 118 (1E-I5)
 Hasan, Tariqul, 34 (1E-C1), 38 (2B-I1), 119, 120 (1E-C1), 134 (2B-I1)
 Hasler, Caren, 53, 209 (2E-C4)
 He, Fang, 43, 156 (2C-P)
 He, Genming, 27 (1C-D1), 28 (1C-D2)
 He, Qiming, 55, 214 (3B-I1)
 He, Zhouqin, 51, 197 (2E-I4)
 Heffernan, Jane, 31, 102 (1D-C2)
 Henrey, Andrew, 24, 73 (1B-I4)
 Heritier, Stephane, 25, 83 (1B-C3)
 Hill, Kathleen, 44, 159 (2C-P)
 Hiraki, Linda, 52, 204 (2E-C2)
 Hockey, Daniel Evan, 48, 182 (2D-C1)
 Hong, Brian Younho, 58, 226 (3B-C2)
 Hossain, Shakhawat, 26, 86 (1B-C4)
 Hu, Joan, 47, 178 (2D-I5)
 Huang, Chao, 46, 172 (2D-I2)
 Huang, Mei Ling, 26, 83 (1B-C3)
 Huang, Shiqiong, 28, 88 (1D-I1)
 Hughes, Edward, 38, 133 (2B-I1)
 Huitema, Bradley, 49, 186 (2D-C2)
 Husain, Syed, 51, 198 (2E-I5)
 Huynh, Mirabelle, 46, 170 (2D-I1)
- Iijima, Hiroaki, 27 (1C-D1)
 Im, Jongho, 39, 140 (2B-I4)
 Indralingam, Maheshan, 27 (1C-D1)
 Iorio, Alfonso, 36, 130 (1E-C4)
 Islam, Shofiqul, 41, 149 (2B-C3)
 Iyengar, Satish, 30, 99 (1D-C1)
- Jacobbe, Tim, 33, 113 (1E-I3)
 Jang, Gun Ho, 41, 149 (2B-C3)
 Jankowski, Hanna, 31, 102 (1D-C2)
 Janssen, Jeannette, 22, 66 (1B-I1)
 Jazi, Omidali Aghababaei, 40, 143 (2B-C1)
 Jermyn, Ian H., 46, 172 (2D-I2)
 Jewell, Sean, 24, 76 (1B-C1)
 Ji, Kexin, 59, 233 (3B-C4)
 Jiang, Bei, 58, 227 (3B-C2)
 Jiang, Jiming, 29 (1D-I4), 62 (3D-I1), 94 (1D-I4), 239 (3D-I1)
 Jiang, Wenyu, 40 (2B-C1), 52 (2E-C2), 144 (2B-C1), 203 (2E-C2)
 Jiayi, Liang, 53, 206 (2E-C3)
 Jin, Jiashun, 28, 88 (1D-I1)
 Jing, Jia, 27 (1C-D1), 28 (1C-D2)
- Jiongo, Valéry Dongmo, 53, 210 (2E-C4)
 Jocelyn, Wisner, 57, 220 (3B-I4)
 Johnson, Timothy, 46, 173 (2D-I2)
 Jonsen, Ian D., 57, 222 (3B-I5)
 Juarez-Colunga, Elizabeth, 35, 121 (1E-C1)
- Kaczmirek, Lars, 43, 155 (2C-P)
 Kalbfleisch, John D., 61, 237 (3C-A3)
 Kalinin, Alexandr, 51, 198 (2E-I5)
 Kalyaniwalla, Nauzer, 22, 66 (1B-I1)
 Kan, Megan, 27 (1C-D1)
 Kang, Sohee, 42, 152 (2B-C4)
 Karim, Mohammad Ehsanul, 30, 101 (1D-C2)
 Karlis, Dimitris, 38, 133 (2B-I1)
 Karunamuni, Rohana, 48, 185 (2D-C2)
 Kazimer, Scarlett, 52, 203 (2E-C2)
 Kenney, Toby, 24 (1B-C1), 30 (1D-C1), 41 (2B-C3), 51 (2E-I4), 77 (1B-C1), 98 (1D-C1), 150 (2B-C3), 196 (2E-I4)
 Khalidi, Nader, 58, 226 (3B-C2)
 Khalili, Abbas, 28 (1D-I1), 40 (2B-C1), 89 (1D-I1), 143 (2B-C1)
 Khan, Shahedul, 59, 232 (3B-C4)
 Kiberd, James, 43--45, 156, 159, 165 (2C-P)
 Kim, Jae-Kwang, 39, 139, 140 (2B-I4)
 Kinniburgh, Susan, 33, 113 (1E-I3)
 Klar, Neil, 52, 203 (2E-C2)
 Kong, Linglong, 46, 172 (2D-I2)
 Kopciuk, Karen A, 43 (2C-P), 58 (3B-C2), 157 (2C-P), 227 (3B-C2)
 Kortbi, Othmane, 41, 146 (2B-C2)
 Koudstaal, Mark, 63, 246 (3D-I5)
 Kounta, Moussa, 51, 200 (2E-C1)
 Kouritzin, Michael, 49, 189 (2D-C3)
 Kranakis, Evangelos, 34, 117 (1E-I5)
 Krewski, Daniel, 31 (1D-C2), 31 (1D-C3), 103 (1D-C2), 106 (1D-C3)
 Kristensen, Kasper, 57, 222 (3B-I5)
 Kubokawa, Tatsuya, 29, 94 (1D-I4)
 Kulperger, Reg, 43, 44, 155, 159 (2C-P)
 Kunwar, Pratheep, 27 (1C-D1)
 Kurosawa, Takuma, 44, 157, 162 (2C-P)
 Kurtek, Sebastian, 46, 172 (2D-I2)
 Kye, Yisub, 46, 170 (2D-I1)
- Labbe, Aurélie, 41 (2B-C3), 50 (2E-I3), 150 (2B-C3), 194 (2E-I3)
 Labrecque-Synnott, Félix, 57, 220 (3B-I4)

- Lai, Chu Shing, 30, 96 (1D-I5)
 Landriault, David, 46, 170 (2D-I1)
 Laperriere, Christiane, 53, 211 (2E-C4)
 Larche, Maggie, 58, 226 (3B-C2)
 Lawless, Jerry, 31 (1D-C3), 56 (3B-I3), 58 (3B-C2), 104 (1D-C3), 218 (3B-I3), 227 (3B-C2)
 Le, Nhu D., 25, 80 (1B-C2)
 Leblanc, Alexandre, 32, 108, 109 (1D-C4)
 LeBlanc, Michael, 23, 72 (1B-I4)
 Lee, Bonnie, 43, 154 (2C-P)
 Lee, Joon, 58, 226, 228 (3B-C2)
 Lee, Thomas, 30, 96 (1D-I5)
 Lee, Wan-Chen, 44, 158 (2C-P)
 Lefebvre, Geneviève, 50 (2E-I2)
 Lefebvre, Isabelle, 28 (1C-D2)
 Lemyre, Félix Camirand, 32, 107 (1D-C4)
 Leroux, Brian, 26, 85 (1B-C4)
 Lesperance, Mary, 24, 78 (1B-C1)
 Li, Deli, 62, 241 (3D-I2)
 Li, Fuxiao, 50, 192 (2E-I1)
 Li, Haocheng, 35, 121 (1E-C1)
 Li, Hongzhe, 24, 74 (1B-I5)
 Li, Li, 41, 150 (2B-C3)
 Li, Longhai, 44, 161 (2C-P)
 Li, Michael, 44, 158 (2C-P)
 Li, Pengfei, 34, 120 (1E-C1)
 Li, Weili, 32, 107 (1D-C4)
 Li, Weiming, 57, 224 (3B-C1)
 Li, Xuan, 31, 102 (1D-C2)
 Liang, Han-Ying, 62, 241 (3D-I2)
 Liang, Kun, 41, 150 (2B-C3)
 Liao, Chendi, 27 (1C-D1), 28 (1C-D2)
 Lin, Chang-Yun, 35, 125 (1E-C2)
 Lin, Laura, 27 (1C-D1)
 Lin, Xihong, 24, 74 (1B-I5)
 Ling, Shen, 51, 196 (2E-I4)
 Liu, Hai Yan, 40, 143 (2B-C1)
 Lix, Lisa, 29 (1D-I3), 60 (3B-C4), 93 (1D-I3), 234 (3B-C4)
 Loeppky, Jason, 44, 158 (2C-P)
 Loredo-Osti, J Concepción, 26 (1B-C4), 59 (3B-C4), 85 (1B-C4), 232 (3B-C4)
 Lou, Wendy, 30, 31, 101, 102 (1D-C2)
 Loughin, Thomas, 24, 73 (1B-I4)
 Lu, Danny, 27 (1C-D1)
 Lu, TongYu, 52, 205 (2E-C2)
 Lu, Zihang, 31, 102 (1D-C2)
 Lundy, Erin, 35, 121 (1E-C1)
 Luo, Bin, 44, 159 (2C-P)
 Lysy, Martin, 25 (1B-C2), 36 (1E-C3), 36 (1E-C4), 52 (2E-C1), 81 (1B-C2), 127 (1E-C3), 129 (1E-C4), 201 (2E-C1)
 Ma, Renjun, 25 (1B-C3), 34 (1E-C1), 38 (2B-I1), 82 (1B-C3), 119, 120 (1E-C1), 133, 134 (2B-I1)
 Ma, Yanyuan, 56, 216 (3B-I2)
 Mackay, Anne, 23, 71 (1B-I3)
 MacKay, Jock, 28 (1D-I2), 35 (1E-C2), 90 (1D-I2), 123 (1E-C2)
 MacKay Altman, Rachel, 26, 85 (1B-C4)
 Madras, Neal, 22, 67 (1B-I1)
 Maity, Arnab, 56, 216 (3B-I2)
 Mandal, Abhyuday, 29, 94 (1D-I4)
 Marceau, Etienne, 30 (1D-C1), 45 (2D-I1), 98 (1D-C1), 169 (2D-I1)
 Marchand, Éric, 32 (1D-C4), 40, 41 (2B-C2), 107 (1D-C4), 146, 147 (2B-C2)
 Mariathas, Hensley Hubert, 35, 124 (1E-C2)
 Marriott, Paul, 34, 120 (1E-C1)
 Martsynuk, Yuliya, 49 (2D-C3), 59 (3B-C3), 189 (2D-C3), 231 (3B-C3)
 Mattern, Paul, 40 (2B-I5), 63 (3D-I3), 141 (2B-I5), 243 (3D-I3)
 McDonald, Ellen, 58, 226 (3B-C2)
 McFayden, Colin, 58, 228 (3B-C2)
 McIsaac, Michael, 31, 105 (1D-C3)
 McKean, Joseph, 49, 186 (2D-C2)
 McLeish, Don, 48, 181 (2D-I6)
 McNeil, Shelly, 45, 166 (2C-P)
 McNicholas, Paul, 50, 196 (2E-I4)
 McPherson, John, 41, 149 (2B-C3)
 Melnikov, Alexander, 30, 98 (1D-C1)
 Meng, Xiao-Li, 22, 65 (1A-A2)
 Metzler, Adam, 55, 214 (3B-I1)
 Meyer, Renate, 40 (2B-I5), 63 (3D-I3), 141 (2B-I5), 243 (3D-I3)
 Mian, Mohammad Rajibul Islam, 25, 82 (1B-C3)
 Millar, Michelle, 23, 69 (1B-I2)
 Mills, Shirley, 28 (1D-I2), 48 (2D-C1), 90 (1D-I2), 182 (2D-C1)
 Mills Flemming, Joanna, 35 (1E-C3), 57 (3B-I5), 126 (1E-C3), 222, 223 (3B-I5)
 Minin, Vladimir, 34, 115 (1E-I4)
 Miville, Marie-Hélène, 57, 220 (3B-I4)
 Miyaoka, Etsuo, 44, 157, 162 (2C-P)
 Molloy, Mike, 22, 66 (1B-I1)
 Moodie, Erica, 48, 183 (2D-C1)

- Moon, Nathalie, 40, 143 (2B-C1)
 Morales, Manuel, 23, 70 (1B-I3)
 Morin, Lealand, 30, 99 (1D-C1)
 Mostafaiy, Behdad, 53, 207 (2E-C3)
 Mottola, Justyne, 26, 83 (1B-C3)
 Mtalai, Itre, 30 (1D-C1), 45 (2D-I1), 98 (1D-C1), 169 (2D-I1)
 Muir, Adam, 42 (2B-C4), 43--45 (2C-P), 153 (2B-C4), 156, 159, 165 (2C-P)
 Mukherjee, Gourab, 28, 88 (1D-I1)
 Mukherjee, Partha Sarathi, 46, 172 (2D-I2)
 MULLOWNEY, Darrell, 36, 127 (1E-C3)
 Murdoch, Duncan, 43 (2C-P), 63 (3D-I4), 156 (2C-P), 245 (3D-I4)
 Murkerjee, Rajarshi, 24, 74 (1B-I5)
- Nadeem, Khurram, 43, 154 (2C-P)
 Nathoo, Farouk, 24 (1B-C1), 32 (1E-I1), 46 (2D-I2), 78 (1B-C1), 110 (1E-I1), 173 (2D-I2)
 Neish, Drew, 44, 160 (2C-P)
 Nešlehová, Johanna, 38, 135 (2B-I2)
 Newton, Michael, 48, 182 (2D-C1)
 Nguimkeu, Pierre, 53, 210 (2E-C4)
 Nguyen, Christine, 26, 83 (1B-C3)
 Nguyen, Thuan, 62, 239 (3D-I1)
 Nicosia, Aurélien, 44, 160 (2C-P)
 Nie, Zixin, 27 (1C-D1)
 Nielsen, Anders, 57, 223 (3B-I5)
 Nielsen, Jason, 63, 246 (3D-I5)
 Nkurunziza, Sévérien, 49, 188 (2D-C3)
 Nordell-Markovits, Alexei, 27 (1C-D1)
 Nsoesie, Elaine, 39, 137 (2B-I3)
- Ogden, Todd, 58, 227 (3B-C2)
 Oldford, Wayne, 41 (2B-C2), 52 (2E-C3), 147 (2B-C2), 206 (2E-C3)
 Oliver, Broderick, 57, 220 (3B-I4)
- Pannu, Yuvraj, 27 (1C-D1)
 Paterson, Andrew, 52, 204 (2E-C2)
 Paul, Sudhir, 25, 82 (1B-C3)
 Pehlivan, Lerna, 22, 67 (1B-I1)
 Pellegrini, Tiago, 34, 119 (1E-C1)
 Pencer, Matthew, 41, 148 (2B-C2)
 Peng, Jianan, 52, 205 (2E-C2)
 Petkau, John, 22 (1A-A1), 30 (1D-C2), 55 (3A-A2), 101 (1D-C2), 213 (3A-A2)
 Petkova, Eva, 58, 227 (3B-C2)
 Picka, Jeffrey, 53, 206 (2E-C3)
- Pillai, Natesh, 34 (1E-I5), 52 (2E-C1), 117 (1E-I5), 201 (2E-C1)
 Platt, Robert, 59, 229 (3B-C3)
 Pohlkamp-Hartt, Joshua, 59, 232 (3B-C4)
 Poon, Wai Yin, 52, 205 (2E-C2)
 Porter, Andrew, 44, 161 (2C-P)
 Pratola, Matthew T., 25, 79 (1B-C2)
 Pullenayegum, Eleanor, 29, 92 (1D-I3)
- Qian, Wei, 59, 233 (3B-C4)
 Qin, Lin, 36, 127 (1E-C3)
 Qin, Yingli, 57, 224 (3B-C1)
 Qin, Yongsong, 53, 209 (2E-C4)
 Qiu, Shi, 44, 161 (2C-P)
 Quan, Hude, 59, 230 (3B-C3)
 Quessy, Jean-François, 32, 107 (1D-C4)
- Rad, Narges Nazeri, 58, 227 (3B-C2)
 Rahman, Adam, 41, 147 (2B-C2)
 Rai, Kamal, 27 (1C-D1)
 Ramsay, James, 51 (2E-C1), 63 (3D-I5), 200 (2E-C1), 246 (3D-I5)
 Ranjan, Pritam, 35, 123 (1E-C2)
 Rao, J. N. K., 53 (2E-C4), 59 (3B-C4), 209 (2E-C4), 233 (3B-C4)
 Rao, Sunil, 62, 239 (3D-I1)
 Regier, Michael, 48, 183 (2D-C1)
 Reid, Nancy, 29, 96 (1D-I5)
 Remaining members of the mode effect working group, , 57, 220 (3B-I4)
 Rémillard, Bruno, 20, 38, 135 (2B-I2)
 Ren, Jiandong, 55, 214 (3B-I1)
 Revesz, Pal, 62, 241 (3D-I2)
 Rhéaume, Maxime, 58, 226 (3B-C2)
 Riegert, David L, 25 (1B-C2), 59 (3B-C4), 79 (1B-C2), 232 (3B-C4)
 Rivest, Louis-Paul, 25 (1B-C2), 36 (1E-C4), 38 (2B-I2), 44, 45 (2C-P), 80 (1B-C2), 131 (1E-C4), 135 (2B-I2), 160, 164 (2C-P)
 Romanescu, Razvan, 39, 138 (2B-I3)
 Rosenthal, Jeffrey, 26, 86 (1B-C4)
 Rosychuk, Rhonda, 44, 162 (2C-P)
 Rothstein, Steven, 44, 163 (2C-P)
 Ruan, Ji, 43 (2C-P), 45 (2D-I1), 157 (2C-P), 168 (2D-I1)
 Rusmevichientong, Paat, 28, 88 (1D-I1)
- Sabalaiuskas, Kelly, 30, 101 (1D-C2)
 Sadeghkhani, Nasser, 41, 147 (2B-C2)
 Safari, Abdollah, 26, 85 (1B-C4)

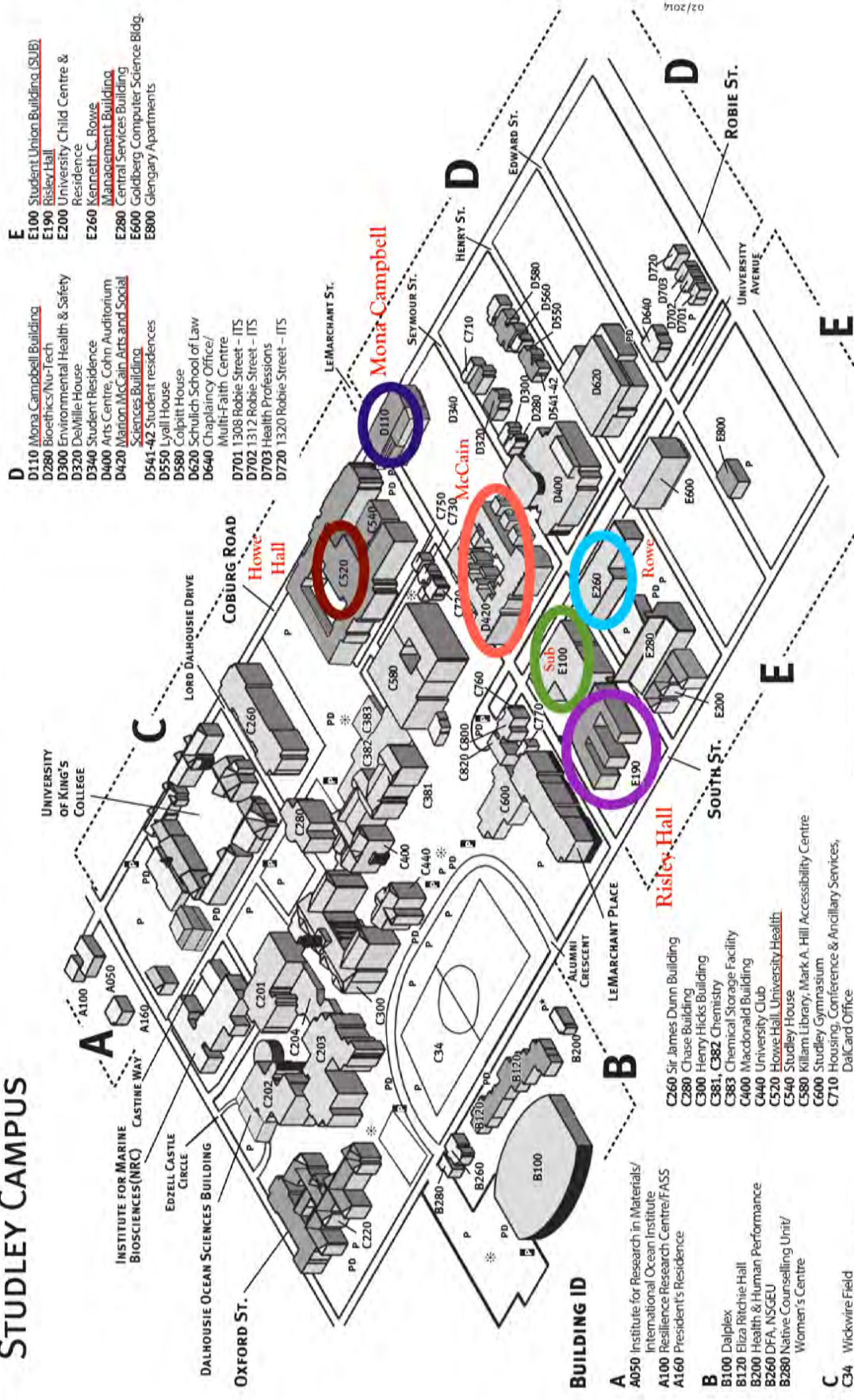
- Saha, Sudipta, 27 (1C-D1)
 Saïdi, Abdelnasser, 20
 Sajobi, Tolulope, 29 (1D-I3), 60 (3B-C4), 93 (1D-I3), 234 (3B-C4)
 Salibián-Barrera, Matías, 36 (1E-C3), 55 (3A-A1), 127 (1E-C3), 212 (3A-A1)
 Sango, Joel, 51, 201 (2E-C1)
 Sarhan, Ammar, 49, 189 (2D-C3)
 Saunders, David, 30, 99 (1D-C1)
 Schaubel, Douglas, 61, 237 (3C-A3)
 Schiopu-Kratina, Ioana, 40, 143 (2B-C1)
 Schnitzer, Mireille, 47 (2D-I4), 59 (3B-C3), 176 (2D-I4), 229 (3B-C3)
 Schonlau, Matthias, 43, 155 (2C-P)
 Sébille, Véronique, 29, 92 (1D-I3)
 Shaikhet, Gennady, 34, 117 (1E-I5)
 She, Xichen, 53, 210 (2E-C4)
 Shen, Bingqing, 28 (1C-D2)
 Shi, Jenny, 30, 96 (1D-I5)
 Shi, Tianji, 40, 142 (2B-I5)
 Shi, Tianxiang, 46, 170 (2D-I1)
 Shimokawa, Asanao, 44, 157, 162 (2C-P)
 Shokoohi, Farhad, 41, 150 (2B-C3)
 Shrier, Ian, 47, 176 (2D-I4)
 Shu, Xu, 61, 237 (3C-A3)
 Shuangge, Ma, 32, 110 (1E-I1)
 Sikdar, Khokan, 59, 230 (3B-C3)
 Silva, Anjali, 44, 163 (2C-P)
 Simoneau, Gabrielle, 27 (1C-D1), 44, 163 (2C-P)
 Sivathayalan, Ahalya, 31, 103 (1D-C2)
 Small, Christopher, 53, 206 (2E-C3)
 Smith, Aaron, 34, 117 (1E-I5)
 Smith, Bruce, 45, 166 (2C-P)
 Sneddon, Gary, 34 (1E-C1), 38 (2B-I1), 120 (1E-C1), 134 (2B-I1)
 Sobrino, Rafael, 27 (1C-D1)
 Sohrabi, Maryam, 48, 185 (2D-C2)
 Solís-Lemus, Claudia, 33, 115 (1E-I4)
 Soltanifar, Mohsen, 27 (1C-D1)
 Spencer, Neil, 48, 183 (2D-C1)
 Spiegelman, Donna, 56, 216 (3B-I2)
 Springford, Aaron, 25, 79 (1B-C2)
 Srivastava, Anuj, 46, 172 (2D-I2)
 Stafford, James, 47, 178 (2D-I5)
 Staicu, Ana-Maria, 56, 216 (3B-I2)
 Starszyk, Peter, 25, 81 (1B-C2)
 Steele, Russell, 47, 176 (2D-I4)
 Stein, Lincoln, 41, 149 (2B-C3)
 Steiner, Stefan, 28 (1D-I2), 35 (1E-C2), 50 (2E-I2), 90 (1D-I2), 123 (1E-C2)
 Steingrimsson, Jon, 24, 72 (1B-I4)
 Stephens, David, 20, 41 (2B-C3), 50 (2E-I3), 150 (2B-C3), 194 (2E-I3)
 Stephenson, Matthew, 52, 203 (2E-C2)
 Stevens, Nathaniel, 35, 123 (1E-C2)
 Stewart, Connie, 36, 128 (1E-C3)
 Steyn, Douw G., 40, 142 (2B-I5)
 Stingo, Francesco, 41, 151 (2B-C3)
 Stojkova, Biljana Jonoska, 48, 182 (2D-C1)
 Strawderman, Robert, 24, 72 (1B-I4)
 Strawderman, William, 41, 146 (2B-C2)
 Strug, Lisa, 32 (1D-C4), 56 (3B-I3), 107 (1D-C4), 218 (3B-I3)
 Studholme, Katharine, 57, 222 (3B-I5)
 Su, Wanhua, 52, 204 (2E-C2)
 Subedi, Sanjeena, 44, 160 (2C-P)
 Suchard, Marc, 34, 115 (1E-I4)
 Sugasawa, Shonosuke, 29, 94 (1D-I4)
 Sun, Lei, 56, 218 (3B-I3)
 Sunderland, Kelly, 28 (1C-D2)
 Sung, Yubin, 45, 164 (2C-P)
 Susko, Ed, 22 (1A-A1)
 Swartz, Philippa, 36, 129 (1E-C4)
 Swartz, Tim, 36, 129 (1E-C4)
 Szefer, Elena, 24, 78 (1B-C1)
 Taback, Nathan, 36 (1E-C4), 48 (2D-I6), 129 (1E-C4), 180 (2D-I6)
 Takahara, Glen, 58, 225 (3B-C1)
 Tao, Sile, 27 (1C-D1)
 Tarpey, Thaddeus, 58, 227 (3B-C2)
 Tchahou, Herbert Nkwimi, 25, 79 (1B-C2)
 Teismann, Holger, 43, 154 (2C-P)
 Tekougang, Thierry Chekouo, 41, 151 (2B-C3)
 Thabane, Lehana, 33 (1E-I2), 36 (1E-C4), 41 (2B-C3), 48 (2D-I6), 58 (3B-C2), 112 (1E-I2), 130 (1E-C4), 149 (2B-C3), 180 (2D-I6), 226 (3B-C2)
 Tharshanna, Nadarajah, 59, 232 (3B-C4)
 Thompson, Jenny, 57, 220 (3B-I4)
 Thompson, Keith, 63, 244 (3D-I3)
 Thompson, Mary, 39 (2B-I4), 61 (3C-A2), 139 (2B-I4), 236 (3C-A2)
 Thompson, Steve, 53, 209 (2E-C4)
 Thomson, David J, 25, 79 (1B-C2)
 Thomson, Trevor, 26, 86 (1B-C4)
 Thornton, Jennifer, 42, 152 (2B-C4)

- Tillé, Yves, 25, 80 (1B-C2)
Tomal, Javed, 58, 224 (3B-C1)
Tong, Jinjun, 26, 85 (1B-C4)
Torabi, Mahmoud, 62, 240 (3D-I1)
Tounkara, Fode, 45, 164 (2C-P)
Tremlett, Helen, 30, 101 (1D-C2)
Tsai, Cary, 23 (1B-I3), 45 (2D-I1), 70 (1B-I3), 168 (2D-I1)
Tsala, Eric, 27 (1C-D1)
Tsui, Lung Kwan, 30, 99 (1D-C1)
Tu, Wei, 27 (1C-D1), 48, 185 (2D-C2)
Tusche, Marco, 46, 174 (2D-I3)
- Uggenti, Chelsea, 58, 228 (3B-C2)
Umphrey, Gary, 44, 161 (2C-P)
Usman, Iram, 44, 162 (2C-P)
- Vallée, Audrey-Anne, 25, 80 (1B-C2)
Vanniyasingam, Thuva, 36, 130 (1E-C4)
Variyath, Asokan Mulayath, 31 (1D-C3), 36 (1E-C4), 42 (2B-C4), 59 (3B-C4), 105 (1D-C3), 130 (1E-C4), 152 (2B-C4), 232 (3B-C4)
Varu, Abhishek, 27 (1C-D1)
Vasudevan, Chithran, 31, 105 (1D-C3)
Veilleux, Déry, 30 (1D-C1), 45 (2D-I1), 98 (1D-C1), 169 (2D-I1)
Vernic, Raluca, 46, 170 (2D-I1)
Verret, François, 38, 135 (2B-I2)
Vicente, Sergio, 28 (1C-D2)
Vidyapeetham, Amrita Vishwa, 26, 87 (1B-C4)
Vidyashankar, Anand, 28 (1D-I1), 32 (1E-I1), 89 (1D-I1), 111 (1E-I1)
Vincent, Kyle, 53, 209 (2E-C4)
Volny, Dalibor, 46, 174 (2D-I3)
- Waddell, Adrian, 52, 206 (2E-C3)
Wang, Chunlin, 34, 120 (1E-C1)
Wang, Liangliang, 47, 177 (2D-I4)
Wang, Liqun, 59, 231 (3B-C3)
Wang, Shijia, 34, 119 (1E-C1)
Wang, Sunny, 51, 197 (2E-I4)
Wang, Xiaogang, 31, 102 (1D-C2)
Ward, Lawrence, 62, 242 (3D-I2)
Watier, François, 51, 200 (2E-C1)
Weerasinghe, Swarna, 42, 153 (2B-C4)
Weiss, William, 26, 86 (1B-C4)
Welch, William, 40 (2B-I5), 58 (3B-C1), 142 (2B-I5), 224 (3B-C1)
Weng, Chengguo, 55, 215 (3B-I1)
West, Brady, 57, 221 (3B-I4)
- Whitaker, Douglas, 33 (1E-I3), 45 (2C-P), 113 (1E-I3), 164, 165 (2C-P)
White, Bethany, 51, 199 (2E-I5)
Whoriskey, Kim, 57, 223 (3B-I5)
Willmot, Gordon E, 46, 170 (2D-I1)
Wilson, Kevin, 42 (2B-C4), 43--45 (2C-P), 153 (2B-C4), 156, 159, 165 (2C-P)
Wilton, Bree, 33, 113 (1E-I3)
Winiszewska, Malgorzata, 53, 209 (2E-C4)
Wolters, Mark, 24, 75 (1B-I5)
Woolford, Douglas, 58, 228 (3B-C2)
Wu, Changbao, 39 (2B-I4), 53 (2E-C4), 139 (2B-I4), 210 (2E-C4)
Wu, Dan, 28, 90 (1D-I2)
Wu, Yilei, 57, 224 (3B-C1)
Wu, Ying, 31, 105 (1D-C3)
Wüthrich, Mario, 23, 71 (1B-I3)
- Xia, Michelle, 59, 230 (3B-C3)
Xiao, Bowei, 27 (1C-D1)
Xie, Ming-ge, 30, 97 (1D-I5)
Xie, Qian, 46, 172 (2D-I2)
Xiong, Teng Ke, 27 (1C-D1)
Xiong, Yi, 47, 178 (2D-I5)
Xu, Joanne, 27 (1C-D1)
Xu, Mai, 28 (1C-D2)
Xu, Ximing, 62, 243 (3D-I3)
Xu, Yingying, 58, 228 (3B-C2)
- Yadegari, Iraj, 40, 146 (2B-C2)
Yan, Ying, 61, 238 (3C-A4)
Yang, Jinyoung, 26, 86 (1B-C4)
Yang, Po, 35, 125 (1E-C2)
Yang, Qian, 27 (1C-D1)
Yang, Shu, 39, 139 (2B-I4)
Yao, Fang, 63, 246 (3D-I5)
Yao, Zhigang, 28, 88 (1D-I1)
Ye, Lingyun, 45, 166 (2C-P)
Yi, Grace, 56, 216 (3B-I2)
Yilmaz, Yildiz, 47, 176 (2D-I4)
Yin, Yihao, 63, 244 (3D-I3)
Yu, Hao, 43 (2C-P), 50 (2E-I1), 155 (2C-P), 192 (2E-I1)
Yuan, Yan, 52, 204 (2E-C2)
Yurkowski, David, 57, 223 (3B-I5)
- Zamar, Ruben H., 58, 224 (3B-C1)
Zarepour, Mahmoud, 48, 185 (2D-C2)
Zeng, Leilei, 31 (1D-C3), 40 (2B-C1), 104 (1D-C3), 143 (2B-C1)

Zhai, Yongliang, 24, 76 (1B-C1)
Zhang, Hui, 61, 237 (3C-A3)
Zhang, Sabrina, 45, 168 (2D-I1)
Zhang, Yichi, 56, 216 (3B-I2)
Zhang, Ying, 43 (2C-P), 46 (2D-I3), 154 (2C-P), 174 (2D-I3)
Zhang, Yukun, 29 (1D-I3), 60 (3B-C4), 93 (1D-I3), 234 (3B-C4)
Zhao, Yang, 49, 186 (2D-C2)
Zhou, Xingcai, 46, 172 (2D-I2)
Zhu, Hongtu, 46, 172 (2D-I2)
Zhu, Jie, 28 (1C-D2)
Zhu, Mu, 52 (2E-C2), 57 (3B-C1), 204 (2E-C2), 224 (3B-C1)
Zhu, Yayuan, 31, 104 (1D-C3)
Zhu, Yeying, 59, 229 (3B-C3)
Zhu, Yunan, 24, 76 (1B-C1)
Zidek, James V., 25, 80 (1B-C2)
Zlatev, Boyko, 45, 166 (2C-P)
Zwiers, Francis, 49, 191 (2E-I1)

DALHOUSIE UNIVERSITY CAMPUS

STUDLEY CAMPUS



- KEY TO SYMBOLS**
- P — Permit Parking
 - PD — Metered Parking
 - PD — Parking for people with disabilities
 - * — Emergency "Code Blue" Phone



- E**
- E100 Student Union Building (SUB)
 - E190 Ristey Hall
 - E200 University Child Centre & Residence
 - E260 Kenneth C. Rowe Management Building
 - E280 Central Services Building
 - E600 Goldberg Computer Science Bldg.
 - E800 Goldengay Apartments

- D**
- D110 Mona Campbell Building
 - D280 Bioethics/Nu-Tech
 - D300 Environmental Health & Safety
 - D320 DeMillie House
 - D340 Student Residence
 - D400 Arts Centre, Cohn Auditorium
 - D420 Marion McCain Arts and Social Sciences Building
 - D541-42 Student residences
 - D550 Lyall House
 - D580 Colpitt House
 - D620 Schulich School of Law
 - D640 Chaplaincy Office/Multi-Faith Centre
 - D701 1308 Robie Street - ITS
 - D702 1312 Robie Street - ITS
 - D703 Health Professions
 - D720 1320 Robie Street - ITS

- C**
- C260 Sir James Dunn Building
 - C280 Chase Building
 - C300 Henry Hicks Building
 - C381, C382 Chemistry
 - C383 Chemical Storage Facility
 - C400 MacDonald Building
 - C440 University Club
 - C520 Howe Hall, University Health
 - C540 Sturdley House
 - C580 Killam Library, Mark A. Hill Accessibility Centre
 - C600 Sturdley Gymnasium
 - C710 Housing, Conference & Ancillary Services, DalCard Office
 - C720 European Studies
 - C730 APICS, Science Co-op
 - C750 Transition Year Program
 - C760, C800, C820 Economics
 - C770 Grad House
 - C790 LeMarchant Place

- B**
- B100 Dalplex
 - B120 Eliza Ritchie Hall
 - B200 Health & Human Performance
 - B260 DIFA, NSGEU
 - B280 Native Counselling Unit/Women's Centre

- A**
- A050 Institute for Research in Materials/International Ocean Institute
 - A100 Resilience Research Centre/FASS
 - A160 President's Residence

- C**
- C34 Wickwire Field
 - C201 LSC - Biology & Earth Science
 - C202 LSC - Oceanography
 - C203 LSC - Psychology
 - C204 Life Sciences Centre (LSC)
 - C220 Shirreff Hall
 - C230 Dalhousie-Ocean Sciences Building