

# Table of Contents • Table des matières

Table of Contents • Table des matières	3
Welcome • Bienvenue	4
Sponsors • Commanditaires	5
Organizers • Organisateurs	6
Exhibitors • Exposants	7
General Information • Information générale	7
Committees and Meetings • Comités et réunions	13
List of Sessions • Liste des sessions	17
Program • Programme	22
Abstracts • Résumés	54
Author Index • Index des auteurs	153
Author Index (per session) • Index des auteurs (par session)	158

## Welcome • Bienvenue

The University of Saskatchewan, the City of Saskatoon, and the Province of Saskatchewan welcome you. This year marks the Province's centennial and it is good for the members of the Statistical Society of Canada to join the local people in celebrating this anniversary. A little more than three weeks ago, the Queen honoured us with a visit. The Annual Meeting of the SSC does not generate as much public excitement as the presence of Her Majesty. Nevertheless, for statisticians, the talks and events of the coming days will create a similar enthusiasm.

We hope that you will find some time to explore Saskatoon and maybe other parts of Saskatchewan. They call it the land of the living skies. Instead of mountains we have clouds, instead of the sea, there is the prairie. Walking along the South Saskatchewan River you realize that one can experience nature even in the city.

We wish you a stimulating and enjoyable meeting as you continue to build the community of statisticians in Canada and the world.

L'Université de la Saskatchewan, la ville de Saskatoon, et la province de la Saskatchewan vous souhaitent la bienvenue. Cette année est le centenaire de la province et nous sommes heureux que les membres de la Société statistique du Canada joignent les gens locaux pour célébrer cet anniversaire. Il y a un peu plus de trois semaines la reine nous a honorés de sa visite. Le Congrès annuel de la SSC n'engendre pas une émotion aussi vive parmi la population que la présence de Sa Majesté. Néanmoins, pour les statisticiens et statisticiennes, les conférences et activités des prochains jours exprimeront une passion comparable.

Nous espérons que vous aurez du temps d'explorer Saskatoon et peut-être d'autres parties de la Saskatchewan. On l'appelle le pays des cieux vivants. Au lieu de montagnes nous avons des nuages, au lieu de mer, il y a la prairie. En faisant une promenade le long de la rivière Saskatchewan Sud, on constate que même dans la ville il est possible d'éprouver la nature.

Nous souhaitons que vous passiez des journées stimulantes et agréables tout en édifiant la communauté de statisticiens et de statisticiennes du Canada et du monde.

## Sponsors • Commanditaires

Le SSC 2005 thanks the following sponsors for their extremely generous contributions for the success of this meeting:

SSC 2005 remercie les commanditaires suivants pour leurs très généreuses contributions qui ont mené au succès du congrès :

- Le Centre de recherches mathématiques,
- the Fields Institute,
- the Pacific Institute for the Mathematical Sciences,
- MITACS,
- University of Saskatchewan,
- AON Consulting,
- MyTravel Sinfonia,
- W. H. Freeman.

# Organizers • Organisateurs

## Local Arrangements Committee • Comité organisateur local

**Chair • Président:** Mik Bickis

The Chair would like to express his sincerest thanks to his colleagues Vaneeta Grover, Sakti Jana, Tulay Koru-Sengul, Haydar Sengul, Chris Soteris, and Raj Srinivasan, for their generous help of various sorts over the last several months. He also wants to acknowledge the support of Robert Baker, Winfried Grassmann and Punam Pahwa.

Many people have contributed to the success of this conference. The Local Arrangements Chair is particularly grateful to Lynn Guina and the staff of the University of Saskatchewan Conference Services for taking care of all the details of registration, logistics, and feeding of the conference participants. Dion Rowney of the Division of Media and Technology created the excellent on-line registration system.

Le Centre de recherches mathématiques at the Université de Montréal kindly provided the system for capturing the abstracts found in this booklet. Its production was possible thanks to the advice and assistance of Christian Léger who produced it last year. Daniel Ouimet and André Montpetit of CRM prepared the data base and translated it into L<sup>A</sup>T<sub>E</sub>X code. Louis François Poirier carefully translated into French abstracts that had been submitted only in English. Jane Bickis developed the cover concept, which was implemented by Lori Verishagen of the University of Saskatchewan Printing Services. The final printing was accomplished by Maria Jochmaring.

We must also thank Ronda Pletz and Julie Zerebeski of MyTravel Sinfonia for facilitating hotel bookings; Rhonda Schindel and Veronica Nessman of the Sheraton Hotel; JoAnne Paydli, Jacqui McAfee, and Colleen Kisby of the Park Town Hotel; Marion Smale of the Western Development Museum; Don Turnmire of Sharp's Audio-Visual; and Janice Graham of Impact Marketing.

Le président souhaite offrir ses plus sincères remerciements à ses collègues Vaneeta Grover, Sakti Jana, Tulay Koru-Sengul, Haydar Sengul, Chris Soteris, et Raj Srinivasan, pour leur généreuse aide sous plusieurs formes au cours des derniers mois. Il veut également souligner l'appui qu'il a reçu de Robert Baker, Winfried Grassmann et Punam Pahwa.

Beaucoup de personnes ont contribué au succès du congrès. Le président du comité organisateur local est particulièrement reconnaissant à Lynn Guina et le personnel du service des congrès de l'Université de la Saskatchewan de prendre soin de tous les détails de l'inscription, de la logistique et de l'alimentation des congressistes. Dion Rowney de la Division média et technologie a construit l'excellent système d'inscription en ligne.

Le Centre de recherches mathématiques de l'Université de Montréal a gracieusement fourni le système pour capturer les résumés qui se trouvent dans ce cahier. Sa production a été possible grâce au conseil et à l'aide de Christian Léger qui l'avait produit l'an passé. Daniel Ouimet et André Montpetit du CRM ont préparé la base de données et l'ont transformée en code de L<sup>A</sup>T<sub>E</sub>X. Louis François Poirier a scrupuleusement traduit en français les résumés soumis uniquement en anglais. Jane Bickis a conçu le dessein de la couverture ; sa réalisation est accompli par Lori Verishagen du Service d'imprimerie de l'Université de la

Saskatchewan. Maria Jochmaring a réussi le tirage final des cahiers.

On doit aussi remercier Ronda Pletz et Julie Zerebeski de MyTravel Sinfonia pour faciliter l'hébergement; Rhonda Schindel et Veronica Nessman de l'Hôtel Sheraton; JoAnne Paydli, Jacqui McAfee et Colleen Kisby de l'Hôtel Park Town; Marion Smale du Western Development Museum; Don Turnmire de Sharp's Audio-Visual; et Janice Graham de Impact Marketing.

## **Program Committee • Comité du programme**

**Chair • Président:** Ausustine Wong, *York University*

**Members • Membres:**

**Biostatistics Section • Groupe de biostatistique**

Gordon Fick, *University of Calgary*

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

Stefan Steiner, *University of Waterloo*

**Survey Methods Section • Groupe de méthodologie d'enquête**

Changbao Wu, *University of Waterloo*,

## **Exhibitors • Exposants**

Displays of books for examination and purchase may be found in room 202 of the Arts building. Representatives of John Wiley, Pearson, and Thompson-Nelson will be present. W. H. Freeman did not send a representative, but their material may be viewed in the coffee area.

Une exposition des livres que l'on peut examiner ou acheter est située dans la salle 202 de l'Édifice des Arts. Les représentants des maisons d'édition John Wiley, Pearson, et Thompson-Nelson y assisteront. Il n'y a pas de représentant de W. H. Freeman, mais on peut consulter son matériel de promotion à l'endroit où à lieu la pause-café.

## **General Information • Information générale**

Registration will be in the Conference Foyer of the Sheraton Hotel on Sunday, June 12, during the hours 8:00–10:30, 12:30–14:00, and 16:00–19:00. From Monday to Wednesday, the registration desk will be located in the Arts Building at the University of Saskatchewan, in the area near room 133. It will be open on Monday from 7:30–15:00, on Tuesday from 8:00–12:30, and on Wednesday from 8:00–11:00.

L'inscription aura lieu dans le Conference Foyer de l'Hôtel Sheraton pendant les heures 8 :00–10 :30, 12 :30–14 :00, et 16 :00–19 :00. De lundi à mercredi, le comptoir d'inscription sera situé dans l'Édifice

des Arts à l'Université de la Saskatchewan près de la salle 133. Les heures d'ouverture seront de 7 :30–15 :00 le lundi, 8 :00–12 :30 le mardi, et 8 :00–11 :00 le mercredi.

## Directions

At a brisk walk, you can get to the University from the Sheraton Hotel in twenty minutes—fifteen from the Park Town. Leaving the Sheraton, turn left and walk along the river to the Park Town, which lies at the foot of the University Bridge. You should walk under the bridge and cross the river on the sidewalk on the north side of the bridge. On the other side you will find a dusty footpath going up a small rise. Follow the path (watch out for gopher holes!) to the grounds of the Royal University Hospital. Pass by St. Andrew's College onto Clinic Place and keep going until you pass the Dental Clinic. There you will see a passage going under a skywalk. Going through the passage, the observatory will be straight ahead, with an intersection of three roads to your left. On the other side of the intersection you will see a paved path with a large monolithic building—the library—to the right. Follow the path until you come to a patio on the left. A tall white building—the so-called “Arts Tower” is on the other side of the patio. That is where the conference will take place.

If you don't want to walk, a charter bus will leave the hotels at 7:45 and 8:15 on Monday, and at 8:15 and 8:45 on Tuesday and Wednesday. You can also take public transport. Bus numbers 6A and 25 go to the University. There is a bus stop beside the Park Town Hotel. Each route provides 30 minute service, so there is a bus every 15 minutes. On Sunday, however, you have to wait 30 minutes between buses. The fare is \$2.10, exact change.

If you are travelling by car, be advised that parking near the Arts Building is quite expensive: \$1.50 per hour. You can take advantage of parkades that are a bit farther but cheaper. The Stadium Parkade lies on the south side of College Drive, the street that comes off the bridge. You can park there for \$2, regardless of the time—you pay when you exit. A bit closer is the underground parkade in the Agriculture Building. After crossing the bridge, turn left at Wiggins Road (there are traffic lights), left again at Campus Drive and continue to Science Place on the right. At the end of this street is the parking area No. 5. The cost is \$3. It's about a 500 m walk to the Arts Building. (If it's raining, you can get from one place to the other without going outside.)

En marchant un peu rapidement, on peut arriver à l'Université de l'Hôtel Sheraton en vingt minutes ; de l'Hôtel Park Town ça prend quinze minutes. En quittant l'Hôtel Sheraton, tournez à gauche et marchez le long de la rivière jusqu'au Hôtel Park Town. Celui-ci est près du pont de l'Université. Il faut passer au dessous du pont et traverser la rivière sur le trottoir du côté nord du pont. A l'autre rive vous trouverez un chantier un peu poussiéreux qui monte une petite colline. Suivez ce chantier (attention aux terriers de spermophiles !) jusqu'au terrain du Royal University Hospital. Passez à côté du St. Andrew's College sur Clinic Place, et passez aussi le Dental Clinic. Là vous verrez un passage au dessous d'une passerelle. En sortant le passage il y aura l'observatoire tout droit. À gauche il y aura une intersection des trois chaussées. À l'autre côté de cette intersection vous verrez un trottoir avec un gros bâtiment monolithe—la bibliothèque—à droite. Suivez le trottoir jusqu'à une petite terrasse qui se présente de gauche. A l'autre côté il y aura un édifice blanc en hauteur—le “Arts Tower”. Le lieu du Congrès se trouve dans cet édifice.

Si vous ne voulez pas marcher, il y aura un autobus nolisé qui partira des deux hôtels à 7h45 et 8h15 le lundi, et à 8h15 et 8h45 les mardis et mercredis. Il y a également le transport public. Les autobus N<sup>os</sup> 6A et 25 vont à l'Université. Il y a un point d'arrêt à coté de l'Hôtel Park-Town. Chaque ligne a une fréquence de 30 minutes—alors il y a un autobus chaque 15 minutes. Le dimanche, cependant, il faut attendre 30 minutes entre les autobus. Le tarif est 2,10\$, monnaie exacte requise.

Si vous voyagez en voiture, soyez averti que le stationnement près de l'Édifice des Arts est cher : 1,50\$ par heure. Vous pouvez profiter des parcs de stationnement un peu plus loin qui sont moins coûteux. Le

Stadium Parkade se trouve à côté sud du College Drive, la rue qui amène du pont de l'Université. On peut y stationner pour 2\$, sans égard pour le temps—vous payez lorsque vous sortez. Un peu plus près il y a un parc de stationnement souterrain dans l'Édifice d'Agriculture. Après avoir traversé le pont, tournez à gauche à Wiggins Road (il y a des feux de circulation), à gauche encore à Campus Drive et continuez jusqu'au Science Place à droite. Au bout de cette rue se trouve le parc de stationnement N° 5. Le prix est 3\$. C'est une marche de 500 m environ à l'Édifice des Arts. (S'il pleut, il est possible de passer d'un endroit à l'autre sans aller à l'extérieur.)

## Rooms • Salles

All the sessions take place in the Arts Building at the University of Saskatchewan. Room 133 is on the ground floor, to the right of the registration desk. Room 241 is on top of the ramp behind the desk. The other rooms are located on the second floor off the hallway facing you as you come up the ramp.

Toutes les séances ont lieu dans l'Édifice des Arts à l'Université de la Saskatchewan. La salle 133 est située au rez-de-chaussée, à droite du comptoir d'inscription. La salle 241 se trouve au dessus de la rampe derrière le comptoir. Les autres salles s'étendent du couloir situé au deuxième étage, tout droit en montant la rampe.

## Workshops • Ateliers

The workshops organized by the three sections will take place in the Sheraton Hotel. Note that the Bio-statistics Section workshop starts at 9:00 while the other two start at 13:00.

Les ateliers organisés par les trois groupes auront tous lieu le dimanche à l'Hôtel Sheraton. Notez bien que l'atelier du groupe de biostatistique débutera à **9h30** alors que les deux autres débuteront à **13h00**.

## Poster Session • Séance d'affichage

The poster session will be held on Sunday, 12 June in the Conference Foyer of the Sheraton Hotel, between the hours of 13:00 and 18:00.

Le séance d'affichage aura lieu le dimanche 12 juin dans le Conference Foyer de l'Hôtel Sheraton, entre **13h00** et **18h00**.

## Synchrotron Tour • Visite au synchrotron

The Candian Light Source is a unique research facility in Canada, the second in the world of its kind. A tour of this installation has been arranged for Monday, June 13 at 15:30. Advance registration is required which may be done at the Conference registration desk. The tour group is limited to twenty people. The group should assemble at the registration desk during the coffee break so as to be able to leave for the synchrotron (a 1 km walk) at 15:10.

Le Centre canadien de rayonnement synchrotron est une installation de recherche unique au Canada, le deuxième au monde de sa catégorie. Une visite à cette installation est organisée le lundi, 13 juin à 15h30. L'inscription à l'avance est nécessaire. On peut s'inscrire au comptoir d'inscription du Congrès. Le groupe des visiteurs est limité à vingt personnes. Le groupe doit se rassembler au comptoir d'inscription pendant la pause-café afin que l'on puisse partir vers le synchrotron à 15h10 ; prévoir une marche d'un kilomètre.

## **NSERC Workshop • Atelier offert par le CRSNG**

### **Tips to Prepare Your Next Discovery Grant Application**

**Monday, June 13, 17:00–18:30**

**Arts 214**

NSERC representatives and Grant Selection Committee members will present an interactive overview of the peer review process, inform you of the latest changes at NSERC, give useful advice for the preparation of your next NSERC application and answer your questions on the functioning of grant selection committees.

The workshop is open to all researchers. It is particularly helpful for new faculty members and researchers likely to apply (or re-apply) in the fall.

**Conseils pour l'élaboration de votre prochaine demande de subvention à la découverte**

**Lundi, le 13 juin, 17h00–18h30**

**Arts 214**

Des employés du CRSNG et des membres des comités de sélection des subventions du CRSNG donneront une présentation en mode interactif sur le processus d'évaluation par les pairs, vous renseigneront sur les derniers changements opérés au CRSNG, vous donneront des conseils utiles pour l'élaboration de votre prochaine demande de subvention au CRSNG et répondront à vos questions sur le fonctionnement des comités de sélection des subventions.

Tous les chercheurs peuvent participer à l'atelier, qui sera particulièrement utile aux professeurs et aux chercheurs récemment embauchés qui comptent présenter une demande (ou une nouvelle demande) à l'automne.

## **Probability Section • Groupe de probabilité**

There will be a meeting during lunch on Wednesday for those interested in forming an SSC Section on Probability. For more details, see Andre Dabrowski or Gail Ivanoff.

Pour ceux qui s'intéressent à la constitution d'un Groupe de probabilité, il y aura une assemblée d'organisation pendant le dîner mardi. Pour des plus amples renseignements, on peut s'informer à Andre Dabrowski ou à Gail Ivanoff.

## **Opening reception • Réception d'accueil**

The welcoming reception will take place at the "Top of the Inn" at the Sheraton Hotel on Sunday, 12 June from 18:00 to 20:00. All conference attendees and companions are warmly invited to come and meet old friends and make new acquaintances.

La réception d'accueil aura lieu dans le « Top of the Inn » au dernier étage de l'Hôtel Sheraton, le dimanche 12 juin de 18h00 à 20h00. Les participants au congrès et leurs compagnons sont tous invités à venir retrouver de vieux amis et faire de nouvelles connaissances.

## **Food Service on Campus • Restauration sur le campus**

Lunch will be served from 12:30 to 13:30, Monday to Wednesday in Marquis Hall, situated at the end of the elm-lined avenue to your left as you leave the Arts Building. The cost of lunch is included in the registration.



Le dîner sera servi de 12h30 à 13h30, lundi à mercredi, au Marquis Hall, situé au bout de l'allée bordée d'ormes qui se trouve à votre gauche quand vous quittez l'Édifice des Arts. Les dîners sont compris dans les frais d'inscription.

## **Barbecue**

The annual barbecue will be held at the Western Development Museum on Monday, 13 June. Buses will take all those who registered for the barbecue to the museum. Buses will leave the University at 18:00 and 18:30. For the benefit of those who skipped the late-afternoon meetings and returned to their hotel, the buses will stop at the Park Town and Sheraton Hotels on the way to the museum. (One can walk to the museum but it is over 4 km from the Sheraton.)

At the museum, you may enjoy the display while you take a drink and are entertained by fiddlers Cammy and Sarah Romanuck. The meal will be served at 19:30. Buses will return to the hotels as well as to the University residences around 22:00.

Le Barbecue annuel aura lieu au Musée de développement de l'Ouest (Western Development Museum), le lundi 13 juin. Des autobus conduiront au musée tous ceux qui se sont inscrits au barbecue. Les autobus quitteront l'université à 18h00 et à 18h30. Dans l'intérêt de ceux qui auront raté les dernières réunions de l'après-midi et seront retournés à l'hôtel, les autobus s'arrêteront aux hôtels Park-Town et Sheraton en route vers le musée. (Il est possible d'aller au musée à pied, mais c'est une distance de plus de 4 km de l'Hôtel Sheraton.)

Au musée, vous pouvez regarder l'exposition, prendre un apéritif pendant que Cammy et Sarah Romanuck, deux jeunes violoneuses, vous divertiront. Le repas commencera à 19h30. Les autobus retourneront aux hôtels et aussi à la résidence de l'université autour de 22h00.

## **Women in Statistics Reception • Réception pour les femmes en statistique**

The Canadian Section of the Caucus for Women in Statistics and the SSC Committee on Women in Statistics invite interested men and women to attend an informal dinner at 7pm on Monday June 13 at Mykonos Restaurant (416 - 21st Street East, near Sheraton). This gathering is a pay event and to enable reservations, those who have not previously reserved a spot need to leave a message for Rhonda Rosychuk at the conference registration desk as soon as possible.

Le Caucus for Women in Statistics, en collaboration avec le Comité sur les femmes en statistique de la SSC, souperont à 19h00, le lundi 13 juin, au restaurant Mykonos, 416 21st Street East, près de l'Hôtel Sheraton.

## **Banquet**

All participants are cordially invited to attend the Conference Banquet on Tuesday evening, June 14. A cash bar will be available from 18:30. Dinner will be served at 19:30. After the meal, there will a little entertainment by the duo of Deborah Buck and Brad Grass, which will be followed by the Society awards. At 23:00, a bus will return to the University for the benefit of those staying at the residence.

Tous les participants sont cordialement invités à prendre part au Banquet du congrès qui aura lieu le mardi 14 juin en soirée dans l'Hôtel Sheraton. Un bar payant sera disponible dans le Conference Foyer dès 18h30. Le repas commencera à 19h30. Après le dîner il y aura un petit divertissement présenté par le

duo Deborah Buck et Brad Grass, suivi par la remise des prix de la Société. Un autobus ramena ceux qui demeurent à la résidence de l'université à 23h00.

# Committees and Meetings

## Comités et réunions

### Calendar Listing • Calendrier journalier

#### Saturday June 11 • Samedi 11 juin

18:00–22:00 Executive Committee • Comité exécutif Sheraton Board Room

#### Sunday, June 12 • Dimanche 12 juin

9:00–11:00 Finance Committee • Comité des finances Sheraton East Room  
 11:00–12:00 Publications Committee • Comité des publications Sheraton Board Room  
 12:00–16:00 Board of Directors • Conseil d'administration Sheraton East Room

#### Monday, June 13 • Lundi 13 juin

7:00–8:00 Biostatistics Section (Section Officers) • Groupe de bio- Park Town South Dining Room  
 statistique (directeurs du groupe)  
 Business and Industrial Statistics Section (Section Offi- Park Town South Dining Room  
 cers) • Groupe de statistique industrielle et de gestion  
 (directeurs du groupe)  
 Survey Methods Section (Section Officers) • Groupe des Park Town South Dining Room  
 méthodes d'enquête (directeurs du groupe)  
 12:30–13:30 Accreditation Appeals Committee • Comité des appels Private Dining Room,  
 d'accréditation Marquis Hall  
 CJS Editorial Board • Comité de rédaction de la RCS Private Dining Room,  
 Marquis Hall  
 Public Relations Committee • Comité des relations Private Dining Room,  
 publiques Marquis Hall  
 Southern Ontario Regional Association (General Meet- Private Dining Room,  
 ing) • Association régionale du sud de l'Ontario (assem- Marquis Hall  
 blée générale)  
 13:30–15:00 Accredited Members General Meeting (open) • Assem- Arts 200  
 blée générale des membre accrédités (publique)  
 17:00–18:00 Business and Industrial Statistics Section (General Arts 217  
 Meeting) • Groupe de statistique industrielle et de ges-  
 tion (assemblée générale)  
 Survey Methods Section (General Meeting) • Groupe Arts 212  
 des méthodes d'enquête (assemblée générale)

**Tuesday, June 14 • Mardi 14 juin**

7:00–8:00	Research Committee • Comité de la recherche Accreditation Committee • Comité sur l'accréditation	Park Town South Dining Room Park Town South Dining Room
12:30–13:30	NPCDS • PNSDC  Bilingualism Committee • Comité du bilinguisme  Committee on Women in Statistics • Comité sur les femmes en statistique Statistical Education Committee • Comité d'éducation en statistique	Private Dining Room, Marquis Hall Private Dining Room, Marquis Hall Private Dining Room, Marquis Hall Private Dining Room, Marquis Hall
15:30–17:00	SSC General Meeting • Assemblée générale de la SSC	Arts 241
17:00–18:00	Biostatistics Section (General Meeting) • Groupe de biostatistique (assemblée générale)	Arts 212

**Wednesday, June 15 • Mercredi 15 juin**

7:00–8:00	Biostatistics Section (Section Officers) • Groupe de biostatistique (directeurs du groupe) Business and Industrial Statistics Section (Section Officers) • Groupe de statistique industrielle et de gestion (directeurs du groupe) Survey Methods Section (Section Officers) • Groupe des méthodes d'enquête (directeurs du groupe) Professional Development Committee • Comité sur le perfectionnement professionnel	Park Town South Dining Room Park Town South Dining Room Park Town South Dining Room Park Town South Dining Room
12:30–13:30	Liaison  Probability Section Organizational Meeting • Assemblée d'organisation du groupe de probabilité Program Committee • Comité du programme	Private Dining Room, Marquis Hall Private Dining Room, Marquis Hall Private Dining Room, Marquis Hall
17:30–19:30	Board of Directors • Conseil d'administration	Private Dining Room, Marquis Hall
19:30	Executive Committee • Comité exécutif	Private Dining Room, Marquis Hall

**Alphabetical Listing • liste alphabétique****Accreditation Committee • Comité sur l'accréditation**

Tuesday, June 14 • mardi 14 juin, 7:00–8:00, Park Town South Dining Room

**Accreditation Appeals Committee • Comité des appels d'accréditation**

Monday, June 13 • lundi 13 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Accredited Members General Meeting (open) • Assemblée générale des membres accrédités (publique)**

Monday, June 13 • lundi 13 juin, 13:30–15:00, Arts 200

**Bilingualism Committee • Comité du bilinguisme**

Tuesday, June 14 • mardi 14 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Biostatistics Section (General Meeting) • Groupe de biostatistique (assemblée générale)**

Tuesday, June 14 • mardi 14 juin, 17:00–18:00, Arts 212

**Biostatistics Section (Section Officers) • Groupe de biostatistique (directeurs du groupe)**

Monday, June 13 • lundi 13 juin, 7:00–8:00, Park Town South Dining Room

Wednesday, June 15 • mercredi 15 juin, 7:00–8:00, Park Town South Dining Room

**Board of Directors • Conseil d'administration**

Sunday, June 12 • dimanche 12 juin, 12:00–16:00, Sheraton East Room

Wednesday June 15 • mercredi 15 juin, 17:30–19:30, Private Dining Room, Marquis Hall

**Business and Industrial Statistics Section (General Meeting) • Groupe de statistique industrielle et de gestion (assemblée générale)**

Monday, June 13 • lundi 13 juin, 17:00–18:00, Arts 217

**Business and Industrial Statistics Section (Section Officers) • Groupe de statistique industrielle et de gestion (directeurs du groupe)**

Monday, June 13 • lundi 13 juin, 7:00–8:00, Park Town South Dining Room

Wednesday, June 15 • mercredi 15 juin, 7:00–8:00, Park Town South Dining Room

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

Monday, June 13 • lundi 13 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Committee on Women in Statistics • Comité sur les femmes en statistique**

Tuesday, June 14 • mardi 14 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Executive Committee • Comité exécutif**

Saturday June 11 • samedi 11 juin, 18:00–22:00, Sheraton Board Room

Wednesday June 15 • mercredi 15 juin, 19:30, Private Dining Room, Marquis Hall

**Finance Committee • Comité des finances**

Sunday, June 12 • dimanche 12 juin, 9:00–11:00, Sheraton East Room

**NPCDS • PNSDC**

Tuesday, June 14 • mardi 14 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Liaison**

Wednesday, June 15 • mercredi 15 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Probability Section (Organizational Meeting) • Groupe de probabilité (Assemblée d'organisation)**

Wednesday June 15 • mercredi 15 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Professional Development Committee • Comité sur le perfectionnement professionnel**

Wednesday June 15 • mercredi 15 juin, 7:00–8:00, Park Town South Dining Room

**Program Committee • Comité du programme**

Wednesday June 15 • mercredi 15 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Public Relations Committee • Comité des relations publiques**

Monday, June 13 • lundi 13 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Publications Committee • Comité des publications**

Sunday, June 12 • dimanche 12 juin, 11:00–12:00, Sheraton Board Room

**Research Committee • Comité de la recherche**

Tuesday, June 14 • mardi 14 juin, 7:00–8:00, Park Town South Dining Room

**Southern Ontario Regional Association (General Meeting) • Association régionale du sud de l'Ontario (assemblée générale)**

Monday, June 13 • lundi 13 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**SSC General Meeting • Assemblée générale de la SSC**

Tuesday, June 14 • mardi 14 juin, 15:30–17:00, Arts 241

**Statistical Education Committee • Comité d'éducation en statistique**

Tuesday, June 14 • mardi 14 juin, 12:30–13:30, Private Dining Room, Marquis Hall

**Survey Methods Section (General Meeting) • Groupe des méthodes d'enquête (assemblée générale)**

Monday, June 13 • lundi 13 juin, 17:00–18:00, Arts 212

**Survey Methods Section (Section Officers) • Groupe des méthodes d'enquête (directeurs du groupe)**

Monday, June 13 • lundi 13 juin, 7:00–8:00, Park Town South Dining Room

Wednesday, June 15 • mercredi 15 juin, 7:00–8:00, Park Town South Dining Room

## List of Sessions • Liste des sessions

<b>Session 0</b>	Contributed Poster Session <i>Présentation d'affiches</i>	June 12 juin	13:00 – 18:00	Sheraton Conference Foyer
<b>Session 01A</b>	Welcome <i>Bienvenu</i>	June 13 juin	08:30 – 09:00	241
<b>Session 01B</b>	SSC Presidential Invited Address <i>Allocution de l'invitée de la Présidente de la SSC</i>	June 13 juin	09:00 – 10:30	241
<b>Session 02A</b>	Statistics models for discrete longitudinal data <i>Modèles statistiques pour des données longitudinales discrètes</i>	June 13 juin	11:00 – 12:30	214
<b>Session 02B</b>	Strong asymptotics <i>Asymptotiques fortes</i>	June 13 juin	11:00 – 12:30	200
<b>Session 02C</b>	How and how well do we train future consultants <i>Comment et avec quelle qualité entraînons-nous nos futurs consultants</i>	June 13 juin	11:00 – 12:30	212
<b>Session 02D</b>	Bayes and likelihood <i>Bayes et la vraisemblance</i>	June 13 juin	11:00 – 12:30	133
<b>Session 02E</b>	The Joint Canada/United States Survey of Health: Analytic results and methodological issues <i>L'enquête conjointe Canada/États-Unis sur la santé: Résultats analytiques et questions méthodologiques</i>	June 13 juin	11:00 – 12:30	211
<b>Session 02F</b>	Current advancement in statistical genetics <i>Progrès récents en génétique statistique</i>	June 13 juin	11:00 – 12:30	241
<b>Session 02G</b>	Probability <i>Probabilité</i>	June 13 juin	11:00 – 12:15	217
<b>Session 03A</b>	Special Invited Session of the Biostatistics Section <i>Séance spéciale de l'invité du groupe de biostatistique</i>	June 13 juin	13:30 – 15:00	241

<b>Session 03B</b>	Quality improvement <i>Amélioration de la qualité</i>	June 13 juin	13:30 – 15:00	214
<b>Session 03C</b>	Measurement error models <i>Modèles d'erreurs de mesure</i>	June 13 juin	13:30 – 15:00	133
<b>Session 03D</b>	Recent development of finite mixture models <i>Développements récents dans les mélanges de modèles finis</i>	June 13 juin	13:30 – 15:00	211
<b>Session 03E</b>	Elaboration of a calibration strategy: a case study <i>Élaboration d'une stratégie de calage: une étude de cas</i>	June 13 juin	13:30 – 14:30	217
<b>Session 03F</b>	Bayesian methodology <i>Méthodologie bayésienne</i>	June 13 juin	13:30 – 14:30	212
<b>Session 03G</b>	General Meeting of SSC Accredited members (open) <i>Assemblée générale des membres accrédités par la SSC</i>	June 13 juin	13:30 – 15:00	200
<b>Session 04A</b>	Statistical climatology <i>Climatologie statistique</i>	June 13 juin	15:30 – 17:00	241
<b>Session 04B</b>	Inference from multiple surveys and multiple frame surveys <i>Inférence à partir d'enquêtes multiples et d'enquêtes à cadres multiples</i>	June 13 juin	15:30 – 17:00	214
<b>Session 04C</b>	Methodology on missing data in longitudinal studies <i>Méthodologie pour les données manquantes dans les études longitudinales</i>	June 13 juin	15:30 – 17:00	200
<b>Session 04D</b>	Spatial analysis of survival data <i>Analyse spatiale de données de survie</i>	June 13 juin	15:30 – 17:00	133
<b>Session 04E</b>	Dependent data analysis <i>Analyses des données dépendantes</i>	June 13 juin	15:30 – 16:45	211
<b>Session 04F</b>	Methods with applications <i>Méthodes et applications</i>	June 13 juin	15:30 – 17:00	212
<b>Session 04G</b>	Statistical genetics and genomics <i>Statistique génétique et génomique</i>	June 13 juin	15:30 – 16:15	217
<b>Session 04H</b>	NSERC Workshop <i>Atelier du CRSNG</i>	June 13 juin	17:00 – 18:30	214
<b>Session 05</b>	Gold Medal Address <i>Allocution de la Médaille d'or</i>	June 14 juin	09:00 – 10:30	241
<b>Session 06A</b>	Bayesian model selection <i>Sélection de modèles bayésiens</i>	June 14 juin	11:00 – 12:30	133



<b>Session 06B</b>	Extensions of policy-relevant analytical methods for complex survey data <i>Extensions des méthodes analytiques dépendant des politiques pour les données d'enquêtes complexes</i>	June 14 juin	11:00 – 12:30	217
<b>Session 06C</b>	Statistical genomics: New platforms, new data, new analysis methods for high-throughput experiments <i>La génomique statistique: Nouvelles plateformes, nouvelles données, nouvelles méthodes d'analyse pour les expériences à haut débit.</i>	June 14 juin	11:00 – 12:30	214
<b>Session 06D</b>	Computer network modeling <i>Modélisation de réseaux informatiques</i>	June 14 juin	11:00 – 12:30	200
<b>Session 06E</b>	Statistics in cancer, AIDS, and other medical researches <i>Statistique pour le cancer, le SIDA, et autres recherches médicales</i>	June 14 juin	11:00 – 12:30	241
<b>Session 06F</b>	Statistics and finance <i>Statistique et finance</i>	June 14 juin	11:00 – 12:30	211
<b>Session 06G</b>	Estimation <i>Estimation</i>	June 14 juin	11:00 – 12:30	212
<b>Session 07A</b>	Statistics and technology <i>Statistique et technologie</i>	June 14 juin	13:30 – 15:00	133
<b>Session 07B</b>	Limit theorems <i>Théorèmes limites</i>	June 14 juin	13:30 – 15:00	212
<b>Session 07C</b>	Statistical methodology in clinical trial design <i>Méthodologie statistique dans la planification d'essais cliniques</i>	June 14 juin	13:30 – 15:00	241
<b>Session 07D</b>	Applications of mixture models in genetic linkage analysis <i>Application des mélanges de modèles en analyse de liaison génétique</i>	June 14 juin	13:30 – 15:00	200
<b>Session 07E</b>	Case Study 1: Antidepressant Medication and Stability in Seniors <i>Analyse de cas 1: Sur les antidépresseurs et la stabilité chez les personnes âgées</i>	June 14 juin	13:30 – 14:40	214
<b>Session 07F</b>	Applied survey methods <i>Méthodes d'enquête appliquées</i>	June 14 juin	13:30 – 14:45	211
<b>Session 07G</b>	Advanced Regression Methods <i>Méthodes de régression avancées</i>	June 14 juin	13:30 – 15:00	217

<b>Session 09A</b>	Pierre Robillard Award Address <i>Allocution du prix Pierre Robillard</i>	June 15 juin	09:00 – 09:45	241
<b>Session 09B</b>	Canadian Journal of Statistics Award Address <i>Allocution du prix de La Revue canadienne de statistique</i>	June 15 juin	09:45 – 10:30	241
<b>Session 10A</b>	Imputation <i>Imputation</i>	June 15 juin	11:00 – 12:30	214
<b>Session 10B</b>	Recent Advances in Lifetime Data <i>Progrès récents pour les données de durée</i>	June 15 juin	11:00 – 12:30	200
<b>Session 10C</b>	Estimation in constrained parameter spaces <i>Estimation dans des espaces de paramètres avec contraintes</i>	June 15 juin	11:00 – 12:30	133
<b>Session 10D</b>	Graphical Gaussian models <i>Modèles gaussiens graphiques</i>	June 15 juin	11:00 – 12:30	217
<b>Session 10E</b>	Teaching statistics in the health and life sciences <i>L'enseignement de la statistique dans les sciences de la santé et de la vie</i>	June 15 juin	11:00 – 12:30	241
<b>Session 10F</b>	Model Selection and Robust Modelling <i>Sélection des modèles et la modélisation robuste</i>	June 15 juin	11:00 – 12:30	212
<b>Session 10G</b>	Sampling and Design Issues <i>Questions de sondage et de planification</i>	June 15 juin	11:00 – 12:30	211
<b>Session 11A</b>	Special Invited Session of the Survey Section <i>Séance spéciale de l'invité du groupe d'enquête</i>	June 15 juin	13:30 – 15:00	133
<b>Session 11B</b>	Case Study II: The National Longitudinal Study on Children and Youth (NLSCY) <i>Étude de cas II: L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ)</i>	June 15 juin	13:30 – 15:00	214
<b>Session 11C</b>	Analysis of gene expression data <i>L'analyse des données d'expression génétique.</i>	June 15 juin	13:30 – 15:00	241
<b>Session 11D</b>	Business and industrial statistics methods and applications <i>Méthodologie et applications de la statistique industrielle et de gestion</i>	June 15 juin	13:30 – 14:30	217
<b>Session 11E</b>	Clinical Trials and Related Issues <i>Essais cliniques et questions associées</i>	June 15 juin	13:30 – 15:00	200
<b>Session 11F</b>	Correlated Discrete Data <i>Données discrètes corrélées</i>	June 15 juin	13:30 – 14:30	212

<b>Session 12A</b>	CRM-SSC Award Address <i>Allocution du lauréat du Prix CRM-SSC 2005</i>	June 15 juin	15:30 – 17:00	241
<b>Session 12B</b>	Design of experiments <i>Planification d'expériences</i>	June 15 juin	15:30 – 17:00	214
<b>Session 12C</b>	Large deviations in statistics <i>Déviations importantes en statistique</i>	June 15 juin	15:30 – 17:00	217
<b>Session 12D</b>	Model and design based inference for sample surveys <i>Inférence fondée sur le modèle et sur le plan pour les sondages</i>	June 15 juin	15:30 – 17:00	133
<b>Session 12E</b>	Theoretical survey methods <i>Méthodes d'enquête théoriques</i>	June 15 juin	15:30 – 16:45	211
<b>Session 12F</b>	Statistical modeling and applications <i>Modélisation statistique et ses applications</i>	June 15 juin	15:30 – 16:45	212
<b>Session 12G</b>	Methods with Censored Data <i>Méthodologie pour les données tronquées</i>	June 15 juin	15:30 – 17:00	200

# Program • Programme

**Dimanche 12 juin • Sunday, June 12**

---

**13:00 – 18:00 Session 0**

**Sheraton Conference Foyer**

Contributed Poster Session

Présentation d'affiches

Contributed Paper Session • Séance de communications libres

13:00 Keramat NOURIJELYANI, *Tehran University of Medical Sciences*

Statistical analysis of association between pre-term delivery and low weight gain • Analyse statistique de l'association entre le faible gain de poids durant la grossesse et l'accouchement prématuré. [MS-1]

13:00 Judy-Anne CHAPMAN, *NCIC Clinical Trials Group, Queens University*, Edward YU, *University of Western Ontario*, Melanie ROYCE, *University of New Mexico*, Patricia TAI, *University of Saskatchewan*

Effects of statistical method of assessment on prognosis of limited-stage small-cell lung cancer • Effets des méthodes d'analyse statistique sur le pronostic du cancer à petites cellules du poumon à niveau limité [MS-2]

13:00 Pengfei GUO & Alwell OYET, *Memorial University of Newfoundland*

Tests for equality of curves via wavelets • Tests d'égalité de courbes par les ondelettes [MS-3]

13:00 Kayo WALSH, *University of Colorado at Denver*, Eduard GAMITO, Richard JONES, Shannon PRETZEL, Colin O'DONNELL & Sheryl OGDEN, *University of Colorado Health Sciences Center*

A randomized study to determine the effectiveness of interventions to improve participant protocol compliance in a large clinical study. • Une étude randomisée pour déterminer l'efficacité d'interventions pour améliorer le protocole de conformité des participants dans une importante étude clinique [MS-4]

13:00 Mireille GUAY, Betty EDWARDS & Mark RAIZENNE, *Santé Canada / Health Canada*

Long-term effects of air pollution on lung function of children and adolescents in Canada • Effets d'une Exposition Prolongée à la Pollution Atmosphérique sur la Capacité Respiratoire de Jeunes Canadiens [MS-5]

13:00 Michael SZAFRON & Christine SOTEROS, *University of Saskatchewan*

The probability of knotting after a local strand in a self-avoiding polygon • Probabilité de nouage après un brin local dans un polygone qui s'auto évite [MS-6]

13:00 P. YU, *University of Saskatchewan*

Applications of statistical methods - principal component (PCA) and cluster analysis (CLA) in synchrotron-based ftir spectra analysis for feed/plant inherent structure research • Applications de méthodes statistiques - Analyse en composantes principales (ACP) et classification automatique (CA) en analyse de spectre FTIR basé sur le synchrotron pour la recherche sur les structures inhérentes aliments/plantes [MS-7]

13:00 Peter HOOPER, *University of Alberta*

Remedies for serial correlation and multiple local optima in period analysis of variable stars. • Remède pour la corrélation sérielle et les optimums locaux multiples dans l'analyse périodique d'étoiles variables [MS-8]

13:00 Chandima KARUNANAYAKE & William H. LAVERTY, *University of Saskatchewan*

Application of multivariate poisson models for analysis of weed counts • Application de modèles de Poisson multivariés pour l'analyse du nombre de variétés de plantes [MS-9]

13:00 Qiaohao ZHU & Narasimha PRASAD, *University of Alberta*

On the estimation of number of exceedances over a threshold • Estimation de la fréquence de dépassement d'un seuil [MS-10]

13:00 M. MIKET, W. H. LAVERTY, C. PRINGLE-NELSON & I. W. KELLY, *University of Saskatchewan*

Expressions of life meaning among college students • Le sens de la vie selon des étudiants collégiaux [MS-11]

13:00 Jing LIN & Miao FANG, *McMaster University*

Case Study II: The National Longitudinal Study on Children and Youth (NLSCY) • Étude de cas II : L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) [MS-12]

## Lundi 13 juin • Monday, June 13

---

### 08:30 – 09:00 Session 01A

241

Welcome

Bienvenu

Special Session • Conférence spéciale

Chair • Présidente : Nancy REID *University of Toronto*08:30 Peter MACKINNON, *President, University of Saskatchewan*08:40 Jo-Anne DILLON, *Dean, College of Arts and Science*08:50 Jim BASINGER, *Associate Dean, Science*

### 09:00 – 10:30 Session 01B

241

SSC Presidential Invited Address

Allocution de l'invitée de la Présidente de la SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et présidente : Nancy REID *University of Toronto*09:00 Barbara KEYFITZ, *Fields Institute and University of Houston*

Deterministic and statistical models for turbulence: What could burgers have said to Kolmogorov?

• Modèles déterministe et statistique pour la turbulence : qu'est-ce que Burgers aurait pu dire à Kolmogorov ? [MS-13]

### 10:30 – 11:00 Pause-café • Coffee Break

### 11:00 – 12:30 Session 02A

214

Statistics models for discrete longitudinal data

Modèles statistiques pour des données longitudinales discrètes

Invited Paper Session • Présentations sur invitation

Caucus for Women

Organizer and Chair • Responsable et présidente : Julie HORROCKS *University of Guelph*11:00 Rachel ALTMAN, *Simon Fraser University*, Stefan REINKER, *University of Freiburg*, John STAUDENMAYER, *University of Massachussetts*

Some unusual applications of hidden Markov models • Quelques nouvelles applications des chaînes de Markov cachées [MS-14]

11:30 Diana MIGLIORETTI, *Group Health Cooperative*, Elizabeth BROWN, *University of Washington*

A marginalized model for combining state and national level survey data to estimating screening trends • Un modèle marginalisé pour combiner des données de sondages au niveau des états et au niveau national pour estimer les tendances de dépistage [MS-15]

12:00 Rhonda ROSYCHUK, *University of Alberta*

Comparison of goodness-of-fit tests based on transition counts for misclassified, binary, alternating longitudinal data • Comparaison de tests de qualité d'ajustement basés sur les compteurs de transition pour des données longitudinales binaires, alternantes et mal classifiées [MS-16]

### 11:00 – 12:30 Session 02B

200

Strong asymptotics

Asymptotiques fortes

Invited Paper Session • Présentations sur invitation

Bernoulli Society

Organizer and Chair • Responsable et président : Andre DABROWSKI *University of Ottawa*

11:00 Andrew HEUNIS, *University of Waterloo*, Seyed HASHEMI, *University of Manitoba*

Asymptotic limits for coupled stochastic DE's • Limites asymptotiques pour des équations différentielles stochastiques couplées [MS-17]

11:45 Raluca BALAN, *University of Ottawa*

A strong invariance principle for associated random fields • Un principe d'invariance forte pour des champs aléatoires associés [MS-18]

### 11:00 – 12:30 Session 02C

212

How and how well do we train future consultants

Comment et avec quelle qualité entraînons-nous nos futurs consultants

Invited Paper Session • Présentations sur invitation

Biostatistics Section

Organizer and Chair • Responsable et présidente : Jeanette O'HARA HINES *University of Waterloo*

11:00 Jeanette O'HARA HINES, *University of Waterloo*, David BELLHOUSE, *University of Western Ontario*, John PETKAU, *University of British Columbia*, Heather THIESSEN, *London Health Sciences Centre*, Jamie MYLES, *Pfizer*

How and how well do we train future consultants? • Comment, et comment bien entraînons nous les futures consultants [MS-19]

### 11:00 – 12:30 Session 02D

133

Bayes and likelihood

Bayes et la vraisemblance

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Don FRASER *University of Toronto*

11:00 Tom SEVERINI, *Northwestern University*

Elimination of nuisance parameters by integration • Élimination de paramètres de nuisance par intégration [MS-20]

11:30 Judith ROUSSEAU, *Université Paris Dauphine*, Don FRASER, *University of Toronto*

$p$ -values with nuisance parameters: A unified approach between Bayesian and frequentists • Quelques valeurs  $p$  en présence de paramètres de nuisance : une approche unifiée entre bayésiens et fréquentistes [MS-21]

12:00 Don FRASER, *University of Toronto*

What model information is appropriate for the Bayesian paradigm • Quelle est l'information appropriée sur le modèle pour le paradigme bayésien [MS-22]

## 11:00 – 12:30 Session 02E

211

The Joint Canada/United States Survey of Health: Analytic results and methodological issues

L'enquête conjointe Canada/États-Unis sur la santé : Résultats analytiques et questions méthodologiques

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Jane F. GENTLEMAN *National Center for Health Statistics*

11:00 Jane GENTLEMAN, Debra L. BLACKWELL & Michael E. MARTINEZ, *National Center for Health Statistics, CDC*

Obesity rate differences between American and Canadian adults: Findings from the Joint Canada/United States Survey of Health • Les différences des taux d'obésité entre les adultes américains et canadiens : Constatations de l'enquête conjointe Canada/États-Unis sur la santé [MS-23]

11:30 Jean-Marie BERTHELOT, *Statistics Canada*, Nancy Ross, *McGill University & Statistics Canada*

Health disparities in the U.S. and Canada • Les disparités de santé aux États-Unis et au Canada [MS-24]

12:00 Stephen BLUMBERG, Karen DAVIS, Meena KHARE & Michael MARTINEZ, *National Center for Health Statistics*

The effect of survey follow-up on nonresponse bias • L'effet du suivi après sondage sur le biais de non réponse [MS-25]

## 11:00 – 12:30 Session 02F

241

Current advancement in statistical genetics

Progress récents en génétique statistique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Xin GAO *York University*

11:00 Hongyu ZHAO & Ning SUN, *Yale University School of Medicine*

Genetic association studies in isolated populations • Études d'associations génétiques dans des populations isolées [MS-26]

11:30 Shili LIN & Zhenqiu LIU, *Ohio State University*

Multilocus LD measures and tagSNP selection with generalized mutual information • Mesure de déséquilibre de liaison et sélection de tagSNP avec une information mutuelle généralisée [MS-27]



12:00 Shelley BULL, *University of Toronto*, Joanna BIERNACKA, *Cambridge Institute for Medical Research*, Lei SUN, *University of Toronto*, Long Yang WU, *Samuel Lunenfeld Research Institute*

Improving the validity and efficiency of statistical genetic analysis • Amélioration de la validité et de l'efficacité des analyses génétiques statistiques [MS-28]

## 11:00 – 12:15 Session 02G

217

Probability

Probabilité

Contributed Paper Session • Séance de communications libres

Chair • Président : Pritan Ranjan (Simon Fraser University)

11:00 Rafal KULIK, *University of Ottawa*

Rate of convergence in invariance principle for associated sequences • Taux de convergence du principe d'invariance pour des séquences dépendantes [MS-29]

11:15 Supranee LISAWADI & Andrei VOLODIN, *University of Regina*

On the Gut's general weak law of large numbers • Sur la loi faible des grands nombres générale de Gut [MS-30]

11:30 Janusz KAWCZAK, *University of North Carolina at Charlotte*

Weak convergence of nilpotent Markov chain • Convergence faible d'une chaîne de Markov impotente [MS-31]

11:45 Juan ALVAREZ & Chris SOTEROS, *University of Saskatchewan*

A second order Morita approximation to random copolymer localization • Une approximation de Morita du deuxième ordre pour la localisation aléatoire de copolymère [MS-32]

12:00 Simo PUNTANEN & Ka Lok CHU, *University of Tampere, Finland*, Jarkko ISOTALO & George P. H. STYAN, *McGill University*

On the Watson efficiency in a weakly singular linear model • Sur l'efficacité de Watson dans un modèle linéaire faiblement singulier [MS-33]

## 12:30 – 13:30 Dîner • Lunch

Marquis Hall

## 13:30 – 15:00 Session 03A

241

Special Invited Session of the Biostatistics Section

Séance spéciale de l'invité du groupe de biostatistique

Special Session • Conférence spéciale

Biostatistics Section • Groupe de biostatistique

Organizer and Chair • Responsable et président : Gordon FICK *University of Calgary*

13:30 Gerald VAN BELLE, *University of Washington*

Biostatistics as narrative • Narration sur la biostatistique [MS-34]

**13:30 – 15:00 Session 03B****214**

Quality improvement

Amélioration de la qualité

Invited Paper Session • Présentations sur invitation

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

Organizer and Chair • Responsable et président : François PAGEAU *SNC Technologies*13:30 Jock MACKAY & Stefan STEINER, *University of Waterloo*

The power of a baseline investigation I • La puissance d'une étude de base I [MS-35]

14:00 Stefan STEINER & Jock MACKAY, *University of Waterloo*

The power of a baseline investigation II • La puissance d'une étude de base II [MS-36]

14:30 Gemai CHEN, *The University of Calgary*, Lingyun ZHANG, *Massey University, New Zealand*

Control charts for monitoring process mean and variability simultaneously • Diagrammes de contrôle pour suivre simultanément la moyenne et la variabilité d'un processus [MS-37]

**13:30 – 15:00 Session 03C****133**

Measurement error models

Modèles d'erreurs de mesure

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Liqun WANG *University of Manitoba*13:30 Paul GUSTAFSON, *University of British Columbia*

Bayesian modelling for measurement error problems • Modélisation bayésienne pour des problèmes d'erreurs de mesures [MS-38]

14:00 Susanne SCHENNACH, *University of Chicago*

Instrumental variable estimation of nonlinear errors-in-variables models • Estimation instrumentale de variable pour des modèles non linéaires d'erreurs dans les variables [MS-39]

14:30 Liqun WANG, *University of Manitoba*, Cheng HSIAO, *University of Southern California*

Identifiability and estimation of semiparametric nonlinear errors-in-variables models • Identifiabilité et estimation de modèles semi paramétriques non linéaires d'erreurs dans les variables [MS-40]

**13:30 – 15:00 Session 03D****211**

Recent development of finite mixture models

Développements récents dans les mélanges de modèles finis

Invited Paper Session • Présentations sur invitation

Chair • Présidente : Cindy FU *York University*Organizer • Responsable : Jiahua CHEN *University of Waterloo*13:30 Jiayang SUN, *Case Western Reserve University*, Jiahua CHEN, *University of Waterloo*

Adaptive estimation for mixture distributions • Estimation adaptative de mélanges de distributions [MS-41]

14:00 Hongtu ZHU, *Columbia University and New York Psychiatric Institute*, Heping Zhang ZHANG, *Yale University*

Hypothesis testing in mixture regression models • Tests d'hypothèses pour des mélanges de modèles de régressions [MS-42]

14:30 Abbas KHALILI & Jiahua CHEN, *University of Waterloo*

Variable selection in finite mixture of regression models • Sélection de variables pour un mélange fini de modèles de régression [MS-43]

### 13:30 – 14:30 Session 03E

217

Elaboration of a calibration strategy: a case study

Élaboration d'une stratégie de calage : une étude de cas

Contributed Paper Session • Séance de communications libres

Survey Methods Section • Groupe de méthodologie d'enquête

Chair • Président : Milorad KOVACEVIC *Statistics Canada*

13:30 Johanne TREMBLAY & Michel LATOUCHE, *Statistique Canada*

Overview of the harmonised calibration strategy for income statistics • Aperçu de la stratégie de calage harmonisée pour les statistiques du revenu [MS-44]

13:45 Sylvie AUGER & Johanne TREMBLAY, *Statistique Canada*

Evaluation of the quality of the demographic estimations and the tax data used in the calibration of some different surveys in Statistics Canada (STC) • Évaluation de la qualité des estimations démographiques et des données fiscales utilisées lors du calage de différentes enquêtes de Statistique Canada (STC) [MS-45]

14:00 Sylvie LAROCHE & Michel LATOUCHE, *Statistique Canada*

Calibration strategy of the Survey of Labour and Income Dynamics • Stratégie de calage de l'Enquête sur la dynamique du travail et du revenu [MS-46]

14:15 Michel LATOUCHE & Johanne TREMBLAY, *Statistique Canada*

Lessons to be learned from the use of administrative data at the weighting stage • Leçons à tirer de l'utilisation de données administratives au stade de la pondération [MS-47]

### 13:30 – 14:30 Session 03F

212

Bayesian methodology

Méthodologie bayésienne

Contributed Paper Session • Séance de communications libres

Chair • Président : Georges MONETTE *York University*

13:30 Alain DESGAGNÉ, *Université du Québec à Montréal & Drake University*, Jean-François ANGERS, *Université de Montréal*

Importance sampling with the generalized exponential power density • Simulations Monte Carlo avec fonction d'importance à l'aide de la densité GEP [MS-48]

13:45 ÉRIC MARCHAND, *Université de Sherbrooke et University of New Brunswick*, William E. STRAWDERMAN, *Rutgers University*

On the behaviour of Bayesian credible intervals for some restricted parameter space problems • Sur le comportement de divers estimateurs par intervalles pour des espaces paramétriques contraints [MS-49]

14:00 Michael EVANS & Hadas MOSHONOV, *University of Toronto*

Checking for prior-data conflict • Vérification de l'existence d'un conflit a priori-données [MS-50]

14:15 Hui SHEN & Will WELCH, *University of British Columbia*

Model selection by cross-validation in linear model • Sélection de modèle par validation croisée dans des modèles linéaires [MS-51]

### 13:30 – 15:00 Session 03G

200

General Meeting of SSC Accredited members (open)  
Assemblée générale des membres accrédités par la SSC

Special Session • Conférence spéciale

Initial Accreditation Committee • Comité initial sur accréditation

Organizer and Chair • Responsable et présidente : Judy-Anne CHAPMAN *Présidente du comité*

### 15:00 – 15:30 Pause-café • Coffee Break

### 15:30 – 17:00 Session 04A

241

Statistical climatology  
Climatologie statistique

Invited Paper Session • Présentations sur invitation

CAIMS

Organizer and Chair • Responsable et président : Samuel SHEN *University of Alberta*

15:30 Allan HOWARD & Allan HOWARD, *Agriculture and Agri-Food Canada*, Samuel SHEN, *University of Alberta*, Ralph WRIGHT, *Alberta Agriculture Food and Rural Development*

Use of climate data in a drought reporting system for agriculture • Utilisation de données climatiques dans un système de suivi des sécheresses pour l'agriculture [MS-52]

16:00 Hee-Seok OH, *University of Alberta*

Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptive method • Estimation de champs de température globaux à partir d'observations parsemées avec une méthode spatialement adaptative basée sur les ondelettes sphériques [MS-53]

16:30 Samuel SHEN, *University of Alberta*

Interpolating daily precipitation data and regression variance correction • Interpolation de données sur les précipitations journalière et correction par régression de la variance [MS-54]

**15:30 – 17:00 Session 04B****214**

Inference from multiple surveys and multiple frame surveys

Inférence à partir d'enquêtes multiples et d'enquêtes à cadres multiples

Invited Paper Session • Présentations sur invitation

Survey Methods Section • Groupe de méthodologie d'enquête

Organizer and Chair • Responsable et président : Narasimha PRASAD *University of Alberta*15:30 Sharon LOHR, *Arizona State University*, J.N.K. RAO, *Carleton University*

Bootstrap variance estimation for multiple frame surveys • Estimation bootstrap de la variance pour des sondages à plusieurs bases [MS-55]

16:00 Avinash SINGH, *Statistics Canada*, Shiyong WU, *RTI International*

Dual Frame Calibration for Possibly Dependent Samples under Different sampling Designs • Calibration pour deux bases de sondage avec des échantillons possiblement dépendants sous différents plans d'échantillonnage [MS-56]

16:30 Changbao WU, *University of Waterloo*, Jon N.K. RAO, *Carleton University*

Empirical likelihood inference for multiple surveys and multiple frame surveys • Inférence par vraisemblance empirique pour des sondages multiples et des sondages à bases multiples [MS-57]

**15:30 – 17:00 Session 04C****200**

Methodology on missing data in longitudinal studies

Méthodologie pour les données manquantes dans les études longitudinales

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Grace Y. YI *University of Waterloo*15:30 Xihong LIN, *University of Michigan*, Raymond CARROLL, *Texas A&M University*

Semiparametric estimation in general clustered/longitudinal data problems • Estimation semi paramétrique pour des problèmes généraux de données de type groupées longitudinales [MS-58]

16:00 Yannis JEMIAI, *Cytel Software Corporation*

Semiparametric methods for inferring treatment effects of outcomes defined only if a post-randomization event occurs • Méthodes semi paramétriques pour inférer les effets de traitement sur les résultats qui ne sont pas définis à moins qu'un événement de post-aléation se réalise. [MS-59]

16:30 Richard COOK & Grace YI, *University of Waterloo*

A missing data problem arising in orthopedic trials of blood thinners • Un problème de données manquantes dans des essais orthopédiques sur les éclaircisseurs sanguins [MS-60]

**15:30 – 17:00 Session 04D****133**

Spatial analysis of survival data

Analyse spatiale de données de survie

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Renjun MA *University of New Brunswick*15:30 Yi LI, *Harvard University*, Andrea Cook,

Spatial cluster detection for censored data • Détection de clusters spatiaux pour des données censurées [MS-61]

16:00 Sudipto BANERJEE, Freda COONER & Brad CARLIN, *University of Minnesota*

Flexible modelling of spatially varying cure rates for survival data • Modélisation flexible pour des données de survie avec des taux de guérison qui varient spatialement [MS-62]

16:30 Renjun MA, *University of New Brunswick*

Survival models for spatial data: a Poisson modelling approach • Modèles de survie pour des données spatiales : une approche par un modèle de Poisson [MS-63]

**15:30 – 16:45 Session 04E****211**

Dependent data analysis

Analyses des données dépendantes

Contributed Paper Session • Séance de communications libres

Chair • Président : Steve OVERDUIN *Simon Fraser University*15:30 Aberazzak MOUIHA & Nourddine RAIS, *Université de Fès, Maroc*

Another method for bootstrapping the arithmetic mean and the median when the observations are dependent • Une autre méthode pour faire du bootstrap sur la moyenne arithmétique et la médiane lorsque les observations sont dépendantes [MS-64]

15:45 Deepak SANJEL, Serge PROVOST & Ian MACNEILL, *University of Western Ontario*

On approximating the distribution of a statistics for detecting serial correlation • Sur l'approximation de la distribution d'une statistique permettant la détection de corrélation sérielle [MS-65]

16:00 Pingguo LU & Hao YU, *University of Western Ontario*

Asymptotic properties of quasi-maximum likelihood estimator of ARMA(P,Q)-GARCH(p,q) • Propriétés asymptotiques de l'estimateur du pseudo-maximum de vraisemblance d'un modèle ARMA(P,Q)-GARCH(p,q) [MS-66]

16:15 Chafik BOUHADDIOUI, *CIRANO*, Jean-Marie DUFOUR, *Université de Montréal*

A consistent test for independence between two infinite order cointegrated series • Tests Convergents d'indépendance de deux séries cointégrées autorégressives d'ordre infini [MS-67]

**15:30 – 17:00 Session 04F****212**

## Methods with applications

## Méthodes et applications

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Zeny FENG *University of Waterloo*15:30 Zhong LIU, James ZIDEK & Nhu LE, *University of British Columbia*

Bayesian melding of ozone level measurements and deterministic model output • Combinaison bayésienne de mesures du niveau d'ozone et de résultats de modèles déterministes [MS-68]

15:45 Daniel GILLIS & John HOLT, *University of Guelph*, Dominique CHARRON, *Public Health Agency of Canada*, David WALTNER-TOEWS, *University of Guelph*, Abdel MAAROUF, *Environment Canada*

Spatial temporal analysis of waterborne disease risk as related to climate change • Analyse spatiale temporelle du risque de maladies d'origine hydrique en relation avec les changements climatiques [MS-69]

16:00 Kirill ANDREEV, *Queen's University*, Robert BOURBEAU, *University of Montreal*

Frailty modeling of Canadian mortality at advanced ages • Modélisation de la mortalité canadienne aux grands âges [MS-70]

16:15 Afisi ISMAILA, Angelo CANTY & Lehana THABANE, *McMaster University*

Comparison of Bayesian and frequentist approaches in modelling risk of preterm birth near the Sydney Tar Ponds, Nova Scotia, Canada • Comparaisons des approches bayésiennes et fréquentistes pour modéliser le risque d'accouchement prématuré à proximité du Sydney Tar Pond, Nouvelle-Écosse, Canada [MS-71]

16:30 Hedy JIANG, Lehana THABANE, Anne HOLBROOK, Kalpana NAIR & Lisa DOLOVICH, *McMaster University*

Analysis of priorities of patients living with diabetes • Analyse des priorités des patients vivant avec le diabète [MS-72]

16:45 Amparo CASANOVA-GURRERA, M. Begoña CAMPOS-BONILLA & Wilma PENZO-GIACCA, *Universitat de Barcelona*, Wendy LOU, *University of Toronto*

The challenge of teaching Biostatistics in a Faculty of Medicine • Les défis d'enseigner la biostatistique dans une faculté de médecine [MS-73]

**15:30 – 16:15 Session 04G****217**

Statistical genetics and genomics  
Statistique génétique et génomique

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Peter HOOPER *University of Alberta*

15:30 Angelo CANTY, *Toronto Hospital for Sick Children and McMaster University*

Selection of a cut-off for differentially expressed genes in the SAM method • Sélection d'un seuil pour des gènes à différentes expressions dans la méthode ASM [MS-74]

15:45 Mary LESPERANCE & Edward CHANG, *University of Victoria*, Karissa JOHNSTON, *University of British Columbia*, Carmen BAILEY & Caren HELBING, *University of Victoria*

A mixture model approach to background correction for cDNA arrays • Approche par mélange de modèles pour la correction du fond de réseaux d'ADN complémentaires [MS-75]

16:00 Haydar SENGUL, *University of Saskatchewan*

A survey of analyses of age-at-onset data in human genetics research • Sondage des méthodes d'analyses de l'âge au début de l'invalidité en recherche sur la génétique humaine [MS-76]

**17:00 – 18:30 Session 04H****214**

NSERC Workshop  
Atelier du CRSNG

Special Session • Conférence spéciale

Chair • Présidente : Carmen GERVAIS *NSERC/CRSNG*

Organizer • Responsable : Judie FOSTER *NSERC/CRSNG*



## Mardi 14 juin • Tuesday, June 14

---

### 09:00 – 10:30 Session 05

241

Gold Medal Address

Allocution de la Médaille d'or

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président : David BRILLINGER *University of California, Berkeley*09:00 Keith WORSLEY, *McGill University*

Correlation random fields, brain connectivity, and cosmology • Champs de corrélation aléatoires, connectivité de cerveau, et cosmologie [MS-77]

### 10:30 – 11:00 Pause-café • Coffee Break

### 11:00 – 12:30 Session 06A

133

Bayesian model selection

Sélection de modèles bayésiens

Invited Paper Session • Présentations sur invitation

Institute of Mathematical Statistics

Organizer and Chair • Responsable et présidente : Nicole LAZAR *University of Georgia*11:00 Marina VANNUCCI, *Texas A&M University*

Feature selection in model-based analysis of functional data • Sélection de caractéristiques dans l'analyse paramétrique de données fonctionnelles [MS-78]

11:30 Paul GUSTAFSON, *University of British Columbia*

Bayesian concept-varying feature selection for classification with partially labelled data • Sélection bayésienne de caractéristiques selon le concept pour la classification de données partiellement étiquetée [MS-79]

12:00 Andrew MERTON, Jennifer HOETING & Richard DAVIS, *Colorado State University*, Sandra THOMPSON, *Pacific Northwest National Lab*

Model selection for geostatistical models • Sélection de modèle pour des modèles géostatistiques [MS-80]

**11:00 – 12:30 Session 06B****217**

Extensions of policy-relevant analytical methods for complex survey data

Extensions des méthodes analytiques dépendant des politiques pour les données d'enquêtes complexes

Invited Paper Session • Présentations sur invitation

Survey Methods Section • Groupe de méthodologie d'enquête

Organizer and Chair • Responsable et présidente : Georgia ROBERTS *Statistics Canada*11:00 Irene LU, *York University*

A comparison of latent variable modeling approaches in social science research: A Monte Carlo simulation and an illustration using data from youth in transition survey • Comparaison des techniques de modélisation de variables latentes en recherche en sciences sociales : une simulation de Monte Carlo et une illustration en utilisant des données d'un sondage sur les jeunes en transition [MS-81]

11:30 Norberto PANTOJA GALICIA & Mary THOMPSON, *University of Waterloo*, Milorad KOVACEVIC, *Statistics Canada*

Bivariate density estimation from a longitudinal complex survey: example of testing for temporal association of job loss and divorce using SLID • Estimation d'une densité bivariée à partir d'un sondage longitudinal complexe : exemple de test pour l'association temporelle entre la perte d'emploi et le divorce en utilisant l'EDTR [MS-82]

12:00 Qunshu REN, *Carleton University*

Some advances in the use of marginal models with data from a longitudinal survey having a complex design • Quelques avancements dans l'utilisation des modèles marginaux avec des données provenant d'un sondage longitudinale avec un plan complexe [MS-83]

Discussant • Modérateur

Harold MANTEL, *Statistics Canada*

Discussant • Modérateur

Milorad KOVACEVIC, *Statistics Canada*

Discussant • Modérateur

Georgia ROBERTS, *Statistics Canada*

**11:00 – 12:30 Session 06C****214**

Statistical genomics: New platforms, new data, new analysis methods for high-throughput experiments

La génomique statistique : Nouvelles plateformes, nouvelles données, nouvelles méthodes d'analyse pour les expériences à haut débit.

Invited Paper Session • Présentations sur invitation

Biostatistics Section • Groupe de biostatistique

Organizer and Chair • Responsable et présidente : Karen KOPCIUK *University of Calgary*

11:00 Steve HORVATH, *University of California, Los Angeles*

Statistical methods for the analysis of tissue microarray data. • Méthodes statistiques pour analyser des données de micro-réseaux de tissus [MS-84]

11:45 Rafal KUSTRA, *University of Toronto*

Genomic databases and microarray data: One statistician's approach at integration. • Bases de données génomiques et données de microréseaux : l'approche d'un statisticien pour l'intégration [MS-85]

**11:00 – 12:30 Session 06D****200**

Computer network modeling

Modélisation de réseaux informatiques

Invited Paper Session • Présentations sur invitation

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

Organizer and Chair • Responsable et président : George MICHAILIDIS *University of Michigan*

11:00 Eric COPE, *UBC Sauder School of Business*

Regret and convergence bounds for continuum-armed bandits • Manque à gagner et frontière de convergence pour le problème du " Continuum-Armed Bandits " [MS-86]

11:30 Jaimyoung KWON, *Cal State East Bay*, Pravin VARAIYA, *University of California Berkeley*

Real-time origin-destination matrix estimation from partially observed packet/vehicle trajectories • Estimation de matrices origine-destination en temps réel à partir de trajectoires de paquets ou véhicules partiellement observées [MS-87]

12:00 George MICHAILIDIS, Lawrence EARL & Nair VIJAYAN, *University of Michigan*

Flexicast network tomography • Tomographie de réseaux à diffusion flexible [MS-88]

**11:00 – 12:30 Session 06E****241**

Statistics in cancer, AIDS, and other medical researches

Statistique pour le cancer, le SIDA, et autres recherches médicales

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Joan HU *Simon Fraser University*11:00 Marvin ZELEN, *Harvard University and the Dana-Farber Cancer Institute*, Sandra LEE & Hui HUANG, *Dana-Farber Cancer Institute*

Early detection of disease and the scheduling of examinations • Détection rapide de maladies et calendrier des examens [MS-89]

11:30 Jian HUANG, *University of Iowa*, Shuangge MA, *University of Washington*

Analysis of censored survival data with high-dimensional covariates • Analyse de données de survie censurées avec plusieurs covariables [MS-90]

12:00 Ping YAN, *Public Health Agency of Canada*

Probability distributions in infectious disease transmission models and risks of major outbreaks • Distributions de probabilité pour des modèles de transmission de maladies infectieuses et risques d'épidémies majeures [MS-91]

**11:00 – 12:30 Session 06F****211**

Statistics and finance

Statistique et finance

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Donna SALOPEK *University of New South Wales*11:00 Don McLEISH & Joe DiCESARE, *University of Waterloo*

Missing and incomplete data in finance • Données manquantes et incomplètes en finance [MS-92]

11:45 Donna SALOPEK, *University of New South Wales*

Self-financing in a fractional Brownian motion market • Autofinancement dans un marché de mouvement brownien fractionnel [MS-93]

**11:00 – 12:30 Session 06G****212**

Estimation

Estimation

Contributed Paper Session • Séance de communications libres

Chair • Président : Abbas KHALILI *University of Waterloo*11:00 Rohana KARUNAMUNI & Jingjing WU, *University of Alberta*

Minimum Hellinger distance estimation for a nonparametric mixture model • Estimation à distance d'Hellinger minimale pour un mélange de modèles non paramétriques [MS-94]

11:15 Reza PAKYARI, *Australian National University*

Nonparametric estimation of mixing proportion and component distributions in multivariate mixtures • Estimation non paramétrique de la proportion de mélange et distributions des composantes dans des mélanges multivariés [MS-95]

11:30 Alexandre LEBLANC & Liqun WANG, *University of Manitoba*

Second-order least-squares estimation in nonlinear regression • Moindres carrés du second ordre en régression non-linéaire [MS-96]

11:45 Elies KOUIDER, Mohammed EL-SAIDI, Michael COOPER & Nathaniel TYMES, *Ferris State University*

Estimating simple linear regression model parameters using two-stage ranked set sampling schemes • Estimation des paramètres d'un modèle de régression linéaire simple en utilisant un plan d'échantillonnage en deux phases d'ensembles ordonnés [MS-97]

12:00 Shojaeddin CHENOURI, *Carleton University*

A multivariate generalization of the Hodges-Lehmann estimator • Généralisation multivariée de l'estimateur de Hodges-Lehmann [MS-98]

12:15 Bashir KHAN, *Saint Mary's University*, A.K. Md. E SALEH, *Carleton University*

Estimation of regression parameters: Parallelism restrictions • Estimation des paramètres de régression : restrictions de parallélisme [MS-99]

## 12:30 – 13:30 Dîner • Lunch

**Marquis Hall**

## 13:30 – 15:00 Session 07A

**133**

Statistics and technology

Statistique et technologie

Special Session • Conférence spéciale

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

Organizer and Chair • Responsable et président : Stefan STEINER *University of Waterloo*

13:30 Jerry LAWLESS, *University of Waterloo*

Statistics and technology • Statistique et technologie [MS-100]

## 13:30 – 15:00 Session 07B

**212**

Limit theorems

Théorèmes limites

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Andrei VOLODIN *University of Regina*

13:30 Allan GUT, *Uppsala University*

Some remarks on the law of large numbers • Quelques remarques sur la loi des grands nombres [MS-101]

14:00 Uwe EINMAHL, *Vrije Universiteit Brussels*, Deli LI, *Lakehead University*

Some new LIL type results for variables with infinite variance • De nouveaux résultats de type LIL pour des variables avec variance infinie [MS-102]

14:30 Ulrich STADTMUELLER, *University of Ulm*

Tail behaviour of weighted sums of random variables • Comportement des queues d'une somme pondérée de variables aléatoires [MS-103]

### 13:30 – 15:00 Session 07C

241

Statistical methodology in clinical trial design

Méthodologie statistique dans la planification d'essais cliniques

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Jianrong WU *St. Jude Children's Hospital*

13:30 Xiaoping XIONG, *St. Jude Children's Research Hospital*, Ming TAN, *University of Maryland Greenebaum Cancer Center*, James BOYETT, *St. Jude Children's Research Hospital*

A sequential procedure that stops clinical trials without later reversal of the early conclusion • Une procédure séquentielle qui arrête un essai clinique sans renverser une conclusion précoce [MS-104]

14:00 Zhengqing LI, *Pfizer*

Confirmatory non-inferiority studies for binary outcomes • Études confirmatoires de non infériorité pour des résultats binaires [MS-105]

14:30 Jianrong WU, *St. Jude Children's Research Hospital*

Sample size determination for testing the mean of skewed distribution • Détermination de la taille de l'échantillon pour tester la moyenne de distributions asymétriques [MS-106]

### 13:30 – 15:00 Session 07D

200

Applications of mixture models in genetic linkage analysis

Application des mélanges de modèles en analyse de liaison génétique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Cindy FU *York University*

13:30 Heping ZHANG & Rui FENG, *Yale University*

Genetic analysis of ordinal traits • Analyse génétique de traits ordinaux [MS-107]

14:00 Yongzhao SHAO, Jing HAN & Mengling LIU, *New York University*, Wenbin LU, *North Carolina State University*

Likelihood ratio tests for mixture models with applications to genetic linkage analysis • Tests du ratio de vraisemblance pour des modèles de mélanges avec des applications aux analyses de liaisons génétiques [MS-108]

14:30 Surajit RAY, *University of North Carolina at Chapel Hill and SAMSI*, Bruce LINDSAY, *Pennsylvania State University*

On using quadratic risk to select high dimensional models • Utilisation du risque quadratique pour sélectionner des modèles à haute dimension [MS-109]

### 13:30 – 14:40 Session 07E

214

Case Study 1: Antidepressant Medication and Stability in Seniors

**Analyse de cas 1 : Sur les antidépresseurs et la stabilité chez les personnes âgées**

Case Study • Étude de cas

Organizer and Chair • Responsable et présidente : Alison GIBBS *University of Toronto*13:30 Amy PETERS, *Centre for Studies in Aging at Sunnybrook and Women's College Health Sciences Centre*

Presentation of the data set • Présentation du jeu de données [MS-110]

13:40 Kimberley COLEMAN, Erika NAHM & Maja GRUBISIC, *Acadia University*, Sofia MOSESOVA & Xu WANG, *University of Waterloo*14:00 Yongtao ZHU, Zhushun YUAN & Meijie GUO, *University of Calgary*14:20 Olga MELNICHOUK, Maia LESOSKY, Gatot ILHAMTO & Maria WALIULLAH, *University of Guelph***13:30 – 14:45 Session 07F****211**

Applied survey methods

Méthodes d'enquête appliquées

Contributed Paper Session • Séance de communications libres

Survey Methods Section • Groupe de méthodologie d'enquête

Chair • Président : Harold MANTEL *Statistics Canada*13:30 Taraneh ABARIN, *University of Manitoba*

Application of two-phase sampling in education survey in Tehran • Application de l'échantillonnage en deux phases dans le sondage sur l'éducation à Téhéran [MS-111]

13:45 Myron KATZOFF, Lillian LIN & Michael MONSOUR, *National Center for Health Statistics/CDC*, Steven THOMPSON, *Simon Fraser University*

Sampling issues in national HIV behavioral surveillance of injecting drug users • Problèmes d'échantillonnage dans le programme national de surveillance du VIH pour les utilisateurs de drogues injectives [MS-112]

14:00 Richard LAROCHE & Barbara ARMSTRONG, *Statistique Canada*

Dual frame strategy in the survey of financial security: Do we need a high income sample? • L'utilisation de bases multiples dans l'Enquête sur la sécurité financière : Avons-nous besoin d'un échantillon de ménages à hauts revenus ? [MS-113]

14:15 Laurie REEDMAN, *Statistics Canada*

Improving the quality of the estimates of "special crops" • L'amélioration de la qualité des estimations des cultures spéciales [MS-114]

14:30 Caroline CAUCHON & Sylvie LAROCHE, *Statistique Canada*

Weighting strategy of the Survey of Labour and Income Dynamics (SLID) • Pondération de l'Enquête sur la Dynamique du travail et du revenu (EDTR) [MS-115]

**13:30 – 15:00 Session 07G****217****Advanced Regression Methods****Méthodes de regression avancées**

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Steven WANG *York University*13:30 Marco GERACI & Matteo BOTTAL, *University of South Carolina*

Use of auxiliary data in semiparametric regression with nonignorable missing responses • Utilisation de données auxiliaires en régression semi paramétriques avec des données manquantes non ignorables [MS-116]

13:45 Julie HORROCKS, *University of Guelph*

Prediction of binary outcomes from longitudinal data • Prévission d'un résultat binaire pour des données longitudinales [MS-117]

14:00 Manon FLEURY, *Public Health Agency of Canada*, John HOLT, *University of Guelph*, Dominique CHARRON, *Public Health Agency of Canada*, Brian ALLEN, *University of Guelph*, Abdel MAAROUF, *Environment Canada*

Time series models for the relationship of foodborne diseases and temperature • Modèles de séries chronologiques pour la relation entre la température et les maladies alimentaires [MS-118]

14:15 Tulay KORU-SENGUL, *University of Saskatchewan*, David S. STOFFER, *University of Pittsburgh*

The time-varying autoregressive model with covariates for analyzing longitudinal studies • Le modèle autorégressif à temps variable avec covariables pour analyser des études longitudinales [MS-119]

14:30 Ying ZHANG, *Acadia University*, Ian McLEOD, *University of Western Ontario*

Reparameterization of the boundary of the admissible region for ARMA models • Reparamétrisation de la frontière de la région d'admissibilité pour des modèles ARMA [MS-120]

14:45 Jiajia ZHANG & Yingwei PENG, *Memorial University of Newfoundland*

An alternative algorithm to the accelerated failure time model • Un algorithme alternatif pour le modèle de temps de vie accéléré [MS-121]

**15:00 – 15:30 Pause-café • Coffee Break**



## Mercredi 15 juin • Wednesday, June 15

---

### **09:00 – 09:45 Session 09A** **241**

Pierre Robillard Award Address

Allocution du prix Pierre Robillard

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président : Richard LOCKHART *Simon Fraser University*

### **09:45 – 10:30 Session 09B** **241**

Canadian Journal of Statistics Award Address

Allocution du prix de La Revue canadienne de statistique

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président : Richard LOCKHART *Simon Fraser University*

### **10:30 – 11:00 Pause-café • Coffee Break**

### **11:00 – 12:30 Session 10A** **214**

Imputation

Imputation

Invited Paper Session • Présentations sur invitation

Survey Methods Section • Groupe de méthodologie d'enquête

Organizer and Chair • Responsable et président : Eric RANCOURT *Statistics Canada*

11:00 Joseph SCHAFER, *Pennsylvania State University*

COM server for multiple imputation • Serveur COM pour l'imputation multiple [MS-122]

11:30 Thomas KRENZKE, David JUDKINS & Zizhong FAN, *Westat*

Vector imputation at Westat • Imputation de vecteurs chez Westat [MS-123]

12:00 Robert KOZAK, *Statistics Canada*

The Banff system for automated editing and imputation • Le système Banff pour l'éditage et l'imputation automatique [MS-124]

**11:00 – 12:30 Session 10B****200**

## Recent Advances in Lifetime Data

## Progrès récents pour les données de durée

Invited Paper Session • Présentations sur invitation

Biostatistics Section • Groupe de biostatistique

Organizer and Chair • Responsable et président : Richard COOK *University of Waterloo*11:00 Thierry DUCHESNE, *Université Laval*, Radu CRAIU, *University of Toronto*, Isabelle MICHAUD, *Université Laval*

Inference methods based on the EM algorithm for the competing risk model with masked causes of failure • Méthodes d'inférence basées sur l'algorithme EM pour le modèle des risques concurrents avec causes de pannes masquées [MS-125]

11:30 Doulgas SCHAUBEL, *University of Michigan*

A sequential stratification method to estimate the effect of a time-dependent treatment in the analysis of recurrent event data • Une méthode de stratification séquentielle pour estimer l'effet d'un traitement temporellement dépendant pour l'analyse de données d'événements récurrents [MS-126]

12:00 Bingshu CHEN, *NCI/NIH*, Richard COOK, *University of Waterloo*

The analysis of multivariate point processes with partially missing event types • L'analyse de processus ponctuels multivariés avec des types d'événements partiellement manquants [MS-127]

**11:00 – 12:30 Session 10C****133**

## Estimation in constrained parameter spaces

## Estimation dans des espaces de paramètres avec contraintes

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Eric MARCHAND *Université de Sherbrooke*11:00 Wolfgang BISCHOFF, *Catholic University Eichstaett-Ingolstadt*

A unified approach for minimax estimation of a bounded parameter • Approche unifiée pour l'estimation minimax d'un paramètre borné [MS-128]

11:30 Michael WOODROOFE, *University of Michigan*

Shape restricted estimation in the search for dark matter • Estimation restreinte sur la forme dans la recherche de la matière noire [MS-129]

12:00 FRANCOIS PERRON, *Université de Montréal*Estimation of the proportion  $p$  for a Binomial( $n, p$ ) distribution when  $p$  is near  $1/2$  • Estimation d'une proportion  $p$  pour une loi Binomiale( $n, p$ ) lorsque  $p$  est près de  $1/2$  [MS-130]

**11:00 – 12:30 Session 10D****217**

## Graphical Gaussian models

## Modèles gaussiens graphiques

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Hélène MASSAM *York University*11:00 Thomas RICHARDSON, *University of Washington*, Mathias DRTON, *University of California Berkeley*

Binary models for marginal independence • Modèles binaires pour l'indépendance marginale [MS-131]

11:30 Milan STUDENY, *Academy of Sciences of the Czech Republic*

A glance at graphical modelling in statistics and artificial intelligence • Un bref regard sur la modélisation graphique en statistique et en intelligence artificielle [MS-132]

12:00 Hélène MASSAM, *York University*, Gérard LETAC, *Université Paul Sabatier*

A new family of conjugate priors for decomposable graphical Gaussian models • Une nouvelle famille de densités a priori conjuguées pour des modèles gaussiens graphiques décomposables [MS-133]

**11:00 – 12:30 Session 10E****241**

## Teaching statistics in the health and life sciences

## L'enseignement de la statistique dans les sciences de la santé et de la vie

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente : Wendy Lou *University of Toronto*11:00 Paul COREY, *University of Toronto*

Comparing apples to oranges - issues in confounding • Comparer des pommes et des oranges : problème d'effets confondants [MS-134]

11:30 Penny BRASHER, *University of Calgary*

Teaching statistics in the presence of heteroscedasticity • Enseigner les statistiques en présence d'hétéroscédasticité [MS-135]

12:00 Carl SCHWARZ, *Simon Fraser University*

An online system for teaching the design (and analysis) of experiments • Un système en ligne pour enseigner la planification et l'analyse d'expériences [MS-136]

**11:00 – 12:30 Session 10F****212****Model Selection and Robust Modelling****Sélection des modèles et la modélisation robuste**

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Taraneh ABARIN *University of Manitoba*11:00 J. N. K. RAO, *Carleton University*, Kalyan DAS, *Calcutta University, India* & Qunshu REN, *Carleton University*

Analysis of Longitudinal Bivariate Binary Data Under a Mixed Effects Model • Analyse de données longitudinales binaires bivariées pour un modèle à effet mixte [MS-137]

11:15 Peng ZHANG & Peter SONG, *University of Waterloo*, Annie QU, *Oregon State University*

Maximum likelihood estimation and inference in non-normal random effects models • Estimation du maximum de vraisemblance et inférence dans des modèles à effet aléatoire non normal [MS-138]

11:30 Lisa LIX, *University of Manitoba*

The Analysis of Multivariate Repeated Measures Data with Structured Covariances • Analyse de données de mesures répétées multivariées avec covariance structurée [MS-139]

11:45 Melanie BUITENDYK, Nicole FERKO, Michele KOHLI & Nancy RYAN, *Innovus Research Inc.*

Clustering countries according to socioeconomic and behavioural measures of disease risk • Classification de pays selon des mesures socio-économiques et comportementales du risque de maladies [MS-140]

12:00 Geneviève LEFEBVRE, Alain C. VANDAL & Russell STEELE, *McGill University*

Model selection for hierarchical linear regression models for longitudinal data • Sélection de modèles de régression linéaire hiérarchique pour données longitudinales [MS-141]

12:15 Afisi ISMAILA, *McMaster University*

Kuhnian perspectives applied to the increasing popularity of Bayesian approaches in Health Care research: Is it a paradigm shift? • Perspectives de Kuhn appliquées à la popularité croissante de l'approche bayésienne en recherche sur la santé : est-ce que ce changement est un paradigme? [MS-142]

**11:00 – 12:30 Session 10G****211****Sampling and Design Issues****Questions de sondage et de planification**

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Angelo CANTY *McMaster University*11:00 Yuanyuan LIANG & Keumhee Cho CARRIÈRE, *University of Alberta*

Two-treatment repeated measurement designs for a model with self and mixed carryover effects

- Designs de mesures répétées pour deux traitements pour un modèle avec des effets résiduels uniques et mixtes [MS-143]

11:15 Ali Reza FOTOUHI, *University College of the Fraser Valley*

Modelling left censored durations in analyzing repeated durations data • Modélisation de durées censurées à gauche dans l'analyse de données de durées répétées [MS-144]

11:30 Ram YADAV & Padam SINGH, *Indian Council of Medical Research*

Sampling schemes for studying growth of children • Plans d'échantillonnage pour étudier la croissance des enfants [MS-145]

11:45 Sunita GHOSH & Punam PAHWA, *University Of Saskatchewan*

Asthma Prevalence and its relationship with age, sex, and location: A National Population Health Survey study: 1994-2001 • La prévalence de l'asthme et sa relation avec l'âge, le sexe et la localisation : une étude sur l'Enquête nationale de santé publique 1994-2001 [MS-146]

## 12:30 – 13:30 Dîner • Lunch

**Marquis Hall**

## 13:30 – 15:00 Session 11A

**133**

Special Invited Session of the Survey Section

Séance spéciale de l'invité du groupe d'enquête

Special Session • Conférence spéciale

Survey Methods Section • Groupe de méthodologie d'enquête

Organizer and Chair • Responsable et président : Changbao Wu *University of Waterloo*

13:30 Randy SITTER, *Simon Fraser University*

Bootstrapping in complex surveys • Bootstrap pour des sondages complexes [MS-147]

Discussant • Modérateur

J.N.K. RAO, *Carleton University*

## 13:30 – 15:00 Session 11B

**214**

Case Study II: The National Longitudinal Study on Children and Youth (NLSCY)

Étude de cas II : L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ)

Case Study • Étude de cas

Organizer and Chair • Responsable et présidente : Peggy NG *York University*

13:30 Peggy NG, *York University*

Presentation of the Data Set • Présentation du jeu de données [MS-148]

13:36 Yuanyuan LIANG, Qiaohao ZHU, Adeniyi ADEWALE, Irina DINU, Julie ZHANG, Kehu ZHOU & Xiufang YE, *University of Alberta*

13:48 Zhang LIU, Guohua YAN, Hui SHEN & Jixin LIU, *University of British Columbia*

14:00 Nancy LEBLANC & Cindy UBARTAS, *Université de Montréal*

14:12 Julien BÉRARD-CHAGNON & Joanne CHARLEBOIS, *Université de Montréal*

14:24 Lisa Wen LE, Matthew KOWGIER, Lucia MIREA & Hadassa BRUNSWIG, *University of Toronto*

14:36 Ivan CARILLO, Christina Chor-Yan CHU, Wanhua SU & Xinlei XIE, *University of Waterloo*

14:48 Ye SUN, Rao FU & Hsiao-Hsuan Alice WANG, *York University*

### 13:30 – 15:00 Session 11C

241

Analysis of gene expression data

L'analyse des données d'expression génétique.

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Wenqing HE *University of Western Ontario*

13:30 Jennifer BRYAN, *University of British Columbia*

Statistical methods for gene clustering from high-throughput data • Méthodes statistiques pour la classification de gènes à partir de données à haute résolution [MS-149]

14:00 Steven WANG, *York University*

Weighted likelihood estimation with application to micro-array data • Estimation par vraisemblance pondérée avec applications pour les données de microréseaux [MS-150]

14:30 Wenqing HE, *University of Western Ontario*

Gene selection and prediction of survival information under the Cox proportional hazards model • Sélection de gènes et prévision de l'information de survie sous le modèle de taux de panne proportionnel de Cox [MS-151]

### 13:30 – 14:30 Session 11D

217

Business and industrial statistics methods and applications

Méthodologie et applications de la statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

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

Chair • Président : Jock MacKAY *University of Waterloo*

13:30 Sofia MOSESOVA, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, Jock MacKAY & Stefan STEINER, *University of Waterloo*

Mining functional data • Exploration de données fonctionnelles [MS-152]

13:45 Xu (Sunny) WANG, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, William WELCH, *University of British Columbia*

A new mixture discriminant model for drug discovery data • Un nouveau modèle discriminant de mélanges pour des données de recherche de nouveaux médicaments [MS-153]

14:00 Cody HYNDMAN, *University of Waterloo*

Affine futures and forward prices • Contrat à termes et évaluation postérieures affines [MS-154]

14:15 Adebolanle ADEDIGBA, *University of Saskatchewan*

Statistical distributions for service times • Distributions statistiques pour des temps d'appels de service [MS-155]

### 13:30 – 15:00 Session 11E

200

#### Clinical Trials and Related Issues

#### Essais cliniques et questions associées

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Yun Yee CHOI *University of Waterloo*

13:30 Eleanor PINTO, *University of Toronto & Hospital for Sick Children*, Andrew WILLAN,

Efficient estimation of semi-parametric models for censored cost-effectiveness data • Estimation efficace de modèles semi paramétriques pour des données censurées coût-efficacité [MS-156]

13:45 Kevin THORPE, *University of Toronto*, Robin ROBERTS, *McMaster University*, Stefan HOHNLOSER, *J.W. Goethe University*, Paul DORIAN, *St. Michael's Hospital*, Stuart CONNOLLY & Michael GENT, *McMaster University*

Competing risks methodology enhances interpretation of the defibrillator in acute myocardial infarction trial (DINAMIT) • La méthodologie de risques concurrents améliore l'interprétation du défibrillateur lors d'essais sur les infarctus du myocarde aigu (DINAMIT) [MS-157]

14:00 Gabriel ESCARELA, *Universidad Autonoma Metropolitana, Mexico City*

The semiparametric mixture model for competing risks data • Modèle semi-paramétrique de mélanges pour des données de risques concurrents [MS-158]

14:15 Gabriel ESCARELA, *Universidad Autonoma Metropolitana*

A copula model for the Markov regression of time series count data • Un modèle de copules pour la régression markovienne de données de comptage de séries chronologiques [MS-159]

14:30 Xiaohui (Helen) GUO, Lehana THABANE, Kullathorn THEPHAMONGKHOL & George BROWMAN, *McMaster University*, Ian HODSON, *McMaster University*

Bayesian meta-analysis of trials of chemotherapy with radiotherapy in the management of patients with newly diagnosed locally advanced squamous cell or undifferentiated nasopharyngeal cancer • Méta-analyse bayésienne d'essais de chimiothérapie avec radiothérapie pour la gestion de patients nouvellement diagnostiqués de cellules squameuses localement avancées ou d'un cancer rhinopharyngien non différencié [MS-160]

14:45 Rong CHU, James DOUKETIS, Lehana THABANE & Kurl CUDDY, *McMaster University*

Assessment of the quality of reporting of obesity trials. A systematic review. • Évaluation de la qualité de diffusion d'essais sur l'obésité, une revue systématique [MS-161]

**13:30 – 14:30 Session 11F****212**

## Correlated Discrete Data

## Données discrètes corrélées

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Rohana KARUNAMUNI *University of Alberta*13:30 Rashid AHMED, Steve BROWN & Paul McDONALD, *University of Waterloo*

Use of Markov transition model to analyze longitudinal smoking (WSSP3) data • Utilisation du modèle de transition markovien pour analyser des données longitudinales sur le tabagisme (WSSP3) [MS-162]

13:45 Farouk NATHOO & Charmaine DEAN, *Simon Fraser University*

Mixed Markov models for spatiotemporal multi-state processes • Modèles markoviens mixtes pour des processus spatio-temporels à plusieurs états [MS-163]

14:00 Bethany GIDDINGS, Richard COOK & Grace YI, *University of Waterloo*

Panel event history data with incompletely observed variables • Données d'événements historiques de panel avec des variables observées incomplètes [MS-164]

14:15 Marc FREDETTE, *HEC Montréal*

Calibrating prediction intervals for recurrent events • Calibrage d'intervalle de prédiction d'événements récurrents [MS-165]

**15:00 – 15:30 Pause-café • Coffee Break****15:30 – 17:00 Session 12A****241**

## CRM-SSC Award Address

## Allocution du lauréat du Prix CRM-SSC 2005

Special Session • Conférence spéciale

Organizer and Chair • Responsable et présidente : Charmaine DEAN *Simon Fraser University***15:30 – 17:00 Session 12B****214**

## Design of experiments

## Planification d'expériences

Invited Paper Session • Présentations sur invitation

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

Organizer and Chair • Responsable et président : Randy SITTER *Simon Fraser University*15:30 Derek BINGHAM, *Simon Fraser University*, Elizabeth KELLY & David OLIVAS, *Los Alamos National Laboratories*, Randy SITTER, *Simon Fraser University*

Designing experiments for multi-stage processes • Planification d'expériences pour des processus multi-niveaux [MS-166]



16:00 George MICHAILIDIS, *University of Michigan*, Derek BINGHAM & Pritam RANJAN, *Simon Fraser University*

Designing efficient simulations for exploring features of response surfaces • Conception de simulations efficaces pour explorer les caractéristiques de surfaces de réponses [MS-167]

16:30 Crystal LINKLETTER, Randy SITTER & Boxin TANG, *Simon Fraser University*

Expected number of follow-up runs and cost for two-level fractional • Nombre de séquences subséquentes espéré et coûts pour un plan fractionnel à deux niveaux [MS-168]

## 15:30 – 17:00 Session 12C

217

Large deviations in statistics

Déviations importantes en statistique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Deli LI *Lakehead University*

15:30 S. Ejaz AHMED, *University of Windsor*, A.I. VOLODIN, *University of Regina*

The complete convergence rates and the law of logarithm for the dependent bootstrap of the mean • Taux de convergence complet et loi du logarithme pour la moyenne bootstrap dépendante [MS-169]

16:00 Andre DABROWSKI, *University of Ottawa*, Rafal KULIK, *University of Ottawa & Wroclaw University*, Gail IVANOFF, *University of Ottawa*

Poisson limits for empirical point processes • Limites pour des processus ponctuels empiriques [MS-170]

16:30 Mahmoud ZAREPOUR, *University of Ottawa*, Hemant ISHWARAN, *Cleveland Clinic Foundation*, L.F. JAMES, *Hong Kong University of Science and Technology*

Asymptotic results for an  $m$  out of  $n$  Bayesian bootstrap • Résultats asymptotiques pour un bootstrap bayésien de  $m$  éléments dans  $n$  [MS-171]

## 15:30 – 17:00 Session 12D

133

Model and design based inference for sample surveys

Inférence fondée sur le modèle et sur le plan pour les sondages

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président : Narasimha PRASAD *University of Alberta*

15:30 Milorad KOVACEVIC, David BINDER & Georgia ROBERTS, *Statistics Canada*

How important is the informativeness of the sample design? • Quelle est l'importance de "l'informativité" du plan d'échantillonnage ? [MS-172]

16:00 Narasimha PRASAD & Subhash LELE, *University of Alberta*

Model robust inference in survey sampling • Inférence robuste basée sur des modèles en échantillonnage [MS-173]

16:30 Andre CYR, Alexander DAVIES & Claude GIRARD, *Statistics Canada*

IRT modeling for latent variable in complex surveys • La modélisation par la TRI de variables latente pour les enquêtes à plan complexe [MS-174]

**15:30 – 16:45 Session 12E****211**

## Theoretical survey methods

## Méthodes d'enquête théoriques

Contributed Paper Session • Séance de communications libres

Survey Methods Section • Groupe de méthodologie d'enquête

Chair • Président : Michel LATOUCHE *Statistique Canada*15:30 Robert PHILIPS, *Statistics Canada*

The outlier detection and modeling method proposed for the gst replacement project of the monthly wholesale and retail trade survey • Détection de valeurs aberrantes et méthode de modélisation pour le projet de remplacement du sondage mensuel sur les ventes aux détails et en gros par des données de la TPS [MS-175]

15:45 Avinash SINGH & Catalin DOCHITOIU, *Statistics Canada*

Estimating proportions from survey data by using estimating equations with random recentering • L'estimation des Proportions avec des Plans de Sondage Complexes, en Utilisant des Équations Estimantes Aléatoirement Recentrées [MS-176]

16:00 Cynthia BOCCI & Jean-François BEAUMONT, *Statistics Canada*

A refinement of the regression composite estimator in the labour force survey • Une alternative à l'estimateur par la régression composite à l'enquête sur la population active [MS-177]

16:15 Mahmoud TORABI & J. N. K. RAO, *Carleton University*

Mean squared error estimation for small area means using survey weights • Estimation par l'erreur quadratique moyenne pour la moyenne sur un petit domaine en utilisant les poids d'échantillonnage [MS-178]

16:30 Harold MANTEL, Avi SINGH & Yong YOU, *Statistics Canada*

Use of generalized design effects for variance function modeling in small area estimation from survey data • L'utilisation des effets de design généralisés pour la modélisation des fonctions de variance dans l'estimation sur de petits domaines à partir de données de sondage [MS-179]

**15:30 – 16:45 Session 12F****212**

## Statistical modeling and applications

## Modélisation statistique et ses applications

Contributed Paper Session • Séance de communications libres

Chair • Président : Afisi ISMAILA *McMaster University*15:30 Jiguo CAO, Gregor FUSSMANN & Jim RAMSAY, *McGill University*

Fitting differential equations to data • Ajuster des équations différentielles à des données [MS-180]

15:45 Bruce JONES, *University of Western Ontario*, Debbie DUPUIS, *HEC Montréal*

Modelling extremal dependence - two actuarial problems • Modélisation de la dépendance de valeurs extrêmes - deux problèmes actuariels [MS-181]

16:00 Jeffrey PICKA & Travis STEWART, *University of New Brunswick*

Conduction-based statistics for description of anisotropy in sphere packings • Statistiques basées sur la conduction pour décrire l'anisotropie dans des arrangements de sphères [MS-182]

16:15 Giles HOOKER & James RAMSAY, *McGill University*

Identifying dynamical systems • Identification des systèmes dynamiques [MS-183]

16:30 David CAMPBELL & Jim RAMSAY, *McGill University*

Multi-signal wavelet coherence • Cohérence d'ondelettes multi-signaux [MS-184]

## 15:30 – 17:00 Session 12G

200

### Methods with Censored Data

### Méthodologie pour les données tronquées

Contributed Paper Session • Séance de communications libres

Biostatistics Section • Groupe de biostatistique

Chair • Président : Kelvin THORPE *University of Toronto*

15:30 Jahida GULSHAN & Alwell OYET, *Memorial University of Newfoundland*

Wavelet estimation of density and hazard rate for randomly right censored data • Estimation par ondelettes de la densité et du taux de panne pour des données censurés à droite aléatoirement [MS-185]

15:45 Yogendra CHAUBEY & Arusharka SEN, *Concordia University*, Pranab K. SEN, *University of North Carolina*

Smooth estimation of mean residual life under random censoring • Estimation lisse de la moyenne de vie résiduelle avec censure aléatoire [MS-186]

16:00 Hwashin SHIN, Timothy RAMSAY & Daniel KREWSKI, *University of Ottawa*

A nonparametric trend test for  $k$  samples of doubly censored data • Un test de tendance non paramétrique pour  $K$  échantillons de données doublement censurés [MS-187]

16:15 Yun Hee CHOI, David E. MATTHEWS & Grace Y. YI, *University of Waterloo*

Residual plots for accelerated life regression modelling of dependent bivariate time-to-event data • Graphiques des résidus pour la modélisation de temps de vie accéléré de données dépendantes et bivariées de temps jusqu'à un événement [MS-188]

16:30 Steve OVERDUIN & Michael STEPHENS, *Simon Fraser University*

Confidence intervals for quantiles of the lognormal distribution • Intervalles de confiance pour les quantile de distributions log-normales [MS-189]

16:45 Xuewen LU, *University of Calgary*, Wanrong LIU, *Hunan Normal University*

Local linear estimators of conditional variance functions under random censorship • Estimateurs linéaires locaux de fonctions de variance conditionnelle avec censure aléatoire [MS-190]

## Abstracts • Résumés

### Session 0 Dimanche 12 juin • Sunday, June 12, 13:00 – 18:00 Sheraton Conference Foyer

#### Contributed Poster Session Présentation d'affiches

[MS-1]

#### **Statistical analysis of association between pre-term delivery and low weight gain**

#### **Analyse statistique de l'association entre le faible gain de poids durant la grossesse et l'accouchement prématuré.**

Keramat NOURIJELYANI, *Tehran University of Medical Sciences*

**Objective:** To examine the effect of maternal weight gain on the risk of preterm birth. **Methods:** We used data on more than 3000 pregnant women receiving care at University Hospitals. Maternal weights and other important variables were recorded at each trimester of prenatal visit. Because of nature of this study many observations had at least one missing variable. Thus, statistical techniques for dealing with such missing values were examined. **Results:** Risks of preterm delivery were calculated using logistic regression with and without missing observations and results were compared.

**Objectif :** Analyser l'effet du gain en poids durant la grossesse sur le risque d'accoucher prématurément.

**Méthode :** Nous avons utilisé des données sur plus de 3000 femmes enceintes suivies dans des centres hospitaliers universitaires. Le poids durant la grossesse et d'autres variables importantes ont été enregistrées à chaque visite prénatale trimestrielle. À cause de la nature de cette étude, plusieurs observations ont au moins une valeur manquante. Par conséquent, les méthodes statistiques pour traiter de telles données manquantes ont été comparées.

**Résultats :** Le risque d'un accouchement prématuré a été estimé en utilisant une régression logistique avec et sans les observations manquantes et les résultats ont été comparés.

[MS-2]

#### **Effects of statistical method of assessment on prognosis of limited-stage small-cell lung cancer**

#### **Effets des méthodes d'analyse statistique sur le pronostic du cancer à petites cellules du poumon à niveau limité**

Judy-Anne CHAPMAN, *NCIC Clinical Trials Group, Queens University*, Edward YU, *University of Western Ontario*, Melanie ROYCE, *University of New Mexico*, Patricia TAI, *University of Saskatchewan*

Prognosis and impact of prognostic factors are standardly assessed with respectively, Kaplan-Meier and Cox proportional hazards. However, there might be substantive differences in prognosis, if different statistical methods were used,

Les facteurs du pronostic et de l'impact du pronostic sont normalement évalués respectivement par les modèles proportionnels aléatoires de Kaplan-Meier et de Cox. Cependant, il peut y avoir des différences de pronostic substantielles si différentes méthodes statistiques d'analyse de sur-

e.g. Boag log-normal, Kaplan-Meier, Cox, and log-normal survival analysis. A cohort of 244 limited-stage small cell lung cancer patients accrued between 1981 and 1998, followed to the end of February, 2002 was used to compare 1-, 3-, and 5-year disease-specific survival obtained with the four methods. The lack of events after 1966 days precluded comparison after 5-years. Apparent prognosis for patients was substantively influenced by the statistical method of assessment.

vie sont utilisées, telles que Boag log-normale, Kaplan-Meier, Cox et log-normale. Une cohorte de 244 patients atteint du cancer à petites cellules du poumon à niveau limité recrutés entre 1981 et 1998 et suivis jusqu'à la fin de février 2002 a été utilisé pour comparer la survie spécifique à la maladie après 1, 3 et 5 ans avec quatre méthodes différentes. Le manque d'événements après 1966 jours a écarté les comparaisons après 5 ans. Les pronostics apparents pour certains patients sont substantiellement influencés par la méthode d'analyse statistique choisie.

### [MS-3]

#### Tests for equality of curves via wavelets

#### Tests d'égalité de courbes par les ondelettes

Pengfei GUO & Alwell OYET, *Memorial University of Newfoundland*

In our study, we develop tests for equality of response curves. We assume that the mathematical expression of the response curves are unknown. Consequently, we exploit the adaptive properties of wavelets to construct a wavelet representation of the curves. Then, we develop an approach which combines the False Discovery Rate (FDR) technique and the universal thresholding approach, widely used in data denoising, for detecting differences between the curves. We also discuss some methods based on the general F-test. We consider several examples under a variety of conditions such as equality of variances, nonconstant variance and unequal number of observations. The size and power performances of the tests are reported.

Dans notre étude, nous développons des tests d'égalité de courbes réponses. Nous supposons que l'expression mathématique des courbes réponses est inconnue. Conséquemment, nous exploitons les propriétés adaptatives des ondelettes pour construire une représentation d'ondelettes des courbes. Ensuite, nous développons une approche qui combine la technique du taux de faux positif et l'approche universelle de seuil, largement utilisée en débruitage de données, pour détecter les différences entre les courbes. Nous discutons également quelques méthodes basées sur le test-F généralisé. Nous considérons plusieurs exemples sous une variété de conditions telles que l'égalité des variances, variance non constante et un nombre inégal d'observations. Nous rapportons les performances de la taille et de la puissance des tests.

### [MS-4]

#### A randomized study to determine the effectiveness of interventions to improve participant protocol compliance in a large clinical study.

#### Une étude randomisée pour déterminer l'efficacité d'interventions pour améliorer le protocole de conformité des participants dans une importante étude clinique

Kayo WALSH, *University of Colorado at Denver*, Eduard GAMITO, Richard JONES, Shannon PRETZEL, Colin O'DONNELL & Sheryl OGDEN, *University of Colorado Health Sciences Center*

This study evaluates five interventions to improve protocol compliance in a large randomized cancer screening trial. Intervention effectiveness was tested with Chi-square and Fisher's exact tests. Markov modeling was used to determine the effects of other variables including age, gender, and previous attendance on compliance. Results indicate that the interventions had no statistically significant effect on compliance ( $p > 0.05$ ). The Markov

Cette étude évalue cinq interventions pour améliorer le protocole de conformité dans un important essai randomisé de dépistage du cancer. L'efficacité de l'intervention est testée par le test du khi-deux et le test exact de Fisher. Une modélisation markovienne est utilisée pour déterminer les effets d'autres variables incluant l'âge, le sexe et l'apparence antérieure de conformité. Les résultats montrent que les interventions n'ont aucun effet statistiquement significatif sur la conformité. ( $p > 0.05$ ). Le modèle markovien indique que

model indicated that participants who missed a screening appointment in a given year, the probability that they would come in the next year (the transition probability) was 0.6377 for females and 0.7072 for males ( $p = 0.0047$ ).

les participants qui ont manqué un rendez-vous de dépistage dans une année donnée ont une probabilité de revenir l'année suivante (la probabilité de transition) de 0.6377 pour les femmes et 0.7072 pour les hommes ( $p = .0047$ )

#### [MS-5]

### **Long-term effects of air pollution on lung function of children and adolescents in Canada Effets d'une Exposition Prolongée à la Pollution Atmosphérique sur la Capacité Respiratoire de Jeunes Canadiens**

Mireille GUAY, Betty EDWARDS & Mark RAIZENNE, *Santé Canada / Health Canada*

The effects of long-term exposure to air pollution on the lung function of 552 children and adolescents were examined in four Canadian cities. In 1990, lung function tests were administered to children in grades 4-5, questionnaires were completed and levels of different pollutants were measured. These children were tested again in 1998-1999. A repeated measures model, stratified by gender and including adjustments for age, weight, height, history of allergy and smoking, was fitted. The data suggest that long-term residence in communities with modest air pollution levels may impair normal lung growth and function and that females are at greater risk.

Une étude portant sur les effets d'une exposition prolongée à la pollution atmosphérique sur la capacité respiratoire de 552 jeunes a été réalisée dans quatre villes canadiennes. Des tests de spirométrie ont été conduits en 1990 auprès d'élèves de quatrième et cinquième année ; des questionnaires ont été administrés et certains polluants mesurés. Les participants ont fait l'objet de réévaluation en 1998-1999. Un modèle de mesures répétées, stratifié selon le sexe et ajusté pour l'âge, le poids, la taille, le tabagisme et l'histoire d'allergie a été conçu. Il ressort qu'une exposition à long terme à des niveaux modérés de pollution atmosphérique pourrait porter atteinte au développement normal des poumons. L'étude révèle aussi un risque accru chez les individus de sexe féminin.

#### [MS-6]

### **The probability of knotting after a local strand in a self-avoiding polygon Probabilité de nouage après un brin local dans un polygone qui s'auto évite**

Michael SZAFRON & Christine SOTEROS, *University of Saskatchewan*

Motivated by experimental studies of enzyme action on DNA, there is interest in determining the expected knot probabilities when local strand-passages are implemented at random locations in a ring polymer. In order to investigate this problem, a self-avoiding polygon lattice model of a ring polymer was studied via Markov Chain Monte Carlo simulation, [which was used] to estimate the probability that the ring polymer had knot-type  $K$  after a local strand passage had occurred within the ring polymer. The model, some estimates of the knotting probabilities, and a comparison with the knotting probabilities estimated by Flammini, Maritan, and Stasiak (2004) for an off-lattice model will be presented.

Motivé par des études expérimentales de l'action d'un enzyme sur l'ADN, il y a un intérêt de déterminer les probabilités de noeuds espérées lorsque des passages de brins locaux sont implémentés aléatoirement dans un polymère en anneau. Pour investiguer ce problème, un modèle de treillis de polygones qui s'auto évitent d'un polymère en anneau a été étudié à l'aide de simulations de Monte Carlo avec chaînes de Markov. Ce modèle a été utilisé pour estimer la probabilité qu'un polymère en anneau ait un type de noeuds  $K$  après le passage d'un brin local à l'intérieur du polymère. Nous présentons le modèle, quelques estimateurs des probabilités de nouages et une comparaison avec les probabilités de nouages estimées par Flammini, Maritan, et Stasiak (2004) pour un modèle sans treillis.

**[MS-7]**

**Applications of statistical methods - principal component (PCA) and cluster analysis (CLA) in synchrotron-based ftir spectra analysis for feed/plant inherent structure research**  
**Applications de méthodes statistiques - Analyse en composantes principales (ACP) et classification automatique (CA) en analyse de spectre FTIR basé sur le synchrotron pour la recherche sur les structures inhérentes aliments/plantes**

P. Yu, *University of Saskatchewan*

Statistical methods - principal component analysis (PCA) and cluster analysis (CLA) were used to analyze synchrotron-based FTIR spectra obtained from various feed/plant inherent structures within intact tissues at a cellular and subcellular level, using the advanced synchrotron-based FTIR microspectroscopy as a novel approach. The spectra data of each feed/plant inherent structure were collected at the National Synchrotron Light Source (NSLS) in Brookhaven National Laboratory (BNL), US Dept of Energy (NSLS-BNL, NY, USA). Both PCA and CLA methods gave satisfactory results and are conclusive in showing that they can discriminate and classify inherent structures between and among feed/plant structures. These statistical analyses place synchrotron-based FTIR microspectroscopy at the forefront of those new potential techniques that could be used in rapid, non-destructive, and non-invasive screening of feed/plant structures in relation to quality and nutritive value of feeds/plant.

Des méthodes statistiques telles l'analyse en composantes principales (ACP) et la classification automatique (CA) ont été utilisées pour analyser les spectres FTIR obtenus à partir de différentes structures inhérentes aliments/plantes de tissus intacts à un niveau cellulaire et sous-cellulaire en utilisant la micro-spectroscopie avancée basée sur le synchrotron comme nouvelle approche. Les données sur les spectres de chaque structure inhérente aliments/plantes ont été amassées au National Synchrotron Light Source (NSLS) du Brookhaven National Laboratory (BNL), US Dept of Energy (NSLS-BNL, NY, USA). Les deux méthodes, soient l'ACP et la CA ont donnés des résultats satisfaisants et concluants quant à la discrimination des structures inhérentes aliments/plantes. Ces analyses statistiques placent la micro-spectroscopie basée sur le synchrotron à l'avant-plan des nouvelles techniques potentielles qui pourraient être utilisées dans le dépistage rapide, non destructif et non invasif des structures aliments/plantes en relation avec la qualité et la valeur nutritive des aliments/plantes.

**[MS-8]**

**Remedies for serial correlation and multiple local optima in period analysis of variable stars.**  
**Remède pour la corrélation sérielle et les optimums locaux multiples dans l'analyse périodique d'étoiles variables**

Peter HOOPER, *University of Alberta*

Problems arise when estimating the period and light curve (or periodic function) of a variable star due to serial correlation and consequent overfitting. My method of analysis employs adaptive logistic basis regression to estimate the light curve for each period, over a range of potential periods. The method produces an estimated loss function and a small set of locally optimal periods. To control overfitting, I use an AIC adjustment of the loss involving an effective sample size based on an estimated correlation model. I also suggest heuristics to assist in selecting a unique period.

Des problèmes surviennent lorsqu'on estime la période et la courbe de la lumière (ou la fonction périodique) d'une étoile variable, dû à à corrélation sérielle et au sur-ajustement. Ma méthode d'analyse utilise la régression adaptative à base logistique pour estimer la courbe de la lumière pour chaque période sur un éventail de périodes possibles. La méthode donne une fonction de perte estimée ainsi qu'un petit ensemble de périodes localement optimales. Pour contrôler le sur-ajustement, j'utilise un ajustement AIC de la perte impliquant une taille d'échantillon efficace basée sur un modèle de corrélation estimé. Je suggère également des heuristiques pour assister la sélection d'une période unique.

**[MS-9]****Application of multivariate poisson models for analysis of weed counts****Application de modèles de Poisson multivariés pour l'analyse du nombre de variétés de plantes**Chandima KARUNANAYAKE & William H. LAVERTY, *University of Saskatchewan*

Multivariate count data occurs in different variety of fields. For modeling such data, one may consider the multivariate Poisson distribution. In an agricultural field survey the count for the number of different weed species has been recorded for several fields. The distribution of counts depends on the underlying sequence of states, which is unobserved (i.e., hidden). Analysis of this data involves the estimation of the number of hidden states, Poisson means and covariances. The estimation of hidden states can be done using a model based clustering method, which involves a multivariate Poisson finite mixture or Markov-dependent finite mixture model.

Les données de dénombrement multivariées surviennent dans plusieurs domaines. Pour modéliser de telles données, on peut considérer la distribution de Poisson multivariée. Dans un sondage des champs en agriculture, le dénombrement des différentes espèces de plantes a été enregistré pour plusieurs champs. La distribution des nombres dépend de la séquence sous jacente des états, qui n'est pas observée (i.e. cachée). L'analyse de ces données implique l'estimation du nombre d'états cachés, des covariances et des moyennes Poisson. L'estimation des états cachés peut être effectuée en utilisant une méthode de cluster basée sur des modèles qui implique un modèle de mélange fini de Poisson multivarié ou bien un modèle de mélange fini de dépendance markovienne.

**[MS-10]****On the estimation of number of exceedances over a threshold****Estimation de la fréquence de dépassement d'un seuil**Qiaohao ZHU & Narasimha PRASAD, *University of Alberta*

This article considers applying the balanced one-way random effects model in estimating the number of exceedances over a given threshold and relative quantities. The classic procedure is based on normality and homogeneity assumptions. We compare several bootstrap procedures through numerical studies. Results on comparison of these bootstrap methods using a real data set on blood pressure from the International Prospective Primary Prevention Study in Hypertension are also reported.

Cet article considère l'application du modèle balancé d'effet aléatoire à un facteur pour estimer la fréquence de dépassement d'un seuil pour une certaine quantité. La procédure classique est basée sur des hypothèses de normalité et d'homogénéité. Nous comparons plusieurs procédures bootstrap à travers des études numériques. Nous présentons également les résultats des comparaisons de ces méthodes bootstrap en utilisant un jeu de données réelles sur la pression sanguine de l'étude internationale de prévention primaire prospective sur l'hypertension.

**[MS-11]****Expressions of life meaning among college students****Le sens de la vie selon des étudiants collégiaux**M. MIKET, W. H. LAVERTY, C. PRINGLE-NELSON & I. W. KELLY, *University of Saskatchewan*

This study examined the views of undergraduate students regarding what they considered to be contributing factors to a worthwhile or meaningful life. A survey was given to 132 students (Male=35, Female=97) where they rated, on a Likert scale (1 to 5) of 40 statements their agreement with each statement. The data were analyzed by cluster anal-

Cette étude examine la vision des étudiants sous-gradués sur ce qu'ils considèrent comme étant les facteurs contribuant à une vie intéressante ou qui fait du sens. Un sondage a été distribué à 132 étudiants (35 hommes et 97 femmes) et ceux-ci ont noté sur une échelle de Likert (1 à 5) leur niveau d'accord avec 40 énoncés. Les données ont été analysées à l'aide de méthodes de regroupement et d'autres techniques supplé-



ysis and other supplementary techniques. It was found that three clusters best described the data. One cluster comprised a religious group, and the other two clusters were non-religious, with one having characteristics of both the religious and non-religious cluster.

mentaires. Il a été trouvé que le nombre idéal de groupes est trois. Un groupe rassemble les gens ayant une vision religieuse, un autre une vision non-religieuse et, finalement, un dernier regroupe des gens ayant des caractéristiques des deux premiers groupes.

**[MS-12]**

**Case Study II: The National Longitudinal Study on Children and Youth (NLSCY)**

**Étude de cas II : L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ)**

Jing LIN & Miao FANG, *McMaster University*

**Session 01A**                      **Lundi 13 juin • Monday, June 13, 08:30 – 09:00**                      **241**  
 Welcome  
 Bienvenu

**Session 01B**                      **Lundi 13 juin • Monday, June 13, 09:00 – 10:30**                      **241**  
 SSC Presidential Invited Address  
 Allocution de l'invitée de la Présidente de la SSC

**[MS-13]**

**Deterministic and statistical models for turbulence: What could burgers have said to Kolmogorov?**

**Modèles déterministe et statistique pour la turbulence : qu'est-ce que Burgers aurait pu dire à Kolmogorov ?**

Barbara KEYFITZ, *Fields Institute and University of Houston*

Providing a satisfactory mathematical description of turbulence has challenged mathematicians for over seventy years. The contributions of Kolmogorov, using statistical techniques, provide the best candidates today for a rigorous theory. On the other hand, earlier work by Burgers resulted in a model that forms the basis of the modern theory of conservation laws. In this talk, I will contrast the deterministic approach pioneered by Burgers to the probabilistic models of Kolmogorov.

Trouver une description mathématique satisfaisante de la turbulence est un défi pour les mathématiciens depuis plus de soixante-dix ans. La contribution de Kolmogorov qui utilise des techniques statistiques est encore le meilleur candidat pour une théorie rigoureuse. D'un autre côté, des travaux plus anciens de Burgers ont résulté dans un modèle qui forme les bases de la théorie moderne des lois de la conservation. Dans cet exposé, je vais confronter l'approche déterministe de Burgers au modèle probabiliste de Kolmogorov.

**Session 02A**                      **Lundi 13 juin • Monday, June 13, 11:00 – 12:30**                      **214**

Statistics models for discrete longitudinal data

Modèles statistiques pour des données longitudinales discrètes

[MS-14]

**Some unusual applications of hidden Markov models**

**Quelques nouvelles applications des chaînes de Markov cachées**

Rachel ALTMAN, *Simon Fraser University*, Stefan REINKER, *University of Freiburg*, John STAUDENMAYER, *University of Massachusetts*

Hidden Markov models (HMMs) are a common tool for modelling time series of data. Recently, this class of models has been extended so as to apply to longitudinal as well as time series data. In this talk, we illustrate two novel ways in which HMM theory may be modified in order to solve practical problems. The first example concerns the use of HMMs to estimate chemical reaction rate parameters in small systems such as a biological cell. The second example involves the application of HMM theory to the problem of predicting individuals' physical activities based on their observed accelerometer counts.

Les chaînes de Markov cachées (CMC) sont des outils communs pour modéliser des données de séries chronologiques. Récemment, cette classe de modèle à été étendue pour être appliquée à des données longitudinales en plus des séries chronologiques. Dans cette présentation, nous illustrons deux nouvelles approches dans lesquelles la théorie des CMC peut être modifiée pour résoudre des problèmes pratiques. Le premier exemple concerne l'utilisation des CMC pour estimer les paramètres reliés aux taux de réaction chimique dans des petits systèmes tels des cellules biologiques. Le deuxième exemple implique l'application des CMC au problème de prédiction de l'activité physique des individus basée sur des mesures prises à l'aide d'un accéléromètre.

[MS-15]

**A marginalized model for combining state and national level survey data to estimating screening trends**

**Un modèle marginalisé pour combiner des données de sondages au niveau des états et au niveau national pour estimer les tendances de dépistage**

Diana MIGLIORETTI, *Group Health Cooperative*, Elizabeth BROWN, *University of Washington*

We propose a marginalized regression approach for combining state (BRFSS) and national (NHIS) level survey data to estimate colorectal cancer screening trends in the United States. Interest is in the cumulative incidence of screening, which is modeled using a mixed-influence diffusion of innovations model. The NHIS data is directly modeled using this national-level model. The BRFSS data is modeled using a state-level (conditional) model to account for clustering within states and missing data. This state-level model is linked to the national model for joint estimation. Extensions to account for survey design are discussed.

Nous proposons une approche par régression marginalisée pour combiner des données de sondages au niveau des états (BRFSS) et au niveau national (NHIS) pour estimer les tendances de dépistage du cancer du colon aux États-Unis. L'intérêt est l'incidence cumulative du dépistage qui est modélisée par un modèle d'influence mixte de la diffusion des innovations. Les données provenant des NHIS sont modélisées directement en utilisant le modèle au niveau national. Les données provenant des BRFSS sont modélisées par un modèle (conditionnel) au niveau des états pour tenir compte de la stratification à l'intérieur des états et des données manquantes. Ce modèle au niveau des états est lié au modèle national pour une estimation conjointe. Nous discutons également de certaines extensions pour tenir compte du plan d'échantillonnage.

[MS-16]

**Comparison of goodness-of-fit tests based on transition counts for misclassified, binary, alternating longitudinal data****Comparaison de tests de qualité d'ajustement basés sur les compteurs de transition pour des données longitudinales binaires, alternantes et mal classifiées**Rhonda ROSYCHUK, *University of Alberta*

The behaviour of transition count based goodness-of-fit tests in an alternating binary Markov model with misclassification is examined. Transition probabilities specify the state transitions for a process that is not directly observable. The state of an observable process, which may incorrectly classify the unobservable process state, is obtained at discrete time points. Misclassification probabilities capture two types of classification errors between the unobservable and observable process. Goodness-of-fit statistics are based on observed transition triplet and two-step pair counts, and for separate transition pair counts at each longitudinal observation number. Simulation studies are used to compare the goodness-of-fit tests.

Nous examinons le comportement des tests d'ajustement basés sur les compteurs de transition dans un modèle de Markov binaire alternatif avec mauvaise classification. Les probabilités de transition spécifient les transitions d'états pour un processus qui n'est pas directement observable. Les états d'un processus observable qui peut classifier incorrectement le processus non observé sont obtenus à des périodes de temps discrètes. Les probabilités de mauvaise classification capturent deux types d'erreur de classification entre les processus observé et non observé. Les statistiques d'ajustement sont basées sur les triplets de transition observés et les compteurs appariés à deux étapes, et pour des compteurs de transition appariés séparés à chaque numéro d'observation longitudinale. Des études de simulations sont utilisées pour comparer les tests d'ajustement.

**Session 02B****Lundi 13 juin • Monday, June 13, 11:00 – 12:30****200**

Strong asymptotics

Asymptotiques fortes

[MS-17]

**Asymptotic limits for coupled stochastic DE's****Limites asymptotiques pour des équations différentielles stochastiques couplées**Andrew HEUNIS, *University of Waterloo*, Seyed HASHEMI, *University of Manitoba*

We study a pair of stochastic DE's, one defining a "fast process", the other defining a "slow process". The fast process evolves unilaterally and is a stable but singular diffusion which conditions the drift and covariance coefficients of the slow process. Motivated by recent results of Pardoux and Veretennikov on nonsingular diffusions, we establish a solvability theory for the Poisson equation associated with the linear differential operator of a stable but singular diffusion, then use this to establish limit results for the slow process as the fast process is speeded up indefinitely.

Nous étudions une paire d'équations différentielles (ED) stochastiques, une définissant un processus rapide et l'autre un processus lent. Le processus rapide évolue unilatéralement et possède une diffusion stable mais singulière qui conditionne les coefficients de dérive et de covariance du processus lent. Motivé par des résultats récents de Pardoux et Veretennikov sur la diffusion non singulière, nous établissons une théorie de solvabilité pour l'équation de Poisson associée à l'opérateur différentiel linéaire d'une diffusion stable mais singulière. Nous utilisons ensuite cette théorie pour établir des résultats limites pour le processus lent quand le processus rapide est accéléré indéfiniment.

**[MS-18]****A strong invariance principle for associated random fields****Un principe d'invariance forte pour des champs aléatoires associés**Raluca BALAN, *University of Ottawa*

In this article we generalize Yu's (Ann. Probab. 24, 1996) strong invariance principle for associated sequences to the multi-parameter case, under the assumption that the covariance coefficient  $u(n)$  decays exponentially as  $n$  goes to infinity. The main tools that we use are: the Berkes and Morrow multi-parameter blocking technique, the Csörgő and Révész quantile transform method, and the Bulinski rate of convergence in the central limit theorem.

Dans cet article, on généralise le principe d'invariance forte de Yu (Ann. Probab. 24, 1996) pour des suites associées aux processus indexés par plusieurs paramètres, sous l'hypothèse de convergence exponentielle du coefficient de covariance  $u(n)$ , lors que  $n$  tend vers infini. Les outils principaux qu'on utilise sont : la technique de «blocking» à plusieurs paramètres de Berkes et Morrow, la méthode de la transformation «quantile» de Csörgő and Révész, et la vitesse de convergence de Bulinski dans le théorème limite centrale.

**Session 02C****Lundi 13 juin • Monday, June 13, 11:00 – 12:30****212**

How and how well do we train future consultants

Comment et avec quelle qualité entraînons-nous nos futurs consultants

**[MS-19]****How and how well do we train future consultants?****Comment, et comment bien entraînons nous les futures consultants**Jeanette O'HARA HINES, *University of Waterloo*, David BELLHOUSE, *University of Western Ontario*, John PETKAU, *University of British Columbia*, Heather THIESSEN, *London Health Sciences Centre*, Jamie MYLES, *Pfizer*

This session is designed to integrate an academic view of how statistical consulting should be taught with a critical view of current practices. It will start with descriptions of consulting courses currently being taught at three universities (Bellhouse, Petkau, O'Hara Hines), an assessment of the suitability of one such course as preparation for her first position as a statistician (Thiessen), and an assessment of the skills needed by graduated statistics students (Myles). This will then be followed by open discussion from the participants and the audience.

Cette session est conçue pour intégrer une vision académique sur comment la consultation statistique devrait être enseignée, avec une vision critique des pratiques actuelles. Nous allons commencer par décrire des cours de consultation qui sont présentement enseignés à trois universités (Bellhouse, Petkau et Ohara Hines), ensuite nous évaluons la validité d'un de ces cours comme préparation pour une première position comme statisticien (Thiessen) et finalement nous évaluons les qualités requises chez les étudiants gradués en statistique (Myles). Ceci sera suivi d'une discussion ouverte entre les participants et l'audience.

**Session 02D****Lundi 13 juin • Monday, June 13, 11:00 – 12:30****133**

Bayes and likelihood

Bayes et la vraisemblance

**[MS-20]****Elimination of nuisance parameters by integration****Élimination de paramètres de nuisance par intégration**Tom SEVERINI, *Northwestern University*

Consider a model with parameter  $\theta$  and component parameter of interest  $\psi = \psi(\theta)$  and consider the problem of constructing a likelihood for  $\psi$ . Many non-Bayesian methods of constructing such a likelihood have been proposed, such as marginal and conditional likelihoods, profile likelihoods, and modified versions of the profile likelihood. From the Bayesian point-of-view, construction of a likelihood for  $\psi$  is easily achieved by ‘integrating out’ the nuisance parameter with respect to a given prior distribution, although selection of this prior distribution may be difficult. In this talk, I consider the problem of constructing an integrated likelihood for use in non-Bayesian likelihood inference. A new type of prior density for this is proposed and its properties will be discussed.

Soit un modèle avec un paramètre  $\theta$  et un paramètre de composante d’intérêt  $\psi = \psi(\theta)$ . Considérons le problème qui consiste à construire la vraisemblance pour  $\psi$ . Plusieurs méthodes non bayésiennes pour construire une telle fonction de vraisemblance ont été proposées, telles que les vraisemblances marginales et conditionnelles, les vraisemblances de profile, et les versions modifiées des vraisemblances de profile. D’un point de vue bayésien, la construction d’une vraisemblance pour  $\psi$  est facilement accomplie en intégrant le paramètre de nuisance par rapport à une certaine distribution a priori. Cependant, la sélection de cette distribution a priori peut être difficile.

Dans cette présentation, je considère le problème de la construction d’une vraisemblance intégrée pour utilisation dans l’inférence de vraisemblance non-bayésienne. Un nouveau type de densité a priori est proposé et ses propriétés seront discutées.

**[MS-21]*****p*-values with nuisance parameters: A unified approach between Bayesian and frequentists****Quelques valeurs *p* en présence de paramètres de nuisance : une approche unifiée entre bayésiens et fréquentistes**Judith ROUSSEAU, *Université Paris Dauphine*, Don FRASER, *University of Toronto*

In many problems with nuisance parameters interesting *p*-values are hard to get, i.e. uniform or approximately so under the null hypothesis and consistent under the alternative. Berger and Bayarri have proposed different types of Bayesian *p*-values, among them the conditional posterior predictive *p*-value,  $p_{cpred}$ .

Here we prove that such *p*-values are asymptotically uniform under general conditions and with great precision. We also compare this Bayesian *p*-value by a frequentist one, based on some ancillary statistic.

The behaviour of the Bayesian *p*-value is then considered under the alternative.

We then illustrate these in a goodness of fit test.

ner des valeurs-*p* intéressantes en présence de paramètres de nuisance est souvent difficile. Berger et Bayarri ont proposé différents types de valeurs-*p* bayésiennes et notamment la valeur-*p* conditionnelle a posteriori. Nous démontrons que cette valeur-*p* est asymptotiquement uniforme sous l’hypothèse nulle, avec une grande précision et nous la comparons à une autre valeur-*p* construite à partir d’une statistique exhaustive.

Nous étudions aussi le comportement de cette valeur-*p* sous l’hypothèse alternative.

Un exemple est présenté dans le cadre d’un test d’adéquation.

**[MS-22]****What model information is appropriate for the Bayesian paradigm  
Quelle est l'information appropriée sur le modèle pour le paradigme bayésien**Don FRASER, *University of Toronto*

Both Bayesian and recent likelihood methodologies use the observed likelihood function; but they differ in what additional model information is added. For this we classify priors as subjective, objective or default and argue that the subjective should be introduced only on an individual basis and after a default analysis. We address the basis for choosing default priors; introduce three desiderata for such priors; and in a general context define neutral priors.

Tant les méthodologies bayésiennes que les méthodologies classiques récentes utilisent la fonction de vraisemblance observée. Cependant, ces méthodologies diffèrent en terme de l'information additionnelle sur le modèle. Pour cela, nous classons les densités a priori comme subjectives, objectives ou par défaut et nous argumentons le fait que les densités subjectives devraient être introduites seulement sur une base individuelle et suite à une analyse par défaut. Nous discutons les questions de bases pour choisir des densités a priori par défaut, nous introduisons trois desiderata pour de telles densités à priori et nous définissons les densités a priori neutres dans un contexte général.

**Session 02E****Lundi 13 juin • Monday, June 13, 11:00 – 12:30****211**

**The Joint Canada/United States Survey of Health: Analytic results and methodological issues**

**L'enquête conjointe Canada/États-Unis sur la santé : Résultats analytiques et questions méthodologiques**

**[MS-23]****Obesity rate differences between American and Canadian adults: Findings from the Joint Canada/United States Survey of Health****Les différences des taux d'obésité entre les adultes américains et canadiens : Constatations de l'enquête conjointe Canada/États-Unis sur la santé**Jane GENTLEMAN, Debra L. BLACKWELL & Michael E. MARTINEZ, *National Center for Health Statistics, CDC*

The 2002-03 Joint Canada/United States Survey of Health was a telephone survey that used the same questionnaire in both countries. One finding in a report released with the microdata by the survey co-sponsors—Statistics Canada and the National Center for Health Statistics—was that 21% of Americans and 15% of Canadians were obese, a gap primarily due to differences between obesity rates for American women (21.9) and Canadian women (13.9). The Winnipeg Sun highlighted that finding with a headline: "women fatter: study". We examine differences between Canadian and American obesity/overweight rates, focusing on differences that might explain observed differentials for women.

**[MS-24]****Health disparities in the U.S. and Canada  
Les disparités de santé aux États-Unis et au Canada**Jean-Marie BERTHELOT, *Statistics Canada*, Nancy Ross, *McGill University & Statistics Canada*

Using the Joint Canada/U.S. Survey of Health, we examined the size of the health disparities across the two countries. The analysis looked at income and wealth gradients in health. One-third (31%) of Americans in the lowest income group reported fair or poor health, compared with 23% among Canadians. Similar differences were observed for mobility limitations, smoking, obesity and unmet needs. No systematic differences were observed among the most affluent households on either side of the border. The social gradient in health and health care services is steeper in the U.S. than it is in Canada.

Nous avons utilisé l'Enquête Canada/États-Unis sur la santé pour examiner les disparités de santé entre les deux pays. L'analyse a considéré les gradients du revenu et de l'avoir sur la santé. Le tiers (31%) des Américains appartenant au groupe de revenus les plus faibles ont indiqué une santé passable ou faible, comparativement à 23% des Canadiens. Des résultats similaires ont été observés pour les restrictions de la mobilité, le tabagisme, l'obésité et les besoins non comblés. Aucune différence systématique n'a été observée entre les ménages les plus riches des deux côtés de la frontière. Le gradient social en matière de santé et de soins de santé est plus prononcé aux États-Unis qu'au Canada.

**[MS-25]****The effect of survey follow-up on nonresponse bias  
L'effet du suivi après sondage sur le biais de non réponse**Stephen BLUMBERG, Karen DAVIS, Meena KHARE & Michael MARTINEZ, *National Center for Health Statistics*

To achieve greater response rates for the U.S. sample, the 2002-2003 Joint Canada/United States Survey of Health fielded additional telephone calls and refusal conversion attempts for three months following the planned close of data collection. Final weighted survey estimates were compared with estimates from the U.S. National Health Interview Survey and with weighted estimates based on interviews completed prior to the planned close of data collection, interviews completed with no more than 10 dials, and interviews completed without refusal conversion efforts. Nonresponse bias remained constant despite efforts to increase response rates.

Pour atteindre des taux de réponse plus élevés pour l'échantillon américain dans le sondage conjoint Canada/États-Unis sur la santé, on a effectué des appels téléphoniques additionnels et d'autres essais pour convertir des refus pour les trois mois suivant la date prévue d'arrêt de collecte de données. Les estimations pondérées finales de sondage ont été comparées avec les estimations du U.S. National Health Interview Survey et avec les estimations pondérées basées sur les interviews complétés avant la date prévue d'arrêt de collecte, les interviews complétées avec au plus 10 appels et les interviews complétées sans efforts pour convertir les refus de réponse. Le biais de non réponse est resté constant malgré les efforts pour augmenter le taux de réponse.

**Session 02F****Lundi 13 juin • Monday, June 13, 11:00 – 12:30****241****Current advancement in statistical genetics  
Progress récents en génétique statistique****[MS-26]****Genetic association studies in isolated populations  
Études d'associations génétiques dans des populations isolées**Hongyu ZHAO & Ning SUN, *Yale University School of Medicine*

The identifications of genetic variants underlying complex diseases represent a major challenge to human geneticists. The use of isolated popula-

L'identification des variantes génétiques sous jacentes à des maladies complexes représente un défi majeur pour les généticiens du corps humain. L'utilisation de populations isolées

tions has turned out to be a very fruitful approach due to the reduced heterogeneity both in genetic background as well as in environmental exposures. However, conducting genetic association studies in an isolated population is complicated by the close relatedness among the cases and controls, and the exact degree of relatedness is often unknown due to the complex genealogy for a given population. In this presentation, we will discuss statistical methods that we have recently developed for genetic association studies in an isolated population. The proposed methods will be illustrated through their applications to a real genetic study.

s'est avérée être une approche intéressante dû à la réduction de l'hétérogénéité du passé génétique et de l'exposition à l'environnement. Cependant, effectuer des analyses d'associations génétiques dans une population isolée est compliqué à cause de la proximité des cas et des témoins. De plus, le degré exact de proximité est souvent inconnu à cause de la généalogie complexe d'une population donnée. Dans cette présentation nous discutons des méthodes statistiques que nous avons récemment développées pour les études d'associations génétiques dans une population isolée. Les méthodes proposées sont illustrées à travers leurs applications pour des études génétiques réelles.

#### [MS-27]

##### **Multilocus LD measures and tagSNP selection with generalized mutual information**

##### **Mesure de déséquilibre de liaison et sélection de tagSNP avec une information mutuelle généralisée**

Shili LIN & Zhenqiu LIU, *Ohio State University*

Linkage disequilibrium (LD) plays a central role in fine mapping and in characterizing haplotype blocks. Although classical LD measures have been successfully applied to block detection and SNP selection, their pairwise nature does not provide a direct characterization of joint LD among multiple loci. We propose a multilocus LD measure using Kullback-Leibler distance. Based on this measure and an entropy measure that characterizes haplotype diversity, we introduce a class of stepwise tagSNP selection algorithms. Applications to simulated and real data indicate satisfactory performance and demonstrate their ability for handling a SNP set too large to be amenable with other methods.

Le déséquilibre de liaison (DL) joue un rôle central dans l'indexation fine et dans la caractérisation des blocs haplotypes. Même si les mesures de DL classiques ont été appliquées avec succès à la détection de bloc et à la sélection de SNP, leur nature par paire ne permet pas une caractérisation directe de DL joint à travers des locus multiples. Nous proposons une mesure de DL multi locus en utilisant la distance de Kullback-Leibler. Basé sur cette mesure et une mesure d'entropie qui caractérise la diversité des haplotypes, nous introduisons une classe d'algorithmes de sélection de tagSNP. Des applications à des données réelles et simulées indiquent une performance satisfaisante et montrent l'habileté des algorithmes à gérer un ensemble de SNP trop grand pour être traité par d'autres méthodes.

#### [MS-28]

##### **Improving the validity and efficiency of statistical genetic analysis**

##### **Amélioration de la validité et de l'efficacité des analyses génétiques statistiques**

Shelley BULL, *University of Toronto*, Joanna BIERNACKA, *Cambridge Institute for Medical Research*, Lei SUN, *University of Toronto*, Long Yang WU, *Samuel Lunenfeld Research Institute*

The accuracy of gene localization and the validity of locus-specific effect estimates are major concerns in both fine-mapping and genome-wide linkage studies of complex diseases. I will discuss some of our recent work aimed at improving statistical methods for analysis of identical-by-descent allele sharing in affected siblings. In fine-mapping, we propose a GEE model that improves localiza-

La précision dans la localisation d'un gène et la validité de l'estimation de l'effet spécifique à un locus sont des préoccupations majeures en cartographie fine et dans les études de liaisons sur le génome pour des maladies complexes. Je discute de certains de nos récents travaux visant à améliorer des méthodes statistiques pour l'analyse du partage des allèles identiques par descendance dans des fratries atteintes de la maladie. En cartographie fine, nous proposons un mo-



tion and efficiency when two disease genes are present in one chromosomal region. Secondly, we apply resampling methods for bias reduction of locus-specific effect estimates when stringent significance criteria to control type I error in genome-wide studies induce selection bias.

dèle d'équations estimantes généralisées qui améliore la localisation et l'efficacité lorsque deux gènes de maladies sont présents dans une même région chromosomique. Deuxièmement, nous appliquons des méthodes de rééchantillonnage pour réduire le biais de l'estimation de l'effet spécifique à un locus lorsque des critères de signification stringents pour contrôler l'erreur de type I introduisent un biais de sélection dans des études sur le génome au complet.

## Session 02G

Lundi 13 juin • Monday, June 13, 11:00 – 12:15

217

### Probability Probabilité

[MS-29]

**Rate of convergence in invariance principle for associated sequences**

**Taux de convergence du principe d'invariance pour des séquences dépendantes**

Rafal KULIK, *University of Ottawa*

In this paper we obtain a rate of convergence in the strong invariance principle for associated random variables, extending Sakhanenko (1984) result for independent random variables. That rate is obtained under power-law decay of covariances. It solves the problem posed by Yu (Annals of Probability, 1996), who obtained a strong invariance principle for associated random variables assuming exponential decay of covariances. Moreover, we also obtain strong invariance principle for negatively associated sequences as well as for linear processes with associated coefficients.

Dans cet article, nous obtenons un taux de convergence pour le principe d'invariance fort pour des variables aléatoires dépendantes en étendant le résultat de Sakhanenko (1984) pour des variables aléatoires indépendantes. Ce taux est obtenu pour une décroissance des covariances suivant la loi de puissance. Il résout le problème posé par Yu (Annals of Probability, 1996) qui avait obtenu un principe d'invariance fort pour des variables aléatoires dépendantes en supposant une décroissance exponentielle de la covariance. De plus, nous obtenons également un principe d'invariance fort pour des séquences négativement dépendantes et pour des processus linéaires avec des coefficients dépendants.

[MS-30]

**On the Gut's general weak law of large numbers**

**Sur la loi faible des grands nombres générale de Gut**

Supraanee LISAWADI & Andrei VOLODIN, *University of Regina*

From the classical notion of uniform integrability of a sequence of random variables, a new concept of integrability (called h-integrability) for an array of random variables, concerning an array of constants, has been introduced by Ordonez Cabrera and Volodin, [J. Math. Anal. Appl., to appear]. This concept is weaker than other previous related notions of integrability. We consider the general weak law of large numbers in the sense of Gut [Statist. Probab. Lett. 14(1992) 49052] under this new condition of integrability.

À partir de la notion classique de l'intégrabilité uniforme d'une séquence de variables aléatoires, un nouveau concept d'intégrabilité (appelé h-intégrabilité) pour une matrice de variables aléatoires concernant une matrice de constantes, a été introduite par Ordonez, Cabrera et Volodin, (J. Math. Anal. Appl., à paraître). Ce concept est plus faible que certaines autres notions d'intégrabilité reliées. Nous considérons la loi faible des grands nombres générale au sens de Gut (Stat. & Prob. Letters 14(1992) 49-52) sous cette nouvelle condition d'intégrabilité.

**[MS-31]****Weak convergence of nilpotent Markov chain****Convergence faible d'une chaîne de Markov impotente**

Janusz KAWCZAK, *University of North Carolina at Charlotte*

Markoum of the covariance operator of the nilpotent Markov Chain. This is a special case of the general Markov Chain with the Doeblin condition. The chains with this property arise naturally in testing Random Number Generators (RNG) when it is understood that a finite sequence of digits is produced to form a word. We specialize our approach to the study of weak convergence with the improved estimation of the remainder term for the permutation type test statistics.

A complete analysis of the spectrum of the covariance operator is presented for the  $L^2$  space. We give an explicit decomposition of  $L^2$  into the direct sum of the eigenspaces associated to the eigenvalues of the covariance operator. This decomposition allows for the development of efficient computational algorithms when establishing the limiting distribution of the functional Central Limit Theorem generated by a general Markov Chain.

We also present some results of Berry-Esseen type for general Markov chains with and without nilpotent property.

Nous étudions le spectre de l'opérateur de covariance d'une chaîne de Markov impotente. Ceci est un cas spécial de la chaîne de Markov générale avec la condition de Doeblin. Les chaînes ayant cette propriété surviennent naturellement quand on teste des générateurs de nombres aléatoires (GNA) lorsqu'il est compris qu'une séquence finie de nombres est produite pour former un mot. Nous spécialisons notre approche pour l'étude de la convergence faible avec l'estimation améliorée du terme résiduel pour des statistiques de tests de type permutation.

Une analyse complète du spectre de l'opérateur de covariance est présentée pour l'espace  $L^2$ . Nous donnons une décomposition explicite de l'espace  $L^2$  en une somme directe des espaces propres associés aux valeurs propres de l'opérateur de covariance. Cette décomposition permet le développement d'algorithmes computationnels efficaces lorsqu'on veut établir la distribution limite du théorème limite central fonctionnel générée par une chaîne de Markov générale.

Nous présentons également quelques résultats de type Berry-Esseen pour les chaînes de Markov générales avec et sans la propriété impotente.

**[MS-32]****A second order Morita approximation to random copolymer localization****Une approximation de Morita du deuxième ordre pour la localisation aléatoire de copolymère**

Juan ALVAREZ & Chris SOTEROS, *University of Saskatchewan*

We use a Morita approximation to analyze Dyck path models of random copolymer localization at an interface between two immiscible liquids. A polymer can be either delocalized, with zero density of monomers at the interface, or localized, with a non-zero density of monomers at the interface.

We constrain the proportion of each type of comonomer and the proportion of pairs of neighboring comonomers. The boundaries in the phase diagram remain unchanged with respect to previous work using a single constraint. However, the average free energy in the localized regions yields a tighter upper bound on the average quenched free energy.

Nous utilisons une approximation de Morita pour analyser des modèles de trajectoires de Dyck pour la localisation aléatoire de copolymères sur la surface entre deux liquides immiscibles. Un polymère peut être soit délocalisé, avec une densité nulle de monomères à la surface, ou localisé avec une densité non nulle de monomères à la surface. Nous appliquons une contrainte sur la proportion de chaque type de comonomère et sur la proportion de paires de comonomères voisins. Les frontières dans le diagramme de phase restent inchangées par rapport aux travaux précédents qui ne considéraient qu'une seule contrainte. Cependant, l'énergie libre moyenne dans les régions localisées donne une borne supérieure plus stricte sur l'énergie libre moyenne amortie.

**[MS-33]****On the Watson efficiency in a weakly singular linear model****Sur l'efficacité de Watson dans un modèle linéaire faiblement singulier**

SIMO PUNTANEN & Ka Lok CHU, *University of Tampere, Finland*, JARKKO ISOTALO & GEORGE P. H. STYAN, *McGill University*

While considering the estimation of regression coefficients in a partitioned weakly singular linear model, Chu, Isotalo, Puntanen and Styan (2004) introduced a particular decomposition for the Watson efficiency of the ordinary least squares estimator. This decomposition presents the "total" Watson efficiency as a product of three factors. In this paper we give new insight into the decomposition showing that all three factors are related to the efficiencies of particular submodels or their transformed versions. Moreover, we prove an interesting connection between a particular split of the Watson efficiency and the concept of linear sufficiency. We shortly review the relation between the efficiency and specific canonical correlations. We also introduce the corresponding decomposition for the Bloomfield–Watson commutator criterion, and give a necessary and sufficient condition for its specific split

Tout en considérant l'estimations de coefficients de régression dans un modèle linéaire faiblement singulier partitionné, Chu, Isotalo, Puntanen et Styan (2004) ont introduit une décomposition de l'efficacité de Watson pour l'estimateur des moindres carrés ordinaire. Cette décomposition présente l'efficacité de Watson "totale" comme un produit de trois facteurs. Dans cet article, nous jetons un regard neuf sur cette décomposition, montrant que les trois facteurs sont reliés à l'efficacité de sous-modèles particuliers ou de leurs versions transformées. De plus, nous établissons un lien entre une décomposition particulière de l'efficacité de Watson et le concept d'exhaustivité linéaire. Nous faisons une brève revue de la relation entre cette efficacité et des corrélations canoniques spécifiques. Nous introduisons également la décomposition correspondante pour le critère commutateur de Bloomfield-Watson et donnons une condition nécessaire et suffisante pour sa séparation spécifique.

**Session 03A****Lundi 13 juin • Monday, June 13, 13:30 – 15:00****241**

Special Invited Session of the Biostatistics Section

Séance spéciale de l'invité du groupe de biostatistique

**[MS-34]****Biostatistics as narrative****Narration sur la biostatistique**

Gerald VAN BELLE, *University of Washington*

I discuss the societal context biostatistics. My thesis is that biostatistics is exciting, relevant to daily experience, and essential for understanding ourselves and our society. The grand themes of biostatistics are the refinement of everyday experience.

If the above is true, why isn't that the perception of our fellow citizens? How can we make their experience the refinement of every day statistics?

Biostatisticians should be the sought-after experts for comment on topics such as treatment efficacy, policy, pollution, validation of cause and effect. This talk will show some ways in which this is done, and ways in which we can improve.

Je discute du contexte social des biostatistiques. Ma thèse porte sur le fait que la biostatistique est excitante, utile aux expériences journalières et essentielle pour nous comprendre nous même et notre société. Les grands thèmes de la biostatistique sont le raffinement des expériences de chaque jour.

Si ce qui est mentionné plus haut est vrai, pourquoi n'est-ce pas la perception des autres citoyens ? Comment pouvons nous rendre leur expérience le raffinement des statistiques.

Les biostatisticiens devraient être demandés par les experts pour commenter sur des sujets tels l'efficacité des traitements, les politiques, la pollution, la validation des causes et effets. Cet exposé va montrer comment ceci est fait et certaines manières pour l'améliorer.

**Session 03B****Lundi 13 juin • Monday, June 13, 13:30 – 15:00****214**

Quality improvement

Amélioration de la qualité

**[MS-35]****The power of a baseline investigation I****La puissance d'une étude de base I**Jock MacKAY & Stefan STEINER, *University of Waterloo*

To reduce variation in a medium to high volume manufacturing process, we recommend a sequence of investigations, guided by an algorithm we call Statistical Engineering. The purpose of the initial investigation is to establish the problem baseline. We propose a plan using a systematic sampling scheme for the baseline investigation to quantify the full extent of output variation and to determine how the output changes over time. We demonstrate the critical role of this baseline information in the planning and analysis of subsequent observational investigations to isolate a dominant cause of the variation.

Pour réduire la variation dans un processus de fabrication à moyen ou haut volume, nous recommandons une séquence d'investigations, guidée par un algorithme que nous appelons ingénierie statistique. Le but de l'investigation initiale est d'établir la base du problème. Nous proposons un plan qui utilise un plan d'échantillonnage systématique pour l'investigation de base, pour quantifier la variation total du produit et pour déterminer comment le produit change dans le temps. Nous démontrons le rôle critique que joue cette information de base dans la planification et l'analyse d'investigations subséquentes pour isoler une cause dominante de variation.

**[MS-36]****The power of a baseline investigation II****La puissance d'une étude de base II**Stefan STEINER & Jock MacKAY, *University of Waterloo*

In part II, we examine the role of the baseline information in the planning, conduct and analysis of experiments used in the later stages of the Statistical Engineering algorithm. In particular, we examine how knowledge of the full extent of variation and the time-dependent nature of the process is useful in defining experimental runs, blocks, and replicates. We also look at the role of randomization in light of the baseline knowledge. We discuss experiments for component swapping, verification of a suspect dominant cause and process robustness, commonly used as part of Statistical Engineering to reduce variation.

Dans la partie 2, nous examinons le rôle de l'information de base dans la planification, la mise en application et l'analyse d'expériences utilisées dans les dernières étapes de l'algorithme d'ingénierie statistique. En particulier, nous examinons comment la connaissance de la variation complète et de la nature dépendante du temps du processus est utile pour définir les lots, les blocs et les répliques expérimentales. Nous examinons également le rôle de la randomisation à la lumière de la connaissance de l'information de base. Nous discutons des expériences pour des échanges de composants, pour la vérification de la cause dominante pour un suspect et pour la robustesse d'un processus, qui sont habituellement utilisés en ingénierie statistique pour réduire la variance.

**[MS-37]****Control charts for monitoring process mean and variability simultaneously****Diagrammes de contrôle pour suivre simultanément la moyenne et la variabilité d'un processus**Gemaï CHEN, *The University of Calgary*, Lingyun ZHANG, *Massey University, New Zealand*

Process mean and variability are usually monitored by two different control charts. Some efforts have been made to use one chart by combining two existing charts, one for mean and one for variability. In this talk, we explore the possibility of designing charts which by nature are meant to monitor both mean and variability simultaneously.

La moyenne et la variabilité d'un processus sont habituellement suivies par deux diagrammes de contrôle différents. Certains efforts ont été effectués pour utiliser un seul diagramme en combinant deux diagrammes existants, un pour la moyenne et un autre pour la variabilité. Dans cette présentation, nous explorons la possibilité de construire des diagrammes qui sont naturellement conçus pour suivre simultanément la moyenne et la variabilité.

**Session 03C****Lundi 13 juin • Monday, June 13, 13:30 – 15:00****133**

## Measurement error models

## Modèles d'erreurs de mesure

**[MS-38]****Bayesian modelling for measurement error problems****Modélisation bayésienne pour des problèmes d'erreurs de mesures**Paul GUSTAFSON, *University of British Columbia*

There is considerable literature on how to adjust inferences from regression models because one or more of the predictor variables are measured imprecisely. We discuss Bayesian approaches to such adjustments, with particular reference to two issues that arise frequently in the literature. First, we comment on the issue of model identifiability. Often there is a gap between what might reasonably be assumed about the measurement error mechanism and what must be assumed to gain a formally identified model. Second, we consider the debate about the desirability of likelihood-based methods for such problems, given that the likelihood requires modelling the distribution of the true but unobserved predictor in the study population.

Il y a beaucoup de littérature sur comment ajuster l'inférence des modèles de régression lorsqu'une ou plusieurs variables prédictives sont mesurées de façon imprécise. Nous discutons d'approches bayésiennes pour de tels ajustements avec une référence particulière à deux problèmes qui apparaissent fréquemment dans la littérature. Premièrement, nous commentons le problème " d'identifiabilité " du modèle. Il y a souvent un écart entre ce qui peut être raisonnablement supposé sur le mécanisme d'erreur et ce qui doit être supposé pour obtenir un modèle formellement identifiable. Deuxièmement, nous considérons le débat sur le désir d'avoir des méthodes basées sur la vraisemblance pour de tels problèmes, sachant que la vraisemblance requiert la modélisation de la distribution du vrai prédicteur non observé dans la population étudiée.

**[MS-39]****Instrumental variable estimation of nonlinear errors-in-variables models****Estimation instrumentale de variable pour des modèles non linéaires d'erreurs dans les variables**Susanne SCHENNACH, *University of Chicago*

This paper establishes that instruments enable the identification of nonparametric regression models in the presence of measurement error and provides a root  $n$  consistent and asymptotically normal es-

Cet article établit que des instruments rendent possible l'identification de modèles de régression non paramétriques lorsqu'il y a des erreurs de mesures et présente un estimateur de racine d'ordre  $n$  consistant et asymptotiquement normal

imator when the regression function is parametrically specified. Both the identification and the estimation methodologies rely on Fourier analysis and on the theory of generalized functions. The proposed estimator takes the well-known form of a generalized method of moment estimator with a plugged-in nonparametric kernel density estimate, thus facilitating the analysis of its asymptotic properties. Monte Carlo simulations and an application to the estimation of the black-white male wage gap are presented.

lorsque la fonction de régression est spécifiée de façon paramétrique. Les méthodes d'identification et d'estimation sont basées sur l'analyse de Fourier et sur la théorie des fonctions généralisées. L'estimateur proposé prend la forme bien connue d'un estimateur par la méthode des moments généralisée avec une estimation non paramétrique par le noyau d'une densité. Cette dernière caractéristique facilite l'analyse des propriétés asymptotiques de l'estimateur. Nous présentons des résultats de simulations de Monte-Carlo et une application de la méthode pour l'estimation de l'écart entre les salaires des hommes de race noire versus les hommes de race blanche.

**[MS-40]**

**Identifiability and estimation of semiparametric nonlinear errors-in-variables models**

**Identifiabilité et estimation de modèles semi paramétriques non linéaires d'erreurs dans les variables**

Liqun WANG, *University of Manitoba*, Cheng HSIAO, *University of Southern California*

We deal with a nonlinear errors-in-variables model where distributions of the unobserved predictor variable and random errors are nonparametric. Using instrumental variable approach, we construct minimum distance estimators for model parameters, which are shown to be consistent and asymptotically normally distributed under fairly general conditions. We also construct simulation-based estimators to overcome the possible computational difficulty of minimizing an objective function which involves multiple integrals. Finally, using the method of moments and Fourier deconvolution we derive a rank condition for model identifiability, which is easy to check in practice.

Nous traitons un modèle non linéaire d'erreurs dans les variables où les distributions de la variable prédictive non observée et des erreurs aléatoires sont non paramétriques. En utilisant une approche instrumentale variable, nous construisons des estimateurs de distance minimale pour les paramètres du modèle. Sous certaines conditions générales, ces estimateurs sont consistents et asymptotiquement normaux. Nous construisons également des estimateurs basés sur des simulations pour éviter certains problèmes informatiques de minimisation d'une fonction objective impliquant des intégrales multiples. Finalement, en utilisant la méthode des moments et la déconvolution de Fourier, nous dérivons une condition d'ordre pour l'identifiabilité du modèle facilement vérifiable en pratique.

**Session 03D**

**Lundi 13 juin • Monday, June 13, 13:30 – 15:00**

**211**

**Recent development of finite mixture models**

**Développements récents dans les mélanges de modèles finis**

**[MS-41]**

**Adaptive estimation for mixture distributions**

**Estimation adaptative de mélanges de distributions**

Jiayang SUN, *Case Western Reserve University*, Jiahua CHEN, *University of Waterloo*

New adaptive estimators for parameters in mixture distributions are proposed. They are especially useful when data come in large quantities (so that they are too large to be inputted into a standard algorithm for (efficient) computation), or sequentially over-time (ie. they are on-line data). The performance

Nous proposons de nouveaux estimateurs adaptatifs pour les paramètres d'un mélange de distributions. Ils sont particulièrement utiles lorsqu'il y a beaucoup de données (lorsque le nombre est trop grand pour que les données soient utilisées directement dans un algorithme standard et pouvoir obtenir un calcul efficace), ou bien lorsque les données sont

of these estimators are compared with the standard EM algorithm in the standard situation when the EM algorithm can be used. Coupling the new estimators with the recent D-test we develop a data mining procedure for intrusion detection.

séquentielles dans le temps (i.e. des données en ligne). La performance de ces estimateurs est comparée avec celle de l'algorithme EM dans la situation habituelle où l'algorithme peut être utilisé. En jumelant les nouveaux estimateurs avec le test-D développé récemment, nous introduisons une procédure de forage de données pour la détection d'intrusions.

[MS-42]

**Hypothesis testing in mixture regression models**  
**Tests d'hypothèses pour des mélanges de modèles de régressions**

Hongtu ZHU, *Columbia University and New York Psychiatric Institute*, Heping Zhang ZHANG, *Yale University*

In this paper, we establish asymptotic theory for both the maximum likelihood and maximum modified likelihood estimators in mixture regression models. Moreover, under specific and reasonable conditions, we show that the optimal convergence rate of  $n^{-\frac{1}{4}}$  for estimating the mixing distribution is achievable for both the maximum likelihood and maximum modified likelihood estimators. We also derive the asymptotic distributions of two log-likelihood ratio test statistics for testing homogeneity and propose a resampling procedure for approximating the  $p$ -value. Simulation studies are conducted to investigate the empirical performance of the two test statistics. Finally, two real data sets are analyzed to illustrate the application of our theoretical results.

Dans cet article, nous établissons une théorie asymptotique pour les estimateurs du maximum de vraisemblance et du maximum de vraisemblance modifié pour des mélanges de modèles de régression. De plus, sous des conditions spécifiques et raisonnables, nous montrons que le taux de convergence optimal de  $n^{-\frac{1}{4}}$  pour estimer la distribution de mélange est atteignable pour les deux estimateurs mentionnés plus haut. Nous obtenons également les distributions asymptotiques de deux statistiques de tests du rapport de vraisemblance pour tester l'homogénéité et nous proposons une procédure de ré-échantillonnage pour approximer la valeur- $p$ . Nous effectuons des études de simulations pour analyser la performance empirique des deux statistiques de tests. Finalement, nous analysons deux jeux de données réelles afin d'illustrer l'application de nos résultats théoriques.

[MS-43]

**Variable selection in finite mixture of regression models**  
**Sélection de variables pour un mélange fini de modèles de régression**

Abbas KHALILI & Jiahua CHEN, *University of Waterloo*

The problem of variable selection is an important model selection problem in statistical applications. We consider the problem of variable selection in Finite Mixture of Regression (FMR) models. Often, at the beginning of a study a large number of potential covariates are of interest. This produces a large and complex FMR model which is not desirable in statistical modeling. We extend the idea of LASSO and SCAD-type estimators proposed by Tibshirani (1996), Fan and Li (2001), to perform variable selection in FMR models. Performance of the method is studied theoretically and via simulations as well.

Le problème de sélection de variables est un problème important de sélection de modèles en statistiques. Nous considérons le problème de la sélection de variables pour un mélange fini de modèles de régression (MFMR). Au début d'une étude, il y a souvent un grand nombre de covariables d'intérêts possibles. Ceci produit un MFMR large et complexe, qui n'est pas souhaitable en modélisation statistique. Nous étendons l'idée des estimateurs de types LASSO ou SCAD proposés par Tibshirani (1996) et Fan et Li (2001) pour effectuer la sélection de variables pour un mélange fini de modèles de régression. La performance de la méthode est étudiée théoriquement et également par des simulations.

**Session 03E****Lundi 13 juin • Monday, June 13, 13:30 – 14:30****217**

Elaboration of a calibration strategy: a case study

Élaboration d'une stratégie de calage : une étude de cas

**[MS-44]****Overview of the harmonised calibration strategy for income statistics****Aperçu de la stratégie de calage harmonisée pour les statistiques du revenu**Johanne TREMBLAY & Michel LATOUCHE, *Statistique Canada*

In order to improve the comparability of income estimates and distributions across Statistics Canada surveys as well as with tax data and National Accounts, the calibration strategies of surveys on income, expenditures and financial security have been evaluated and harmonised. Income controls from administrative sources and new types of demographic controls were then added to the calibration step of these surveys.

An overview of the new strategy will be presented. This presentation will focus on the problems generated by the use of administrative files and by the use of several types of demographic controls.

Dans le but d'améliorer la comparabilité des estimations et des distributions de revenu entre les enquêtes de Statistique Canada ainsi qu'avec les données fiscales et les Comptes Nationaux, les stratégies de calage des enquêtes sur le revenu, les dépenses et la sécurité financière ont été évaluées et harmonisées. Des contrôles de revenu provenant de sources administratives et de nouveaux types de contrôles démographiques ont ainsi été ajoutés à l'étape de calage de ces enquêtes.

Un aperçu de la nouvelle stratégie sera présenté. L'emphase sera mise sur les problèmes engendrés par l'utilisation des fichiers administratifs et par l'utilisation de plusieurs types de contrôles démographiques.

**[MS-45]****Evaluation of the quality of the demographic estimations and the tax data used in the calibration of some different surveys in Statistics Canada (STC)****Évaluation de la qualité des estimations démographiques et des données fiscales utilisées lors du calage de différentes enquêtes de Statistique Canada (STC)**Sylvie AUGER & Johanne TREMBLAY, *Statistique Canada*

When calibrating, different surveys at STC use demographic estimates derived mainly from the last census and some administrative data files. These estimates are revised after each census. As well, some surveys use totals based on revenue coming from tax data files. Within the framework of the development of a strategy for harmonised calibration of income statistics, we did an evaluation of the quality of these estimates and totals.

We will present the results of the evaluation of the demographic estimations based on the 1996 Census and the results of the evaluation of the different tax data files available for the totals based on revenue.

Différentes enquêtes de STC utilisent, pour le calage, des estimations démographiques dérivées principalement du dernier recensement et de fichiers administratifs. Ces estimations sont révisées après chaque recensement. De même, certaines enquêtes utilisent des totaux selon le revenu provenant de fichiers de données fiscales. Dans le cadre du développement de la stratégie de calage harmonisée pour les statistiques du revenu, on a fait une évaluation de la qualité de ces estimations et totaux.

On présentera les résultats de l'évaluation des estimations démographiques basées sur le Recensement 1996 et les résultats de l'évaluation des différents fichiers administratifs disponibles pour les totaux selon le revenu.



**[MS-46]****Calibration strategy of the Survey of Labour and Income Dynamics  
Stratégie de calage de l'Enquête sur la dynamique du travail et du revenu**Sylvie LAROCHE & Michel LATOUCHE, *Statistique Canada*

The Survey of Labour and Income Dynamics (SLID) is a longitudinal panel survey of individuals. The sample is composed of two overlapping panels. In the context of the harmonised calibration strategy for income statistics, SLID developed a regression estimator that uses a combination of demographic and fiscal control counts to adjust survey estimates to known population counts. The purpose of the presentation will be to present the estimator and the auxiliary variables used, along with the impact of the new calibration strategy on the survey estimates.

L'Enquête sur la Dynamique du travail et du revenu (EDTR) est une enquête longitudinale par panel menée auprès des individus. L'échantillon est composé de deux panels qui se chevauchent. Dans le cadre de la stratégie de calage harmonisée pour les statistiques du revenu, l'EDTR a mis sur point un estimateur par régression combinant des comptes démographiques et fiscaux pour ajuster certaines estimations d'enquête à des totaux connus de la population. La présentation portera sur l'estimateur et les variables auxiliaires utilisées ainsi que sur l'impact de la nouvelle stratégie sur les estimations tirées de l'enquête.

**[MS-47]****Lessons to be learned from the use of administrative data at the weighting stage  
Leçons à tirer de l'utilisation de données administratives au stade de la pondération**Michel LATOUCHE & Johanne TREMBLAY, *Statistique Canada*

The use of administrative data at the weighting stage often improves the quality of the estimates. In addition to variance or bias reduction, it assures consistency across surveys and other sources. One should, however, be aware that a calibration strategy may not address an even more fundamental problem, that of sample representativity.

We will show the positive impacts of the Harmonised Calibration of Income Statistics project, the future of the project and the one of the generalised regression estimator. We will also discuss the activities undertaken by Statistics Canada to tackle sample representativity problems at their origin.

L'utilisation de données administratives lors de la pondération permet souvent d'améliorer la qualité des estimations. En plus de réduire la variance ou le biais, elles assurent une meilleure cohérence entre les différentes sources d'estimation. Il faut cependant être prudent lorsqu'on élabore une stratégie de calage. Cette méthode peut ne représenter qu'une solution palliative à un problème plus fondamental : la représentativité de l'échantillon.

Nous présenterons les impacts positifs de la stratégie de calage harmonisée sur les statistiques du revenu, de l'avenir de la régression généralisée et des activités entreprises par Statistique Canada pour résoudre les problèmes de représentativité.

**Session 03F****Lundi 13 juin • Monday, June 13, 13:30 – 14:30****212**Bayesian methodology  
Méthodologie bayésienne**[MS-48]****Importance sampling with the generalized exponential power density  
Simulations Monte Carlo avec fonction d'importance à l'aide de la densité GEP**Alain DESGAGNÉ, *Université du Québec à Montréal & Drake University*, Jean-François ANGERS, *Université de Montréal*

In this paper, the generalized exponential power (GEP) density is proposed as an importance function in Monte Carlo simulations in the context of estimation of posterior moments of a location parameter. The choice of the GEP density as an importance function allows us to obtain reliable and effective results when p-credences of the prior and the likelihood are defined, even if there are conflicting sources of information. Characterization and ordering of the tails of both the GEP density and the posterior can be done using p-credence. Simulation of observations from the GEP density is also addressed.

Dans cet article, la densité GEP (puissance d'exponentielle généralisée) est proposée comme fonction d'importance dans les simulations Monte Carlo dans le contexte d'estimation des moments a posteriori d'un paramètre de position. Le choix de la densité GEP comme fonction d'importance permet d'obtenir des résultats fiables et efficaces lorsque la p-crédence des densités a priori et de la vraisemblance sont définies, même s'il y a des sources d'information conflictuelles. La caractérisation et l'ordination des ailes de la densité GEP et de la densité a posteriori est possible grâce à la p-crédence. La simulation d'observations provenant d'une densité GEP est aussi discutée.

**[MS-49]****On the behaviour of Bayesian credible intervals for some restricted parameter space problems  
Sur le comportement de divers estimateurs par intervalles pour des espaces paramétriques contraints**Éric MARCHAND, *Université de Sherbrooke et University of New Brunswick*, William E. STRAWDERMAN, *Rutgers University*

For estimating a positive normal mean, Zhang and Woodroffe (2003) as well as Roe and Woodroffe (2000) investigate HPD credible sets associated with priors obtained as the truncations of non-informative priors onto the restricted parameter space. Namely, they demonstrate attractive frequentist properties of these procedures. We present and discuss their results before giving various extensions to other models, and to other situations.

Pour le problème de l'estimation du paramètre de moyenne d'une loi normale, lorsque ce paramètre est borné inférieurement, Zhang and Woodroffe (2003), ainsi que Roe and Woodroffe (2000), considèrent des estimateurs par intervalles associés à des lois a priori obtenus comme projections sur l'espace paramétrique de lois a priori non informatives. Notamment, ils en dégagent de belles et intéressantes propriétés fréquentistes. Nous présentons ces résultats et donnons des généralisations à d'autres modèles et situations.

**[MS-50]****Checking for prior-data conflict  
Vérification de l'existence d'un conflit a priori-données**Michael EVANS & Hadas MOSHONOV, *University of Toronto*

Inference proceeds from ingredients chosen by the analyst and data. To validate any inferences drawn, it is essential that the inputs chosen be deemed ap-

Toute inférence est faite à partir de données et d'une sélection d'ingrédients choisis par l'analyste. Pour valider tout résultat d'inférence, il est essentiel que les intrants choisis

appropriate for the data. In the Bayesian context these inputs consist of both the sampling model and the prior. There are thus two possibilities for failure: the data may not have arisen from the sampling model, or the prior may place most of its mass on parameter values that are not feasible in light of the data, referred to here as prior-data conflict. Failure of the sampling model can only be fixed by modifying the model while prior-data conflict can be overcome if sufficient data is available. We examine how to assess whether a prior-data conflict exists and how to assess when its effects can be ignored for inferences. The concept of prior-data conflict is seen to lead to a partial characterization of what is meant by a noninformative prior or a noninformative sequence of priors.

soient jugés appropriés pour les données. Dans le contexte bayésien, ces intrants sont le modèle de vraisemblance et la densité a priori. Il y a donc deux possibilités d'échec : les données peuvent ne pas provenir du modèle de vraisemblance choisi ou la densité a priori peut accorder un poids trop important à des valeurs de paramètres qui semblent peu probables (voire impossibles) en examinant les données, menant ainsi à ce que nous appelons un conflit a priori-données. Un échec du modèle de vraisemblance peut être résolu uniquement en modifiant ce modèle tandis qu'un conflit a priori-données peut être surmonté s'il y a suffisamment de données. Nous examinons comment il est possible d'établir l'existence d'un conflit a priori-données et d'évaluer lorsque l'effet d'un tel conflit peut être négligeable en termes d'inférence. Le concept de conflit a priori-données semble pointer vers une caractérisation partielle de ce que doivent être une densité a priori non-informative et une suite de densités a priori non-informatives.

#### [MS-51]

#### **Model selection by cross-validation in linear model**

#### **Sélection de modèle par validation croisée dans des modèles linéaires**

Hui SHEN & Will WELCH, *University of British Columbia*

Cross-validation (CV) is widely used in model selection. Although the idea and the implementation are simple, how to choose an appropriate cross-validation strategy usually does not have an easy answer. In this talk, we will compare model selection performance for  $v$ -fold CV and repeated  $v$ -fold CV based on linear model. We argue that repeating helps to improve the model selection performance. The criterion we use to select the best model is minimum CVMSE (cross-validated mean squared prediction error).

La validation croisée (VC) est largement utilisée pour faire de la sélection de modèle. Même si l'idée derrière la méthode et l'implémentation sont simples, le choix d'une stratégie de validation croisée n'a habituellement pas de réponse simple. Dans cette présentation, nous comparons la performance de sélection de modèle pour la VC  $v$ -fold et  $v$ -fold avec répétition basé sur des modèles linéaires. Nous argumentons que la répétition améliore la performance de sélection de modèle. Le critère que nous utilisons pour sélectionner le meilleur modèle est l'erreur quadratique moyen de prévision de validation croisée (EQMPVC) minimum.

### **Session 03G**

**Lundi 13 juin • Monday, June 13, 13:30 – 15:00**

**200**

General Meeting of SSC Accredited members (open)

Assemblée générale des membres accrédités par la SSC

**Session 04A****Lundi 13 juin • Monday, June 13, 15:30 – 17:00****241**Statistical climatology  
Climatologie statistique**[MS-52]****Use of climate data in a drought reporting system for agriculture****Utilisation de données climatiques dans un système de suivi des sécheresses pour l'agriculture**Allan HOWARD & Allan HOWARD, *Agriculture and Agri-Food Canada*, Samuel SHEN, *University of Alberta*, Ralph WRIGHT, *Alberta Agriculture Food and Rural Development*

Climate data has been collected in Canada for over 100 years, however few stations have a consistent record over a long time period. Station locations are sparsely distributed and the operational life of the stations is inconsistent, resulting in considerable temporal and spatial variation in the climate record. To enhance the interpretation of the drought indices, historic data for Alberta and the surrounding area was interpolated to develop datasets at 3 scales with daily weather values from 1901-2002. The data was used to model long term soil moisture conditions and precipitation patterns that form the basis for several drought indices used in the Alberta drought monitoring system. Application of this method to the rest of Canada is under review.

Des données sur le climat ont été ramassées depuis plus de 100 ans au Canada, cependant peu de stations ont des enregistrements constants sur une longue période de temps. Les stations sont localisées inégalement et le temps de vie opérationnel de celles-ci est inconsistant, ce qui se traduit par une grande variance autant spatiale que temporelle des enregistrements climatiques. Pour améliorer l'interprétation des indices de sécheresse, des données historiques pour l'Alberta et les endroits environnants ont été interpolées pour développer une base de données à trois échelles avec des données climatiques journalières de 1901 à 2002. Les données ont été utilisées pour modéliser les conditions d'humidité du sol à long terme et les cycles de précipitations, qui forment la base de plusieurs indices de sécheresse utilisés dans le système de suivi des sécheresses de l'Alberta. Nous analysons présentement la possibilité d'appliquer cette méthode au reste du Canada.

**[MS-53]****Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptive method****Estimation de champs de température globaux à partir d'observations parsemées avec une méthode spatialement adaptative basée sur les ondelettes sphériques**Hee-Seok OH, *University of Alberta*

This talk considers the problem of estimating the entire temperature field for every location on the globe from scattered surface air temperatures observed by a network of weather stations. Classical methods such as spherical harmonics and spherical smoothing splines are not efficient in representing the data that have inherent multiscale structures. This talk presents an estimation method that has the capability of adapting to the multiscale characteristics of the data. The method is based on a spherical wavelet approach recently developed for multiscale representation and analysis of scattered data. Spatially adaptive estimators are obtained by coupling

Dans cette présentation, nous considérons le problème de l'estimation du champ de température complet pour chaque point sur le globe à partir d'observations parsemées de la température de l'air observée dans un réseau de stations météorologiques. Les méthodes classiques telles que les harmoniques sphériques et les splines de lissage sphériques ne sont pas efficaces pour représenter les données qui ont des structures multi-échelles inhérentes. Nous présentons une méthode d'estimation qui possède la capacité de s'adapter à la structure multi-échelle des données. La méthode est basée sur une approche par les ondelettes sphériques récemment développée pour la représentation et l'analyse multi-échelle de données parsemées. Les estimateurs spatialement adap-

the spherical wavelets with different thresholding (selective reconstruction) techniques. These estimators are compared for their spatial adaptability and extrapolation performance using the surface air temperature data.

atifs sont obtenus en combinant les ondelettes sphériques avec différentes techniques de seuils (reconstruction sélective). Ces estimateurs sont comparés entre eux selon leur adaptabilité spatiale et leur performance d'extrapolation en utilisant les données de la température de l'air à la surface.

#### [MS-54]

#### **Interpolating daily precipitation data and regression variance correction**

#### **Interpolation de données sur les précipitations journalière et correction par régression de la variance**

Samuel SHEN, *University of Alberta*

A hybrid interpolation method, called RVC (regression variance correction), of retaining variances and using all the station data was developed by S.S.P. Shen, P. Dzikowski, G. Li and D. Griffith [Interpolation of 1961-1997 daily climate data onto Alberta polygons of ecodistrict and soil landscape of Canada, *J. Applied Meteor.* 40, 2162-2177 (2001)] based upon simple mass conservation. The method interpolates the station data onto a dense network of grid points, then averages the grid-point values inside polygons, and finally corrects the variance of the polygon data. The motivation to develop this method was to generate a set of daily precipitation data on 149 Ecodistrict Polygons in Alberta Province, Canada, from the precipitation data observed at 927 stations unevenly distributed within the province. The interpolated data are used as an input to various soil quality models, such as EPIC (Erosion/Productivity Impact Calculator). A key criterion for the precipitation input is the precipitation frequency, i.e., the temporal distribution of precipitation. However, because most of days in Alberta are sunny and have zero precipitation, the conventionally interpolated data often have too low variance in both space and time. Before RVC was developed, a stochastic model, such as Cligen, was used to generate daily precipitation data that have the same mean and variance as the real field. However, the randomly generated precipitation field does not precipitate on the same days as the real field and also cannot generate outliers, i.e., extreme storms or extremely long period of drought. This talk will explain the RVC method from the perspective of statistical regression and demonstrate why the interpolated data can fit both climate mean and climate variance. The talk will also review several commonly used interpolation methods and their applicability.

Une méthode d'interpolation hybride appelé CRV (correction par régression de la variance) a été développé par S.S.P. Shen, P. Dzikowski, G. Li et D. Griffith [Interpolation of 1961-1997 daily climate data onto Alberta polygons of ecodistrict and soil landscape of Canada, *J. Applied Meteor.* 40, 2162-2177 (2001)]. Cette méthode qui préserve la variance et qui utilise toutes les données des stations est basée simplement sur la conservation de la masse. La méthode effectue premièrement une interpolation des données provenant des stations sur un réseau dense de points sur une grille, ensuite moyenne sur les valeurs des points à l'intérieur de polygones et finalement effectue une correction de la variance des données sur ces polygones. La motivation derrière cette méthode était de générer un ensemble de données sur les précipitations journalières pour 149 polygones écodistrictes dans la province d'Alberta au Canada à partir de données observées sur les précipitations dans 927 stations d'observations localisées non uniformément à travers la province. Les données interpolées sont utilisées comme intrants dans plusieurs modèles sur la qualité du sol, tels que EPIC (Erosion/Productivity Impact Calculator). Un critère clé pour les intrants sur les précipitations est la fréquence de précipitation, c'est à dire la distribution temporelle des précipitations. Cependant, puisque la majorité des journées en Alberta sont ensoleillées et n'enregistrent aucune précipitation, les données interpolées de manière conventionnelle ont souvent une trop faible variance tant au niveau spatial que temporel. Avant que le CRV soit développé, un modèle stochastique tel le Cligen était utilisé pour générer des données sur les précipitations journalières qui ont la même moyenne et la même variance que les données réelles. Cependant, le champ de précipitation généré aléatoirement ne prédit pas de précipitations aux mêmes jours que le champ réel et ne peut également pas générer des valeurs à l'écart, c'est à dire des tempêtes extrêmes ou des longues périodes de sécheresse. Cette présentation explique la méthode de CRV d'une perspective de régression statistique et démontre pourquoi les données interpolées peuvent ajuster la moyenne et la variance clima-

tique. Dans l'exposé nous allons également présenter une revue des plusieurs méthodes d'interpolation utilisées couramment ainsi que leurs applications.

## **Session 04B**                      **Lundi 13 juin • Monday, June 13, 15:30 – 17:00**                      **214**

### **Inference from multiple surveys and multiple frame surveys** **Inférence à partir d'enquêtes multiples et d'enquêtes à cadres multiples**

[MS-55]

#### **Bootstrap variance estimation for multiple frame surveys** **Estimation bootstrap de la variance pour des sondages à plusieurs bases**

Sharon LOHR, *Arizona State University*, J.N.K. RAO, *Carleton University*

Multiple frame surveys are increasingly used in situations where several sampling frames may provide better coverage or cost-efficiency for estimating population quantities of interest. The probability sampling designs used to sample from the individual frames must be employed to obtain variance estimates and confidence intervals for these quantities. We propose two bootstrap methods for estimating variances in multiple frame surveys, and derive their properties. A simulation study demonstrates small-sample properties of the methods.

Les sondages à bases multiples sont de plus en plus utilisés dans des situations où plusieurs bases de sondage peuvent fournir une meilleure couverture ou bien réduire les coûts pour estimer des quantités d'intérêts dans une population. Le plan d'échantillonnage utilisé pour échantillonner les bases individuelles doit être utilisé pour obtenir des estimateurs de la variance et des intervalles de confiance pour ces quantités. Nous proposons deux méthodes bootstrap pour estimer les variances dans des sondages à bases multiples et nous dérivons leurs propriétés. Une étude de simulation démontre les propriétés des méthodes pour de petits échantillons.

[MS-56]

#### **Dual Frame Calibration for Possibly Dependent Samples under Different sampling Designs** **Calibration pour deux bases de sondage avec des échantillons possiblement dépendants sous différents plans d'échantillonnage**

Avinash SINGH, *Statistics Canada*, Shiyong WU, *RTI International*

An interesting situation where dependency in samples could arise is in the context of combining information from the main survey with a follow-up of nonrespondents to the main survey as considered by Singh, Iannacchione, and Dever (2003). The usual problem of dual frame estimation deals, however, with independent samples with partial overlap, and is considered, among others, by Fuller and Burmeister (1972), Skinner and Rao (1996), and Singh and Wu (1996). A simpler problem is that of combining independent samples from a single frame or two frames with complete overlap. The Singh-Wu paper attempts to develop regression calibration estimators for dual frames. To account for different designs, they suggested use of inverse relative effective sample size (based on design effects) as scaling factors to differentially weight contribu-

Une situation intéressante où l'on peut observer de la dépendance dans des échantillons est lors de la combinaison d'information provenant d'un sondage principal avec celle d'un suivi des non répondants comme il a été considéré par Singh, Iannacchione et Dever (2003). Habituellement, les défis lors de l'utilisation de base multiples sont reliés à l'utilisation d'échantillons indépendants qui se chevauchent partiellement. Ils sont présentés, entre autre, par Fuller et Burmeister (1972), Skinner et Rao (1996) et Singh et Wu (1996). Un problème plus simple est la combinaison d'échantillons indépendants provenant d'une seule base ou bien de deux bases qui se chevauchent complètement. L'article de Singh et Wu tente de développer des estimateurs de calibration par la régression pour deux bases de sondage. Pour tenir compte des différents plans d'échantillonnage, ils suggèrent l'utilisation de l'inverse de la taille effective de l'échantillon (basé sur l'effet du plan) comme facteur d'échelle pour différen-

tions from the two samples in the regression coefficient. However, an alternative scaling factor as a measure of relative effective sample size is needed without relying on the design effect as the design effect is known to be unsuitable for the multivariate case involving variances and covariances. They also suggested an intuitive value of 0.5 as a combining factor for the two calibration estimators for the common domain. Here, an alternative but objective way to choose the combining factor is needed. The above two limitations were addressed recently by Singh and Wu (2003) by taking a new approach of Dual Frame Calibration which allows for dependent samples as well. Results of an extended simulation study comparing various methods are presented.

tier le poids des deux bases dans le coefficient de régression. Cependant, un autre facteur d'échelle comme mesure de la taille de l'échantillon effective relative qui ne dépend pas du plan est requis, puisqu'il est connu que l'effet du plan n'est pas adapté au cas multivarié impliquant des variances et des covariances. Ils suggèrent également une valeur intuitive de 0.5 comme facteur de combinaison pour les deux estimateurs de calibration pour le domaine habituel. Une méthode alternative et objective est requise pour choisir le facteur de combinaison. Les deux limitations mentionnées ci-dessus ont été étudiées récemment par Singh et Wu (2003) en prenant une nouvelle approche de calibration pour deux bases qui fonctionne également pour des échantillons dépendants. Des résultats d'une vaste étude de simulation comparant plusieurs méthodes seront présentés.

**[MS-57]**

**Empirical likelihood inference for multiple surveys and multiple frame surveys**

**Inférence par vraisemblance empirique pour des sondages multiples et des sondages à bases multiples**

Changbao WU, *University of Waterloo*, Jon N.K. RAO, *Carleton University*

We develop pseudo empirical likelihood (PEL) based estimation methods using data from multiple surveys and multiple frame surveys. For multiple surveys, optimal point estimators and PEL ratio confidence intervals are obtained by using suitably defined PEL function and the design effects for each of the involved surveys; for dual frame or multiple frame surveys, two different approaches, one based on the combined sample and the other based on the post-stratified sample, are explored. Comparisons to existing approaches are pursued, both theoretically and empirically.

Nous développons une méthode d'estimation basée sur la pseudo-vraisemblance empirique (PVE) en utilisant des données de sondages multiples et de sondage à bases multiples. Pour les sondages multiples, les intervalles de confiance pour les estimateurs ponctuels optimaux et le ratio de PVE sont obtenus en utilisant une fonction de PVE bien définie et les effets de plans d'échantillonnage pour chaque sondage impliqué. Pour les sondages à deux ou plusieurs bases, deux approches différentes sont explorées, la première basée sur les échantillons combinés et la deuxième basée sur l'échantillon post-stratifié. Nous effectuons des comparaisons théoriques et empiriques avec d'autres approches existantes.

**Session 04C**

**Lundi 13 juin • Monday, June 13, 15:30 – 17:00**

**200**

**Methodology on missing data in longitudinal studies**

**Méthodologie pour les données manquantes dans les études longitudinales**

**[MS-58]**

**Semiparametric estimation in general clustered/longitudinal data problems**

**Estimation semi paramétrique pour des problèmes généraux de données de type groupées longitudinales**

Xihong LIN, *University of Michigan*, Raymond CARROLL, *Texas A&M University*

We consider a wide class of semiparametric problems with repeated evaluations of a nonparametric function, where estimation can be based on a

Nous considérons une large classe de problèmes semi-paramétriques avec des évaluations répétées d'une fonction non paramétrique où l'estimation peut être basée sur un cri-

criterion function such as a likelihood. Special cases included in our approach include marginal models for longitudinal/clustered data, conditional logistic regression for matched case-control studies, multivariate measurement error models, generalized linear mixed models with a semiparametric component, and many others. We propose profile and backfitting kernel estimation methods for these problems, derive their asymptotic distributions, and show that in likelihood problems the methods are semiparametric efficient. While generally not true, with our methods profiling and backfitting are asymptotically equivalent. We also consider pseudolikelihood methods where some nuisance parameters are estimated from a different algorithm. The proposed methods are evaluated using simulation studies and applied to the Kenya hemoglobin data.

rière fonctionnel tel que la vraisemblance. Les cas spéciaux généralisés par notre approche incluent les modèles marginaux pour des données de type groupées longitudinales, la régression logistique conditionnelle pour des études de cas témoins appariés, les modèles d'erreurs de mesures multivariées, les modèles linéaires mixtes généralisés avec une composante semi-paramétrique et plusieurs autres. Nous proposons des modèles d'estimation par le noyau de profilage et de backfitting pour ces problèmes, nous dérivons leurs distributions asymptotiques et montrons que dans des problèmes de vraisemblance, les méthodes sont semi-paramétriquement efficaces. Même si ce n'est généralement pas le cas, dans notre méthode, le profilage et le backfitting sont asymptotiquement équivalents. Nous considérons également des méthodes de pseudo-vraisemblance où des paramètres de nuisances sont estimés à partir d'un algorithme différent. Les méthodes proposées sont évaluées avec des études de simulations et appliquées à des données sur l'hémoglobine au Kenya.

#### [MS-59]

#### **Semiparametric methods for inferring treatment effects of outcomes defined only if a post-randomization event occurs**

#### **Méthodes semi paramétriques pour inférer les effets de traitement sur les résultats qui ne sont pas définis à moins qu'un événement de post-aléation se réalise.**

Yannis JEMIAI, *Cytel Software Corporation*

We investigate drawing inference, from a randomized experiment, about the treatment effect on an outcome defined only if a post-randomization event  $E$  has occurred. For instance, in an HIV vaccine trial, viral load can only be measured among those subjects who become infected. In such settings, we argue that the only meaningful treatment effect is measured in the subpopulation of subjects who would experience the event  $E$  regardless of treatment assignment. We show this quantity is non-identifiable from the observed data. As a consequence, we adopt two inferential strategies: 1) we place sharp bounds on the relevant quantity that contain a range of plausible values for it, and 2) we describe a set of identifying assumptions extending work by Gilbert et al.(2003), which allow us to estimate under a semiparametric model the treatment effect in the relevant subpopulation conditional on baseline covariates. We develop a class of estimating equations whose solutions comprise, up to asymptotic equivalence, all consistent and asymptotically normal estimators in our model.



**[MS-60]****A missing data problem arising in orthopedic trials of blood thinners****Un problème de données manquantes dans des essais orthopédiques sur les éclaircisseurs sanguins**Richard Cook & Grace Yi, *University of Waterloo*

In most orthopedic surgeries, patients receive some form of blood thinner to reduce the risk of clots during and after surgery. In a small fraction of cases patients develop serological responses following exposure to these blood thinners. Interest lies in i) estimating the proportion of patients developing this serological response, and ii) examining the effect of seroconversion on longer term risk of complications. The challenge in addressing these questions is that patients' seroconversion times are only available through a current status observation scheme since blood samples are taken only upon discharge following surgery. In this talk we will examine the information available to estimate the fraction of patients with this latent binary trait (seroconversion status) and discuss issues surrounding the use of this latent trait as a predictor for long term outcomes.

Dans la majorité des chirurgies en orthopédie. Les patients reçoivent une sorte d'éclaircisseur sanguin pour réduire le risque de caillots après une chirurgie. Dans une petite portion des cas, les patients développent une réponse sérologique suite à l'exposition à ces éclaircisseurs. Il y a un intérêt 1) d'estimer la proportion des patients qui développent une telle réponse sérologique et 2) d'examiner l'effet d'un séroconversion sur le risque de complication à long terme. Le défi relié à ces questions est que les temps de séroconversion du patient sont disponibles à travers un plan d'observation du statut courant puisque les échantillons de sang sont prélevés seulement après la chirurgie. Dans cet exposé, nous examinons l'information disponible pour estimer la proportion de patients avec un trait binaire latent (statut de séroconversion) et discutons des problèmes reliés à l'utilisation de ce trait latent comme prédicteur des résultats à long terme.

**Session 04D****Lundi 13 juin • Monday, June 13, 15:30 – 17:00****133****Spatial analysis of survival data****Analyse spatiale de données de survie****[MS-61]****Spatial cluster detection for censored data****Détection de clusters spatiaux pour des données censurées**Yi LI, *Harvard University*, Andrea Cook,

Numerous methods that have been proposed to test for spatial clustering. But few have focused on censored data. This paper provides an extension of the spatial scan statistic for data with failure time outcomes using both unstratified and stratified semi-parametric methods. It further proposes a robust clustering detection method that utilizes the principle of cumulative martingale residuals for censored outcomes. Application of these methods will be illustrated by the Home Allergens and Asthma prospective cohort study looking at the relationship between environmental exposures and allergic reactions such as asthma and allergic rhinitis/hayfever.

Plusieurs méthodes ont été proposées pour tester la classification spatiale, mais peu d'entre elles se sont arrêtées aux données censurées. Cette présentation fournit une extension à la statistique de balayage spatiale pour des données de temps de vie en utilisant des méthodes semi-paramétriques stratifiées et non stratifiées. Nous proposons ensuite une méthode robuste de détection de clusters qui utilise le principe de résidus martingales cumulatifs pour des résultats censurés. L'application de ces méthodes sera illustrée par l'étude de cohorte prospective sur les allergies et l'asthme qui analyse la relation entre les expositions environnementales et les réactions allergiques telles que l'asthme et les allergies comme le rhume des foins.

**[MS-62]****Flexible modelling of spatially varying cure rates for survival data****Modélisation flexible pour des données de survie avec des taux de guérison qui varient spatialement**Sudipto BANERJEE, Freda COONER & Brad CARLIN, *University of Minnesota*

Survival models have been and continue to be extremely popular in analyzing time-to-event data. With rapid improvements in medical treatment and health-care, survival data reveal a substantial portion of patients being cured. Indeed a proper assessment of such data calls for statistical models that account for cure. Classes of survival models, called cure rate models, account for the probability of a subject getting cured. Popular cure models can be broadly classified into the classical mixture models of Berkson and Gage (1952) (BG type) or the hierarchical classes of Chen, Ibrahim and Sinha (1999) (CIS type). Our present work proposes a unifying class of cure models that facilitates flexible hierarchical model-building while incorporating possible spatial associations for geographically referenced survival data. Issues such as regressing on the cure fraction and spatial distributions are discussed. The models are illustrated with survival data sets from SEER, revealing different underlying mechanisms (including spatial) that lead to relapse and cure.

Les modèles de survie sont et continuent à être extrêmement populaires pour analyser des données de temps vie. Avec les améliorations rapides dans les soins de santé et les traitements médicaux auxquels nous assistons, les données de survie révèlent souvent une portion substantielle de patients qui guérissent. Ainsi, une bonne évaluation de telles données requiert des modèles statistiques qui tiennent compte de la guérison. Des classes de modèles de survie appelés les modèles de taux de guérison tiennent compte de la probabilité qu'un sujet guérisse. Les modèles de guérison populaires peuvent être grossièrement classifiés parmi les modèles de mélanges classiques de Berkson et Gage (1952) (type BG) ou parmi les classes hiérarchiques de Chen, Ibrahim et Sinha (1999) (type CIS). Nous proposons dans ce travail une classe unifiée de modèles de guérison qui facilite la construction de modèles hiérarchiques flexibles, tout en incorporant une corrélation spatiale potentielle pour des données de survie à référence géographique. Nous discutons de problèmes tels que la régression sur la fraction de guérison et les distributions spatiales. Les modèles sont illustrés avec des jeux de données de survie provenant du SEER qui révèlent des mécanismes sous jacents différents (incluant spatial) menant à la rechute ou la guérison.

**[MS-63]****Survival models for spatial data: a Poisson modelling approach****Modèles de survie pour des données spatiales : une approche par un modèle de Poisson**Renjun MA, *University of New Brunswick*

We propose a Poisson modelling approach to Survival models with spatially correlated random effects. An important feature of this approach is that the principal results depend only on the first and second moments of the unobserved random effects. The orthodox best linear unbiased predictor approach to random effects Poisson modelling techniques enables us to justify appropriate consistency and optimality. The explicit expressions for the random effects given by our approach facilitate incorporation of relatively a large number of random effects.

Nous proposons une approche par un modèle de Poisson pour des modèles de survie avec un effet aléatoire corrélé spatialement. Une caractéristique importante de cette approche est que les résultats principaux dépendent seulement des deux premiers moments de l'effet aléatoire non observé. L'approche orthodoxe du meilleur estimateur linéaire sans biais pour les techniques de modélisation de Poisson avec effet aléatoire nous permet de justifier une consistance et une optimalité appropriée. L'expression explicite de l'effet aléatoire donnée par notre approche facilite l'incorporation d'un nombre relativement grand d'effets aléatoires.

**Session 04E****Lundi 13 juin • Monday, June 13, 15:30 – 16:45****211**

Dependent data analysis

Analyses des données dépendantes

**[MS-64]****Another method for bootstrapping the arithmetic mean and the median when the observations are dependent****Une autre méthode pour faire du bootstrap sur la moyenne arithmétique et la médiane lorsque les observations sont dépendantes**Aberazzak MOUIHA & Nourddine RAIS, *Université de Fès, Maroc*

We propose a new bootstrap method based on approximating a stationary process by a  $p$ th order Markov chain. For a given data, we estimate the order  $p$  via Akaike criterion and we generate a bootstrap sample using the estimated kernel transition densities from the data. We show the consistency of this new method in the case of arithmetic mean. We give some simulations to compare its performance with the moving-block bootstrap MBB.

Nous proposons une nouvelle méthode de bootstrap basée sur l'approximation d'un processus stationnaire par une chaîne de Markov d'ordre  $p$ . Pour des données quelconques, nous estimons l'ordre  $p$  par le critère d'Akaike et nous générons un échantillon bootstrap en utilisant les densités de transition du noyau provenant des données. Nous montrons l'efficacité de cette nouvelle méthode pour le cas de la moyenne arithmétique. Nous donnons quelques résultats de simulations pour comparer la performance de cette dernière avec celle du bootstrap à blocs mobiles (BBM).

**[MS-65]****On approximating the distribution of a statistics for detecting serial correlation****Sur l'approximation de la distribution d'une statistique permettant la détection de corrélation sérielle**Deepak SANJEL, Serge PROVOST & Ian MACNEILL, *University of Western Ontario*

Several approximations to the distribution of a statistic which was proposed by De Gooijer and MacNeill (1999) to test for the presence of serial correlation at a given lag in regression residuals are discussed. It is shown that the moments of this statistic, once expressed as linear combinations of product moments of quadratic forms, can be evaluated by making use of a recursive formula or a symbolic computational methodology. Density approximants expressed in terms of beta density functions and Laguerre polynomials are then proposed. Finally, the results of a comparative study of the various approximations are presented.

Nous discutons de plusieurs approximations de la distribution d'une statistique qui a été proposée par De Gooijer et MacNeill (1999) pour tester la présence de corrélation sérielle, à un délai donné, dans les résidus de régression. Il est montré que les moments de cette statistique, une fois exprimés comme une combinaison linéaire de produits de moments de formes quadratiques, peuvent être évalués en utilisant une formule récursive ou une méthode computationnelle symbolique. Nous proposons ensuite des approximations de la densité de cette statistique, exprimées en termes de densités bêta et de polynômes de Laguerre. Finalement, les résultats d'une étude comparative des différentes approximations sont présentés.

**[MS-66]****Asymptotic properties of quasi-maximum likelihood estimator of ARMA(P,Q)-GARCH(p,q)  
Propriétés asymptotiques de l'estimateur du pseudo-maximum de vraisemblance d'un modèle  
ARMA(P,Q)-GARCH(p,q)**Pingguo LU & Hao YU, *University of Western Ontario*

(G)ARCH-type time series models have been extensively used in economics and finance since Engle (1982) and Bollerslev (1986). However, in practice, the constant conditional mean is often not valid. A realistic approach is to fit the data with an ARMA-GARCH model. In this talk, we show the asymptotic consistency and normality of the quasi maximum likelihood estimator (QMLE) for the parameters of an ARMA(P,Q)-GARCH(p,q) model with an arbitrary likelihood kernel. It appears that the moment requirements for ARMA errors and GARCH innovation depend on the choice of the likelihood kernel. By taking student t likelihood kernel, the asymptotic consistency holds with only small positive moments on ARMA errors and GARCH innovation. The asymptotic normality holds with finite variance on ARMA errors and any small positive moment on GARCH innovation. For pure GARCH model with nonzero constant mean, we show that the common practice of using sample mean to center financial data is workable.

Les modèles de séries chronologiques de type GARCH ont été intensivement utilisés en économie et en finance depuis Engle (1982) et Bollerslev (1986). Cependant, en pratique, l'hypothèse d'une moyenne conditionnelle constante n'est souvent pas vérifiée. Une approche réaliste est alors d'ajuster les données à un modèle ARMA-GARCH. Dans cette présentation, nous montrons la convergence et la normalité asymptotique de l'estimateur du pseudo-maximum de vraisemblance pour les paramètres d'un modèle ARMA(P,Q)-GARCH(p,q) avec un noyau de vraisemblance arbitraire. Il s'en dégage que les conditions sur les moments de l'erreur ARMA et de l'innovation GARCH dépendent du noyau de vraisemblance choisi. En considérant des noyaux de type student, la convergence de l'estimateur est obtenue avec l'existence de petits moments de l'erreur ARMA et de l'innovation GARCH. La normalité asymptotique est valide lorsque la variance de l'erreur ARMA est finie et l'innovation GARCH possède n'importe quel moment positif petit. Pour un modèle purement GARCH avec une moyenne constante non nulle, nous montrons que la pratique commune d'utiliser la moyenne échantillonnale pour centrer des données financières est exploitable.

**[MS-67]****A consistent test for independence between two infinite order cointegrated series  
Tests Convergent d'indépendance de deux séries cointégrées autorégressives d'ordre infini**Chafik BOUHADDIOU, *CIRANO*, Jean-Marie DUFOUR, *Université de Montréal*

Here, we propose a semiparametric approach for testing the independence between two infinite-order cointegrated series. We first filter each series by a finite autoregression. The test statistic is a standardized version of the sum of weighted squares of residual cross-correlation matrices. The weights depend on a kernel function and on a truncation parameter. The asymptotic distribution of the statistic test is derived under the null hypothesis and its consistency is also established for a fixed alternative of serial cross-correlation of unknown form. A simulation study is presented to investigate the finite sample properties of the proposed tests.

Dans cette présentation, nous proposons une approche semi-paramétrique pour tester l'indépendance de deux vecteurs cointégrés autorégressifs d'ordre infini. Nous approximons chaque processus par une autorégression d'ordre fini. La statistique de test est une version standardisée d'une somme quadratique pondérée de matrices de corrélations croisées résiduelles. Les poids dépendent d'une fonction de noyau et d'un paramètre de lissage. La distribution asymptotique de la statistique du test est établie sous l'hypothèse nulle. Nous montrons la convergence du test sous une classe générale d'alternatives. Les niveaux de ces tests et leurs puissances sont étudiées par simulation et une application numérique est présentée.

**Session 04F****Lundi 13 juin • Monday, June 13, 15:30 – 17:00****212**

Methods with applications

Méthodes et applications

**[MS-68]****Bayesian melding of ozone level measurements and deterministic model output****Combinaison bayésienne de mesures du niveau d'ozone et de résultats de modèles déterministes**Zhong LIU, James ZIDEK & Nhu LE, *University of British Columbia*

This paper addresses the problem of how to combine measurements from monitoring stations and output from deterministic models. We propose a Bayesian hierarchical model for that purpose. In the model, both the measurements and modeling output are assumed to be related to an unobserved underlying true process plus some bias and random error. The underlying true process and other parameters are estimated via Gibbs sampling and Metropolis-Hasting algorithm. The inference of the estimators will be carried out according to their posterior samples.

Cette présentation traite de la combinaison de mesures de stations d'observations et de résultats de modèles déterministes. Nous proposons un modèle hiérarchique bayésien pour ce problème. Dans ce modèle, nous supposons que les mesures et les résultats modélisés sont reliés à un processus réel sous jacent non observé plus un biais et une erreur aléatoire. Le processus réel sous jacent et les autres paramètres sont estimés par l'échantillonneur de Gibbs et l'algorithme de Metropolis-Hasting. L'inférence sur les estimateurs est faite par rapport aux échantillons a posteriori.

**[MS-69]****Spatial temporal analysis of waterborne disease risk as related to climate change****Analyse spatiale temporelle du risque de maladies d'origine hydrique en relation avec les changements climatiques**Daniel GILLIS & John HOLT, *University of Guelph*, Dominique CHARRON, *Public Health Agency of Canada*, David WALTNER-TOEWS, *University of Guelph*, Abdel MAAROUF, *Environment Canada*

Climate change is a growing facet of our lives. With proposed climate change models, Canada can expect significant changes to weather patterns and seasonal norms. These departures provide impetus to collect knowledge regarding the effects of climate change on the spatial-temporal distribution of waterborne illness, and to determine preventative and response protocols. This talk will discuss links between climate factors and the risk of disease as determined by a hospitalization Case-control study. Interpolation methods to obtain unobserved climate data will be discussed. Regions of significant risk increase as demonstrated via Downscaled climate simulations will be highlighted.

Les changements climatiques font de plus en plus partie intégrante de nos vies. Avec les modèles de changements climatiques proposés, le Canada peut s'attendre à des changements significatifs dans les cycles climatiques et les normes saisonnières. Ce point de départ nous force à accroître nos connaissances par rapport aux effets des changements climatiques sur la distribution spatiale et temporelle des maladies d'origine hydrique et ainsi déterminer des protocoles de prévention et d'actions. Cet exposé traitera des liens entre les facteurs climatiques et les risques de maladies déterminés par une étude cas-témoin d'hospitalisations. Les régions où l'augmentation du risque est significative selon des simulations climatiques micro-échelles seront identifiées.

**[MS-70]****Frailty modeling of Canadian mortality at advanced ages  
Modélisation de la mortalité canadienne aux grands âges**Kirill ANDREEV, *Queen's University*, Robert BOURBEAU, *University of Montreal*

Canadian Mortality Database provides a unique opportunity to study age-specific mortality dynamics in Canada on a cohort basis. Death rates have been computed for cohorts 1850–1888 from age 65 up to the highest age attended. Rate of mortality increase decelerates over age suggesting that frailty modeling might be appropriate. A four parameter model combining Gompertz baseline hazard and Hougaard-Aalen family of frailty distributions have been developed and fitted to the age-specific probabilities of dying. The results suggest that the model fits data better than traditional Gompertz model but its interpretation is sometimes difficult. Possible explanations are provided.

La base de données sur la longévité canadienne (BDLC) permet d'étudier l'évolution du profil de la mortalité canadienne par âge selon une approche longitudinale. Des taux de mortalité ont été calculés pour les générations 1850-1888, à partir de l'âge de 65 ans. On constate qu'aux grands âges, la croissance de la mortalité avec l'âge tend à se ralentir, ce qui rend pertinent l'utilisation d'un modèle de « fragilité ». Un modèle à quatre paramètres alliant la fonction de base de Gompertz et une famille de distributions de « fragilité » de Hougaard-Aalen a été développé et ajusté aux probabilités de décès selon l'âge. Les résultats suggèrent que ce modèle fournit un meilleur ajustement des données que le modèle traditionnel de Gompertz, mais l'interprétation des résultats est parfois difficile. Quelques explications possibles sont avancées.

**[MS-71]****Comparison of Bayesian and frequentist approaches in modelling risk of preterm birth near the Sydney Tar Ponds, Nova Scotia, Canada****Comparaisons des approches bayésiennes et fréquentistes pour modéliser le risque d'accouchement prématuré à proximité du Sydney Tar Pond, Nouvelle-Écosse, Canada**Afisi ISMAILA, Angelo CANTY & Lehana THABANE, *McMaster University*

This study compares a Bayesian and frequentist approaches in the modelling of the association between the risk of preterm birth and maternal proximity to hazardous waste and pollution from the Sydney Tar Pond site in Nova Scotia. In the frequentist approach, the Poisson model for aggregated data and weighted linear regression was used. The Poisson model was also fitted using Bayesian hierarchical modeling. The result of the Bayesian modelling and the frequentist approaches are comparable. The results show that there is no decrease in risk of preterm birth as we move from the Tar Pond site to other regions.

Cette étude compare les approches bayésiennes et fréquentistes pour modéliser l'association entre le risque d'accouchement prématuré et l'exposition à des déchets toxiques et la pollution du site de Sydney Tar Pond en Nouvelle-Écosse au moment de la grossesse. Dans l'approche fréquentiste, le modèle de Poisson pour des données agrégées et la régression linéaire pondérée ont été utilisés. Le modèle de Poisson a également été ajusté à l'aide d'une modélisation bayésienne hiérarchique. Les résultats provenant des deux approches sont comparables. Les résultats montrent que le risque d'accouchement prématuré ne baisse pas lorsqu'on passe du site de Sydney Tar Pond vers une autre région.

**[MS-72]****Analysis of priorities of patients living with diabetes  
Analyse des priorités des patients vivant avec le diabète**Hedy JIANG, Lehana THABANE, Anne HOLBROOK, Kalpana NAIR & Lisa DOLOVICH, *McMaster University*

Diabetes is a chronic disease that afflicts more than 2 million Canadians. Diabetes care is a challenging process that involves self-management activi-

Le diabète est une maladie chronique qui touche plus de deux millions de Canadiens. Le contrôle du diabète est un processus difficile qui demande des activités d'auto gestion

ties including regular monitoring of clinical disease indicators, diet control, regular exercise, and treatment with medication. The primary purpose of this study was to identify priorities for diabetes management from the patient's perspective and factors associated with prioritization. In this presentation, we report on the findings of a cross-sectional study where patients completed a questionnaire to gather their perceptions on the importance, ease and frequency of their preventative and treatment activities for management of diabetes.

incluant le suivi régulier d'indicateurs cliniques de la maladie, le contrôle de la diète, de l'exercice régulier et des traitements avec médication. Le but premier de cette étude est d'identifier les priorités des patients dans le contrôle du diabète et les facteurs associés à ces priorités. Dans cette présentation, nous rapportons les résultats d'une étude à plan croisé où les patients ont complété un questionnaire pour rassembler leurs perceptions sur l'importance, la facilité et la fréquence de leurs activités de prévention et de traitement pour contrôler leur diabète.

[MS-73]

**The challenge of teaching Biostatistics in a Faculty of Medicine**  
**Les défis d'enseigner la biostatistique dans une faculté de médecine**

Amparo CASANOVA-GURRERA, M. Begoña CAMPOS-BONILLA & Wilma PENZO-GIACCA, *Universitat de Barcelona*,  
 Wendy LOU, *University of Toronto*

Students in the Faculty of Medicine at Barcelona University obtain excellent marks in the old high school curriculum, but show poor knowledge of algebraic tools. Teaching Biostatistics requires demonstrating to them that statistical education will improve their medical competence. The main objective of this study is to identify factors that strongly impact grades in the Biostatistics exam. Gender, high school grades, the results of the University's entrance exam, mathematical skills, and the students' preferred learning styles are analyzed. The last one are evaluated via a Spanish modification of the Canfield Learning Styles Inventory (1977).

Les étudiants de la Faculté de médecine de Barcelona University présentent des résultats collégiaux excellents, mais montrent une faiblesse au niveau des outils algébriques. Enseigner la biostatistique exige de leur démontrer que des connaissances statistiques vont améliorer leurs compétences médicales. L'objectif premier de cette étude est d'identifier les facteurs qui influent fortement les résultats du test de biostatistique. Le sexe, les résultats collégiaux, les résultats au test de classement universitaire, le talent mathématique et le style d'apprentissage préféré de l'étudiant sont analysés. Ces derniers sont évalués à partir d'une modification espagnole du Canfield Learning Styles Inventory (1977).

**Session 04G**

**Lundi 13 juin • Monday, June 13, 15:30 – 16:15**

**217**

Statistical genetics and genomics  
 Statistique génétique et génomique

[MS-74]

**Selection of a cut-off for differentially expressed genes in the SAM method**  
**Sélection d'un seuil pour des gènes à différentes expressions dans la méthode ASM**

Angelo CANTY, *Toronto Hospital for Sick Children and McMaster University*

The Significance Analysis of Microarrays (SAM) method is often used to detect differentially expressed genes in a microarray experiment. Genes are declared to be differentially expressed if the test statistic is outside of an acceptance region. In most applications, this region is defined by the user to

La méthode d'analyse de signification de micro-réseaux (ASM) est souvent utilisée pour détecter des gènes exprimés différemment dans une analyse de micro-réseaux. On déclare qu'un gène est exprimé différemment si la statistique de test est à l'extérieur d'une région d'acceptation. Dans la plupart des applications cette région est définie par l'analyste de ma-

balance the number of genes declared differentially expressed and the estimated false positive rate. In this talk I will present a method of automatically choosing the cut-off points for this region by examining the curvature of the plot of observed test statistics against their expected values under a permutation distribution.

**[MS-75]**

**A mixture model approach to background correction for cDNA arrays**

**Approche par mélange de modèles pour la correction du fond de réseaux d'ADN complémentaires**

Mary LESPERANCE & Edward CHANG, *University of Victoria*, Karissa JOHNSTON, *University of British Columbia*, Carmen BAILEY & Caren HELBING, *University of Victoria*

One of the features of DNA arrays is that some background intensity will be detected around gene probe spots as a result of low-level, non-specific target DNA hybridization. Scanning software attempts to estimate this background intensity from the regions surrounding these spots. Traditionally, this estimated local-background intensity is subtracted from the intensity measurement (foreground intensity) of the spot and the resulting target intensity is used as an estimate of the amount of target DNA hybridized to the gene probe spot. One difficulty with this approach is that negative target intensities may result – in practice, the negative readings are often discarded from the dataset and further analysis. Yet, it is these seemingly weak expressors that may be of interest in a given study, especially in experiments where expression levels vary over time or condition. We propose a mixture model approach to background correction which embodies realistic assumptions.

nière à balancer le nombre de gènes qui sont déclarés comme étant exprimé différemment et le taux de faux positif estimé. Dans cet exposé, je vais présenter une méthode automatisée pour sélectionner le seuil pour cette région en examinant la courbure du graphique de la statistique de test en fonction de sa valeur espérée sous une distribution de permutation.

Une des caractéristiques des réseaux d'ADN est qu'une intensité de fond est détectée autour des points de sonde du gène en résultat de l'hybridation à bas niveau non spécifique de l'ADN ciblée. Les logiciels de balayage tentent d'estimer cette intensité de fond à partir des régions autour de ces points. Traditionnellement, cette estimation locale de l'intensité de fond est soustraite de l'intensité mesurée au point et l'intensité de la cible résultante est utilisée comme estimation de la quantité d'ADN cible hybridée au point de sonde du gène. Une des difficultés avec cette approche est qu'il est possible d'obtenir des intensités cibles négatives. En pratique, les lectures négatives sont souvent rejetées du jeu de données et des analyses. Cependant, il se peut que ce soit ces niveaux d'expressions qui soient intéressants dans une étude donnée, spécialement dans les expériences où le niveau d'expression varie dans le temps ou selon la condition. Nous proposons une approche par mélange de modèles pour la correction de fond qui stipule des hypothèses réalistes.

**[MS-76]**

**A survey of analyses of age-at-onset data in human genetics research**

**Sondage des méthodes d'analyses de l'âge au début de l'invalidité en recherche sur la génétique humaine**

Haydar SENGUL, *University of Saskatchewan*

Age at disease onset has turned out to be one of the most important factors in the study of some complex human diseases, which requires substantive attention. Even though the variability in the age at disease onset makes the study of complex human diseases complicated, incorporating age-at-onset data into the analysis of complex human diseases may result in a better model fit in determining genetic and/or environmental effect. Being corre-

L'âge du début de l'invalidité dû à une maladie s'est avéré être un des facteurs les plus importants dans l'étude de maladies humaines complexes, et ce dernier requiert une attention ardue. Même si la variabilité dans l'âge du début de l'invalidité rend l'étude de maladies humaines complexes compliquée, incorporer l'âge au début de l'invalidité dans l'analyse peut résulter en un meilleur ajustement du modèle pour déterminer les effets génétiques et/ou environnementaux. Étant corrélé à l'intérieur d'une famille et dû aux problèmes de



lated within families and due to the problem of censoring, it is very crucial to develop appropriate data analysis techniques taking into account the age-at-onset data that will make it possible to comprehend complex human diseases. In this study, we will survey all the methods, which take into account age-at-onset data, used in the context of genetic linkage and association analyses, and present our method that is based on a version of mixed-effects proportional hazards regression.

censure, il est crucial de développer des techniques d'analyses appropriées pour inclure l'âge au début de l'invalidité de manière à ce que la maladie humaine complexe soit saisie. Dans cette étude, nous allons regarder toutes les méthodes qui intègrent l'âge au début de l'invalidité dans le contexte des analyses de liaisons et d'associations génétiques. Nous allons également présenter notre méthode qui est basée sur une version de la régression à effet mixte de taux de panne proportionnel.

**Session 04H****Lundi 13 juin • Monday, June 13, 17:00 – 18:30****214**

NSERC Workshop  
Atelier du CRSNG

**Session 05****Mardi 14 juin • Tuesday, June 14, 09:00 – 10:30****241**

Gold Medal Address  
Allocution de la Médaille d'or

[MS-77]

**Correlation random fields, brain connectivity, and cosmology****Champs de corrélation aléatoires, connectivité de cerveau, et cosmologie**Keith WORSLEY, *McGill University*

The correlation random field is the 'image' of correlations at all possible pairs of points in two sets of images. We are interested in the topology of the (random) set of high correlations, more specifically, the Euler characteristic. We apply this to networks of fMRI images of brain activity, and to connections between MS lesions and cortical thickness. Similar methods are used in cosmology to look for large-scale structure in the universe, and anomalies in the cosmic microwave background. We apply this to the latest results from the Sloan Digital Sky Survey and the Wilkinson Microwave Anisotropy Probe.

Le champ aléatoire de corrélation est l'image des corrélations à toutes les paires possibles de points dans deux ensembles d'images. Nous sommes intéressés par la topologie de l'ensemble (aléatoire) des corrélations élevées, plus spécifiquement, la caractéristique d'Euler. Nous nous appliquons ceci aux réseaux des images de fMRI d'activité de cerveau, et aux raccordements entre les lésions de sclérose en plaques et l'épaisseur corticale. Des méthodes semblables sont employées dans la cosmologie pour rechercher la structure à grande échelle dans l'univers, et des anomalies dans le fond cosmique de micro-onde. Nous nous appliquons ceci aux derniers résultats de Sloan Digital Sky Survey et de Wilkinson Microwave Anisotropy Probe.

**Session 06A****Mardi 14 juin • Tuesday, June 14, 11:00 – 12:30****133**

Bayesian model selection  
Sélection de modèles bayésiens

**[MS-78]****Feature selection in model-based analysis of functional data****Sélection de caractéristiques dans l'analyse paramétrique de données fonctionnelles**Marina VANNUCCI, *Texas A&M University*

In this talk I will describe methodologies for Bayesian modelling of functional data that incorporate feature extraction. Practical applications will be classification problems that involve functional predictors. Wavelet methods will be used for dimension reduction. Probit models and Bayesian methods will allow the simultaneous classification of the samples as well as the selection of the discriminating features of the data. Applications will involve spectral data. In mass spectrometry, for example, the identification of peaks related to a specific outcome, i.e. peaks that discriminate samples or that predict a clinical response, is of interest. Other practical contexts will come from chemometrics studies that explore the possibility of using NIR spectra to classify samples. In all examples we will find that very small sets of features lead to good classification results.

Dans cette présentation je décrirai des méthodes de modélisation bayésienne de données fonctionnelles permettant l'extraction de caractéristiques. Ces méthodes seront appliquées à des problèmes de classification qui impliquent des prédicteurs fonctionnels.

Des méthodes d'ondelettes seront utilisées pour réduire la dimensionalité du problème. Des modèles probits et des méthodes bayésiennes permettront simultanément de classifier les échantillons et de sélectionner les caractéristiques discriminantes. Les applications impliquent des données spectrales. En spectrométrie de la masse par exemple, l'identification de pics reliés à un résultat spécifique, comme les pics permettant la discrimination ou prédisant une réponse clinique, est d'intérêt. D'autres contextes pratiques viendront d'études chémométriques qui explorent la possibilité d'utiliser des spectres NIR pour classifier des échantillons. Dans tous les exemples considérés, nous verrons que de très petits ensembles de caractéristiques mènent à de bons résultats en termes de classification.

**[MS-79]****Bayesian concept-varying feature selection for classification with partially labelled data****Sélection bayésienne de caractéristiques selon le concept pour la classification de données partiellement étiquetée**Paul GUSTAFSON, *University of British Columbia*

It is well known that Bayesian methods are good at preventing overfitting. Both the use of shrinkage priors to downweight some variables, and the use of Bayesian model averaging to eliminate some variables, have been well-studied in the literature. Most of this work, however, is in the context of regression modelling to do supervised learning. In this talk we discuss modelling and computational issues that arise when applying these techniques to semi-supervised or unsupervised learning scenarios. We particularly emphasize the idea that with many classes, a subset of variables useful in identifying items from one class might differ from a

Il est bien connu que les méthodes bayésiennes sont efficaces pour prévenir la sur-modélisation. L'utilisation de densités a priori de rétrécissement pour enlever du poids à certaines variables et l'utilisation de modèle bayésien pour éliminer des variables ont été très étudiées dans la littérature. La majeure partie de ces travaux est faite dans le contexte de la régression pour faire de l'apprentissage supervisé. Dans cette présentation, nous discutons des problèmes de modélisation et d'informatiques qui surviennent lorsqu'on applique ces techniques à des scénarios d'apprentissage semi supervisés ou non supervisés. Nous mettons un emphase particulier sur l'idée qu'avec plusieurs classes, un sous ensemble de variables qui identifie bien les éléments d'une classes peu être

subset useful for a different class. Ideas are illustrated in the context of an object-recognition problem from machine learning.

bien différent de celui utile pour discriminer les éléments d'une autre classe. Nous illustrons les idées dans le contexte de la reconnaissance d'objets avec des algorithmes d'apprentissage.

**[MS-80]**

**Model selection for geostatistical models**

**Sélection de modèle pour des modèles géostatistiques**

Andrew MERTON, Jennifer HOETING & Richard DAVIS, *Colorado State University*, Sandra THOMPSON, *Pacific Northwest National Lab*

We consider the problem of model selection for geospatial data. Spatial correlation is often ignored in the selection of explanatory variables and this can influence model selection results. The inclusion or exclusion of particular explanatory variables may not be apparent when spatial correlation is ignored. We consider the Akaike Information Criterion (AIC) as applied to a geostatistical model. We offer a heuristic derivation of the AIC in this context and provide simulation results that show that using AIC for a geostatistical model is superior to the often used traditional approach of ignoring spatial correlation in the selection of explanatory variables.

Nous considérons le problème de sélection de modèle pour des données géospatiales. La corrélation spatiale est souvent ignorée pour la sélection de variables explicatives et ceci peut influencer les résultats de la sélection de modèle. L'inclusion ou l'exclusion de certaines variables explicatives peut ne pas être apparent lorsqu'on ignore la corrélation spatiale. Nous considérons le critère d'information d'Akaike (AIC) appliqué au modèle géostatistique. Nous donnons une dérivation heuristique du critère AIC dans ce contexte et présentons des résultats de simulations qui montrent que l'utilisation du critère AIC pour un modèle géostatistique est supérieure à l'approche usuelle qui ignore la corrélation spatiale pour la sélection de variables explicatives.

**Session 06B**

**Mardi 14 juin • Tuesday, June 14, 11:00 – 12:30**

**217**

Extensions of policy-relevant analytical methods for complex survey data

Extensions des méthodes analytiques dépendant des politiques pour les données d'enquêtes complexes

**[MS-81]**

**A comparison of latent variable modeling approaches in social science research: A Monte Carlo simulation and an illustration using data from youth in transition survey**

**Comparaison des techniques de modélisation de variables latentes en recherche en sciences sociales : une simulation de Monte Carlo et une illustration en utilisant des données d'un sondage sur les jeunes en transition**

Irene Lu, *York University*

The study of the relationships among latent variables is very common in social science research. This presentation reviews two approaches for analyzing latent variable regression: a two-step approach featuring regression scale scores, and a combined Item Response Theory and Structural Equation Modeling (IRT-SEM) approach. A Monte Carlo simulation is conducted to quantify and compare the bias in regression estimates for the two ap-

L'étude des relations entre des variables latentes est standard en recherche en sciences sociales. Cette présentation passe en revue deux approches pour analyser une régression de variables latentes : une approche en deux étapes avec des scores d'échelle de régression et une approche combinée de la théorie de réponse aux items et de la modélisation d'équations structurelles (IRT-SEM). Une simulation de Monte Carlo est effectuée pour quantifier et comparer le biais des estimés de régression pour les deux approches. Les différentes proprié-

proaches. The different bias properties of the two methods will also be illustrated using data from Youth In Transition Survey (YITS).

tés de biais des deux méthodes sont également illustrées à partir des données du sondage sur les jeunes en transition (YITS).

**[MS-82]**

**Bivariate density estimation from a longitudinal complex survey: example of testing for temporal association of job loss and divorce using SLID**

**Estimation d'une densité bivariée à partir d'un sondage longitudinal complexe : exemple de test pour l'association temporelle entre la perte d'emploi et le divorce en utilisant l'EDTR**

Norberto PANTOJA GALICIA & Mary THOMPSON, *University of Waterloo*, Milorad KOVACEVIC, *Statistics Canada*

The relationship between job loss and divorce is a topic of interest among the social scientists. The longitudinal data from the Survey of Labour and Income Dynamics (SLID) gives the possibility to approach this issue in a variety of ways. Dates of changes of the marital status (with associated type of change) as well as dates of employment can be obtained for a longitudinal person in a six years panel. For a person who is employed and married let  $T_1$  and  $T_2$  be the times of losing a job and becoming divorced respectively. A formal nonparametric test for a partial order relationship proposed by Thompson and Pantoja Galicia (2002) requires estimation of the joint density for  $T_1$  and  $T_2$ . This test is modified to account for complexities of the sample design, and to allow the observed times-to-events to be interval censored.

La relation entre la perte d'emploi et le divorce est un sujet qui suscite beaucoup d'intérêt chez les sociologues. Les données longitudinales provenant de l'Enquête sur la dynamique du travail et du revenu (EDTR) donnent la possibilité d'aborder cette question de plusieurs manières différentes. Les dates de changement de statut social (avec les types associés) ainsi que les dates concernant l'emploi peuvent être obtenues pour une personne sur un panel de six ans. Pour une personne mariée avec un emploi, établissons  $T_1$  et  $T_2$  les temps respectifs de la perte d'emploi et du divorce. Un test non paramétrique formel pour un relation d'ordre partielle a été proposé par Thompson et Pantoja Galicia (2002) mais requiert l'estimation de la densité conjointe pour  $T_1$  et  $T_2$ . Le test que nous proposons est modifié pour tenir compte de la complexité du plan d'échantillonnage et pour permettre aux temps jusqu'aux événements d'être censurés par intervalle.

**[MS-83]**

**Some advances in the use of marginal models with data from a longitudinal survey having a complex design**  
**Quelques avancements dans l'utilisation des modèles marginaux avec des données provenant d'un sondage longitudinale avec un plan complexe**

Qunshu REN, *Carleton University*

Many longitudinal surveys are being conducted by government agencies to provide data for analytic studies. At Statistics Canada, longitudinal surveys on health, child development, labour, etc. now have available several cycles of data on the same samples. Since the complexity of the survey designs induces cross-sectional dependencies as well as dependence over time, analytical methods for non-survey longitudinal data must be modified for use with survey data. This paper will show some advances in appropriate methods for using marginal models (sometimes called population-average models) with survey data. Assessing goodness-of-fit, using bootstrap variance

Plusieurs sondages longitudinaux sont effectués par des agences gouvernementales pour fournir en données des études analytiques. À Statistique Canada, les sondages longitudinaux sur la santé, le développement des enfants, le travail, etc ont présentement plusieurs cycles de données dans les mêmes échantillons. Puisque la complexité du plan d'échantillonnage induit une dépendance trans-sectionnelle ainsi qu'une dépendance dans le temps, les méthodes analytiques pour de données longitudinales ne provenant pas de sondages doivent être modifiées pour être utilisées sur des données de sondage. Cet article illustre quelques avancements dans les méthodes appropriées pour utiliser les modèles marginaux (également appelé modèles population-moyenne) avec des données de sondage. Nous discutons

estimation and using the odds ratio as a measure of association will all be discussed in the case of binary responses and marginal logistic regression models, and will be illustrated using data from the National Population Health Survey (NPHS).

également de l'évaluation de l'ajustement du modèle en utilisant l'estimation bootstrap de la variance et en utilisant le rapport de cote comme mesure d'association dans le cas d'une réponse binaire avec modèle de régression logistique marginal. La méthode sera illustrée avec des données provenant du sondage national sur la santé de la population (SNSP).

## Session 06C

Mardi 14 juin • Tuesday, June 14, 11:00 – 12:30

214

Statistical genomics: New platforms, new data, new analysis methods for high-throughput experiments

La génomique statistique : Nouvelles plateformes, nouvelles données, nouvelles méthodes d'analyse pour les expériences à haut débit.

[MS-84]

**Statistical methods for the analysis of tissue microarray data.**

**Méthodes statistiques pour analyser des données de micro-réseaux de tissus**

Steve HORVATH, *University of California, Los Angeles*

Tissue microarrays (TMAs) are a new high-throughput tool for the study of protein expression patterns in tissues and are increasingly used to evaluate the diagnostic, prognostic importance of tumor biomarkers. TMA data are rather challenging. Covariates are either ordinal variables (range 0-4) or highly skewed percentages. Tumor marker expression values are usually highly correlated. Since it is standard practice in the TMA community to use cut-off values for tumor marker expression values, it is natural to apply tree-based methods. We describe different supervised and unsupervised learning methods based on survival trees and random forests (Breiman 2001). We describe a novel strategy (random forest clustering) for tumor profiling based on tissue microarray data. Random forest clustering is attractive for tissue microarray and other immunohistochemistry data since it handles highly skewed tumor marker expressions well and weighs the contribution of each marker according to its relatedness with other tumor markers. This is the first tumor class discovery analysis of renal cell carcinoma patients based on protein expression profiles.

Les micro-réseaux de tissus (MRT) sont des nouveaux outils à hauts résultats pour étudier les patterns d'expressions des protéines dans les tissus et sont utilisés de plus en plus pour évaluer l'importance du diagnostique et du pronostique de bio-marqueurs de tissus. Les données de MRT sont un challenge. Les covariables sont soit des variables ordinales (niveaux 0 à 4) ou des pourcentages fortement asymétriques. Les valeurs d'expression des marqueurs de tumeurs sont habituellement fortement corrélées. Puisqu'il est pratique standard dans la communauté des MRT d'utiliser des valeurs de seuils pour les valeurs d'expressions des marqueurs de tumeurs, il est naturel d'appliquer des méthodes d'arborescence. Nous décrivons différentes méthodes d'apprentissage supervisés et non supervisés basées sur les arbres de survie et les forêts aléatoires (Breiman 2001).

Nous décrivons une nouvelle stratégie (classification de forêts aléatoires) pour le profilage de tumeurs basé sur des données de micro-réseaux de tissus. La classification de forêts aléatoires est attrayante pour les MRT et d'autres données immuno-histochimies puisqu'elle supporte bien des expressions de marqueurs de tumeurs fortement asymétriques et pondère la contribution de chaque marqueur selon sa relation avec d'autres marqueurs de tumeurs. Ceci est la première classe d'analyse de découverte de tumeurs pour des patients avec carcinome des cellules rénales basée sur le profil d'expression des protéines.

**[MS-85]****Genomic databases and microarray data: One statistician's approach at integration.****Bases de données génomiques et données de microréseaux : l'approche d'un statisticien pour l'intégration**Rafal KUSTRA, *University of Toronto*

The success of Statistical Genomics will depend on our ability to incorporate into our modelling, external information about the biological processes and other domain knowledge of the field. Although the challenge is not unique to Statistical Genomics, it has a specific meaning and importance due to few factors. With high-dimensional data, we need to control the variance by reducing the complexity of model space. Other factor is the availability of relevant information in public databases. In this talk we will discuss our approaches in incorporating GeneOntology data structure in analysis of microarray data. Initial results indicate that incorporating the information about the relationships of biological processes of known genes, may be an effective variance-controlling device.

Le succès de la génomique statistique dépend de notre habileté à incorporer dans nos modèle de l'information externe sur les processus biologiques et des connaissances provenant de d'autres domaines sur le sujet. Même si le défi n'est pas unique à la génomique statistique, il a une signification et une importance spécifique à cause de certains facteurs. Un d'entre eux est qu'avec des données à haute dimension, nous devons contrôler la variance en réduisant la complexité de l'espace du modèle. Un autre facteur est la disponibilité d'information utile dans les bases de données publiques. Dans cette présentation, nous discutons de notre approche pour intégrer la structure de données GeneOntology dans les analyses de microréseaux. Les résultats initiaux montrent que l'incorporation d'information sur les relations des processus biologiques de gènes connus peut être une bonne approche pour contrôler la variance.

**Session 06D****Mardi 14 juin • Tuesday, June 14, 11:00 – 12:30****200**

## Computer network modeling

## Modélisation de réseaux informatiques

**[MS-86]****Regret and convergence bounds for continuum-armed bandits****Manque à gagner et frontière de convergence pour le problème du " Continuum-Armed Bandits "**Eric COPE, *UBC Sauder School of Business*

We consider immediate-reward reinforcement learning problems with continuous action sets, also known as "continuum-armed bandit problems." Both lower bounds on the growth rate of the regret as well as upper bounds on the rates of convergence of the control values to the optimum are derived. I explicitly characterize the dependence of these convergence rates on the minimal rate of variation of the mean reward function in a neighborhood of the optimal control. The bounds can be used to demonstrate the asymptotic optimality of the Kiefer-Wolfowitz method of stochastic approximation with regard to a large class of possible mean reward functions.

Nous considérons le problème de renforcement d'apprentissage par récompense immédiate avec un ensemble d'actions continues, également connu sous le nom de problème du "Continuum-Armed Bandits". Les bornes inférieures sur le taux de croissance du manque à gagner ainsi que les bornes supérieures sur les taux de convergence des valeurs contrôles vers l'optimum sont dérivées. Nous caractérisons explicitement la dépendance de ces taux de convergence sur le taux minimal de variation de la fonction de récompense moyenne dans un voisinage du contrôle optimal. Les frontières de convergence peuvent être utilisées pour démontrer l'optimalité asymptotique de la méthode d'approximation stochastique de Kiefer-Wolfowitz par rapport à une grande classe de fonctions de récompenses moyennes potentielles.

**[MS-87]****Real-time origin-destination matrix estimation from partially observed packet/vehicle trajectories  
Estimation de matrices origine-destination en temps réel à partir de trajectoires de paquets ou véhicules partiellement observées**Jaimyoung KWON, *Cal State East Bay*, Pravin VARAIYA, *University of California Berkeley*

Accurate and timely information on Origin-Destination (OD) traffic matrices is critical in engineering tasks for both computer and transportation networks. Observing OD traffic matrices directly is practically impossible and they have been estimated either from survey sampling or link counts data using methods such as network equilibrium, gravity, or distribution-assignment models. We consider a relatively novel data type, partially observed packet or vehicle trajectories. We develop Method of Moment (MM) algorithm for estimating time-varying OD matrices from such data in real time. Estimator errors are calculated either from asymptotic approximation or bootstrap. Simulation study shows that the proposed estimator and standard errors estimates perform well for small and simple networks. Recently, such partially observed vehicle trajectories data for transportation network became available from electronic toll collection (ETC) infrastructure in San Francisco Bay Area, called FasTrak. The algorithm is applied to these data successfully to derive hourly OD matrices. Non-trivial extension of the algorithm to larger and more complex networks is also discussed.

Une information précise et rapide sur les matrices de trafic origine-destination (OD) est critique en ingénierie des réseaux informatiques et les réseaux de transport. L'observation directe des matrices de trafic OD est pratiquement impossible. Elles ont été estimées soit par échantillonnage ou par des données de compte de liaisons en utilisant des méthodes telles que l'équilibre de réseau, la gravité ou des modèles d'assignation de distributions. Nous considérons un type de données relativement nouveau : des trajectoires de paquets ou véhicules partiellement observées. Nous développons un algorithme par la méthode des moments (MM) pour estimer en temps réel des matrices OD variables dans le temps à partir de telles données. L'erreur d'estimation est calculée soit à partir d'approximations asymptotiques ou de simulations bootstrap. Nous montrons par une étude de simulations que l'estimateur proposé et l'estimation de l'écart-type performe bien pour des réseaux simples et petits. Récemment, de telles données partiellement observées de trajectoires de véhicules pour des réseaux de transport sont disponibles, provenant de postes de paysage électroniques (PPE) appelé FasTrak dans la région de San Francisco. L'algorithme est appliqué avec succès à ces données pour obtenir des matrices OD horaires. Nous discutons également d'une extension non triviale de l'algorithme pour des réseaux plus gros et plus complexes.

**[MS-88]****Flexicast network tomography  
Tomographie de réseaux à diffusion flexible**George MICHAILIDIS, Lawrence EARL & Nair VIJAYAN, *University of Michigan*

Estimation of quality of service parameters, such as packet loss rates and delay distributions, is of considerable importance to network administrators and service providers. We consider the active network tomography problem where the goal is to estimate packet loss rates and delay distributions of all internal links in a network from measurements obtained from nodes located on its periphery. This is an example of a large-scale statistical inverse problem. We provide a brief review of the literature, including unicast and multicast probing schemes and introduce a class of new probing schemes, that are shown to be computationally efficient for collect-

L'estimation de la qualité de paramètres de service, tels que les taux de pertes de paquets et la distribution des délais est d'une grande importance pour les administrateurs de réseaux et les fournisseurs de services. Nous considérons le problème de tomographie de réseaux actifs où le but est d'estimer les taux de pertes de paquets et la distributions des délais de tous les liens internes d'un réseau de mesures obtenues des nœuds situés sur sa périphérie. Ceci est un exemple d'un problème statistique inverse à grande échelle. Nous effectuons une brève revue de la littérature, incluant les plans de vérifications de diffusions individuelles et de multidiffusions et nous introduisons une classe de nouveaux plans de vérifications qui sont efficaces informatiquement pour col-

ing the necessary data in global network monitoring problems. Several methods of estimation for loss rates and delay distributions will be described, including the use of EM-algorithms for the Maximum Likelihood estimators and several classes of least-squares estimates for loss rates. Theoretical and practical issues regarding the design of probing experiments on different network topologies and results from simulation studies will also be discussed.

lecter les données nécessaires dans des problèmes de suivi de réseaux globaux. Plusieurs méthodes d'estimations pour les taux de pertes et les distributions des délais seront décrites, incluant l'utilisation de l'algorithme EM pour l'estimateur du maximum de vraisemblance et plusieurs classes d'estimateurs des moindres carrés pour les taux de pertes. Nous discuterons également de problèmes théoriques et pratiques par rapport au plan d'expériences de vérifications sur différentes topologies de réseaux et des résultats d'études de simulations.

## Session 06E

Mardi 14 juin • Tuesday, June 14, 11:00 – 12:30

241

Statistics in cancer, AIDS, and other medical researches

Statistique pour le cancer, le SIDA, et autres recherches médicales

[MS-89]

**Early detection of disease and the scheduling of examinations**

**Détection rapide de maladies et calendrier des examens**

Marvin ZELÉN, *Harvard University and the Dana-Farber Cancer Institute*, Sandra LEE & Hui HUANG, *Dana-Farber Cancer Institute*

Special examinations exist for many chronic diseases that can diagnose the disease while it is asymptomatic with no signs or symptoms. The examinations are usually characterized by the age to begin the special exams and the spacing between subsequent examinations. The earlier detection of disease may lead to more cures or reduced mortality. This possibility has led to public health programs which recommend that populations (usually defined by age or risk status) follow a schedule of periodic exams in order to detect specific chronic diseases; e.g., cancer, diabetes, cardiovascular diseases, etc. Such examination schedules, when embedded in a public health program, are invariably costly. The possible candidate number of examinations is so large that it is not feasible to carry out clinical trials in order to identify optimal schedules. Instead this problem may be investigated by developing theoretical models that can predict the eventual disease mortality for different examination schedules. We have developed a stochastic model for this purpose. The model assumptions assume that the natural history of the disease is progressive and that any benefit from earlier diagnosis is due to a change in the distribution of disease stages at diagnosis. The model is general and can be applied to any disease that satisfies the model assumptions.

Il existe des examens spéciaux pour plusieurs maladies chroniques qui peuvent diagnostiquer cette maladie lorsqu'elle est asymptomatique avec aucun symptôme. Les examens sont souvent caractérisés par l'âge pour commencer l'examen spécial et le délai entre les examens subséquents. La détection plus rapide d'une maladie peut mener à un plus grand nombre de traitements possibles ou bien réduire la mortalité. Cette possibilité a mené à la mise sur pieds de programmes de santé publique qui recommandent à la population (habituellement définie par l'âge ou le statut de risque) de suivre un calendrier d'examens périodiques pour détecter des maladies chroniques spécifiques, par exemple le cancer, le diabète, des maladies cardiovasculaires, etc. De tels calendriers d'examens lorsqu'ils font partie d'un programme de santé publique, sont invariablement coûteux. Le nombre possible d'examens est tellement grand qu'il n'est pas possible d'effectuer des essais cliniques pour identifier les calendriers d'examens optimaux. Au lieu de cela, ce problème peut être investigué en développant des modèles théoriques qui peuvent prédire la mortalité éventuelle dû à la maladie pour différents calendriers d'examens. Nous avons développé un modèle stochastique pour cela. Les hypothèses du modèle supposent que l'histoire naturelle de la maladie est progressive et que les avantages d'un diagnostic rapide sont dû à un changement dans la distribution des niveaux de la maladies au moment du diagnostic. Le modèle est général et peut être appliqué à n'importe quelle maladie



The model is applied to breast cancer screening. We address the: (1) benefit of screening women in the age group 40–49 (currently there is controversy over benefit in this age group), (2) schedules for women ages 40–79, (3) schedules for women 50–79 for which there is general agreement about benefit and (4) schedules for women at elevated risk. We also investigate schedules with which are staggered according to age. This may be important for diseases in which the risk increases with age.

qui satisfait les hypothèses du modèle. Le modèle est appliqué au dépistage du cancer du sein. Nous adressons : (1) les avantages de suivre les femmes de 40 à 49 ans (il y a présentement de la controverse sur les avantages de dépister dans ce groupe d'âge), (2) le calendrier pour les femmes de 40 à 79 ans, (3) le calendrier pour les femmes de 50 à 79 pour lesquelles il y a un accord général sur les avantages et (4) le calendrier pour les femmes à risque élevé. Nous investiguons également les calendriers qui sont décalés en fonction de l'âge. Ceci peut être important pour les maladies dont le risque augmente avec l'âge.

### [MS-90]

#### **Analysis of censored survival data with high-dimensional covariates** **Analyse de données de survie censurées avec plusieurs covariables**

Jian HUANG, *University of Iowa*, Shuangge MA, *University of Washington*

Analysis of censored survival data with a large number of covariates is an important practical problem, especially now microarrays that can assay tens of thousands of genes are becoming a routine tool in the studies of various types of cancers and many other diseases. How to associate gene expression data with clinical outcomes such as patients' survival and identify important genes or clusters of genes that are functionally related presents a class of interesting and challenging problems in survival analysis. We propose using the gradient search method of Friedman and Popsecu (2004) in the context of ridge regression for variable selection, parameter estimation, and prediction. This method can be applied to a wide range of survival analysis models including the Cox model, the accelerated failure time mode, and the additive hazards model. We use simulations and two examples to illustrate the application of the proposed method. This is joint work with Dr. Shuangge Ma.

L'analyse de données de survie censurées avec un grand nombre de covariables est un problème pratique important, spécialement depuis que les études de micro-réseaux pouvant évaluer des dizaines de milliers de gènes sont devenues des outils communs pour l'étude de divers types de cancers et plusieurs autres maladies. L'association des données d'expressions de gènes avec des résultats cliniques tels que la survie du patient et l'identification de gènes importants ou de groupes de gènes qui sont fonctionnellement reliés présente une classe de problèmes et des défis intéressants en analyse de survie. Nous proposons d'utiliser la méthode de recherche par le gradient de Friedman et Popsecu (2004) dans le contexte de la méthode de régression ridge pour la sélection de variables, l'estimation des paramètres et la prévision. Cette méthode peut être appliquée à une vaste gamme de modèle de survie incluant le modèle de Cox, le modèle de temps de vie accéléré et le modèle de taux de panne additif. Nous utilisons des simulations et deux exemples pour illustrer l'application de la méthode proposée. Ceci est un travail conjoint avec Dr. Shuangge Ma.

### [MS-91]

#### **Probability distributions in infectious disease transmission models and risks of major outbreaks** **Distributions de probabilité pour des modèles de transmission de maladies infectieuses et risques d'épidémies majeures**

Ping YAN, *Public Health Agency of Canada*

When an infectious agent enters a susceptible population of size  $n$ , with probability  $p$ , the outbreak of a disease terminates with few cases and the outbreak size as a proportion,  $f$ , concentrates at zero for large  $n$ . With probability  $1-p$ , the initial growth

Un agent infectieux entre dans une population de taille  $n$ . Avec probabilité  $p$ , l'épidémie d'une maladie se termine avec quelques propagations et la taille de l'épidémie à une proportion  $f$ , qui converge vers zéro pour un grand  $n$ . Avec probabilité  $1-p$ , la croissance initiale du nombre d'individus

of infected individuals over time may be approximated by an exponential rate  $r$  and the outbreak size as a number,  $nf$ , scales linearly with  $n$ , while  $f > 0$ .  $N$  is the number of infections produced by an infective individual throughout its infectious period. The talk presents models for  $\Pr\{N = n\}$  and compares  $p$  and  $r$  under various assumptions on the point process for contacts, the transmission probability and the infectious period distribution.

infectés dans le temps peut être approximée par une loi exponentielle de taux  $r$  et la taille de l'épidémie par un nombre  $nf$ .  $N$  est le nombre d'infections produites par un individu infecté à travers sa période de contagion. Cette présentation donne des modèles pour  $\Pr\{N = n\}$  et compare  $p$  et  $r$  sous plusieurs hypothèses pour le processus ponctuel de contacts, la probabilité de transmission et la distribution de la période de contagion.

## Session 06F

Mardi 14 juin • Tuesday, June 14, 11:00 – 12:30

211

Statistics and finance  
Statistique et finance

[MS-92]

**Missing and incomplete data in finance**

**Données manquantes et incomplètes en finance**

Don McLEISH & Joe DiCESARE, *University of Waterloo*

In a sense, every data-set in finance involves missing data. Apart from the obvious, missing, deleted, or erroneous values, financial time series are often observed only at discrete time points, stock and bond prices may be considered as established only when a sale takes place; there are long periods of time when markets are closed, at least in our part of the world; and there is always covariate information (highs and lows, prices of other instruments) related to the problem at hand that that is either not collected, or collected and not used. For example when one market is closed, other markets are useful covariates relevant to “filling in” the missing observations. There is, however, little work on the use of missing data techniques for financial time series. In this talk, we investigate some of the uses of imputation, simulation and other techniques to estimate parameters in a diffusion and to conduct inference.

Dans un certain sens, toutes les bases de données en finance impliquent des données manquantes. Mise à part les données manquantes évidentes, effacées ou erronées, les séries chronologiques financières sont souvent observées à des périodes de temps discrètes, on peut considérer que le prix des actions et des obligations peut être établi seulement lors d'une vente, il y a des longues périodes de temps où les marchés sont fermés, du moins dans notre partie du globe et il y a toujours des variables (hauts et bas, prix de d'autres indices) reliées au problème qui ne sont pas collectées ou non utilisées. Par exemple, quand un marché est fermé, l'information provenant des autres marchés peut permettre l'imputation des données manquantes. Il y a cependant peu de travaux effectués sur l'utilisation des techniques de données manquantes pour des séries chronologiques financières. Dans cette présentation, nous étudions l'utilisation de techniques telles l'imputation, la simulation et certaines autres techniques pour estimer les paramètres d'un processus de diffusion et pour faire de l'inférence.

[MS-93]

**Self-financing in a fractional Brownian motion market**

**Autofinancement dans un marché de mouvement brownien fractionnel**

Donna SALOPEK, *University of New South Wales*

In this talk, we discuss the various definitions of self financing in a fractional Brownian motion market and their implications.

Dans cette présentation, nous discutons des différentes définitions d'autofinancement dans un marché de mouvement brownien fractionnel et de leurs implications.

**Session 06G****Mardi 14 juin • Tuesday, June 14, 11:00 – 12:30****212**Estimation  
Estimation**[MS-94]****Minimum Hellinger distance estimation for a nonparametric mixture model****Estimation à distance d'Hellinger minimale pour un mélange de modèles non paramétriques**Rohana KARUNAMUNI & Jingjing WU, *University of Alberta*

We consider the problem of estimating a mixture proportion  $p$  in a nonparametric mixture of the form  $pF(x) + (1 - p)G(x)$ , where  $F$  and  $G$  are unknown distributions. We assume that data from the two distributions  $F$  and  $G$  as well as from the mixture  $pF(x) + (1 - p)G(x)$  are available. In this paper, we develop a minimum Hellinger distance estimator of  $p$ . Theoretical properties such existence, consistency and asymptotic normality of the estimator are investigated.

Nous considérons le problème qui consiste à estimer une proportion de mélange  $p$  dans un mélange non paramétrique de la forme  $pF(x) + (1 - p)G(x)$ , où  $F$  et  $G$  sont des distributions inconnues. Nous supposons que des données provenant des deux distributions  $F$  et  $G$  ainsi que du mélange  $pF(x) + (1 - p)G(x)$  sont disponibles. Dans cet article, nous développons un estimateur à distance d'Hellinger minimale pour  $p$ . Des propriétés théoriques telles l'existence, l'efficacité et la normalité asymptotique de l'estimateur sont analysées.

**[MS-95]****Nonparametric estimation of mixing proportion and component distributions in multivariate mixtures****Estimation non paramétrique de la proportion de mélange et distributions des composantes dans des mélanges multivariés**Reza PAKYARI, *Australian National University*

In this paper, we propose a novel nonparametric estimation method in multivariate mixtures. We consider the nonparametric estimation of the mixing proportion and the component distribution functions in two-term multivariate mixture models. Our method is based on solving simultaneous equations, involving sub-populations, that arise from a multivariate mixture model under the assumption of independent marginals. We show that the proposed estimators are root- $n$  consistent. The numerical properties of the estimators are also studied using Monte Carlo simulations. This is a joint work with Peter Hall at Australian National University.

Dans cet article, nous proposons une nouvelle méthode d'estimation non paramétrique pour les mélanges multivariés. Nous considérons l'estimation non paramétrique de la proportion de mélange et la distribution des composantes dans des mélanges de modèles multivariés à deux termes. Notre méthode est basée sur la solution d'équations simultanées impliquant des sous populations, qui proviennent d'un mélange de modèles multivariés sous l'hypothèse de marginales indépendantes. Nous montrons que les estimateurs proposés sont efficaces au sens de la racine d'ordre  $n$ . Les propriétés numériques des estimateurs sont également étudiées en utilisant des simulations de Monte-Carlo. Ceci est un projet conjoint avec Peter Hall de l'Australian National University.

**[MS-96]****Second-order least-squares estimation in nonlinear regression****Moindres carrés du second ordre en régression non-linéaire**Alexandre LEBLANC & Liqun WANG, *University of Manitoba*

In the context of nonlinear regression, ordinary least-squares estimation (OLS) is based on minimization of the squared distance of the response

Dans un contexte de régression non-linéaire, l'estimation par la méthode des moindres carrés ordinaires (MCO) est basée sur la minimisation du carré de la distance entre la variable

variable from its conditional mean given the predictor variable. We extend this method by also considering the squared distance between the squared response variable and its second conditional moment. Under no parametric distributional assumptions for the random error, the strong consistency and asymptotic normality of the resulting second-order least-squares (SLS) estimator are established. Finally, the asymptotic properties of the SLS estimator are compared with those of the widely used OLS estimator.

indépendante et son espérance conditionnelle étant donné la variable prédictive. Nous généralisons cette méthode en considérant aussi le carré de la distance entre le carré de la variable indépendante et son deuxième moment conditionnel. En ne faisant aucune hypothèse paramétrique sur la distribution des erreurs aléatoires, la convergence presque sûre et la normalité asymptotique de cet estimateur des moindres carrés du second ordre (MCS) sont obtenues. Finalement, les propriétés asymptotiques de l'estimateur MCS sont comparées à celles du très populaire estimateur MCO.

### [MS-97]

#### **Estimating simple linear regression model parameters using two-stage ranked set sampling schemes**

#### **Estimation des paramètres d'un modèle de régression linéaire simple en utilisant un plan d'échantillonnage en deux phases d'ensembles ordonnés**

Elies KOUIDER, Mohammed EL-SAIDI, Michael COOPER & Nathaniel TYMES, *Ferris State University*

For estimating the parameters of a simple linear regression model when replications exist, the ranked set sampling as well as the double ranked set sampling have been used and favored for their advantage over the classical order statistics of a simple random sample. In this paper, we introduce two-stage ranked set sampling schemes and we derive the best linear unbiased estimators of the simple linear regression model for different cases. Additionally, as an illustration, we apply the proposed methods and report the substantial gain in efficiencies when the underlying distribution of the error terms is either normal or Laplace.

Pour estimer les paramètres d'un modèle de régression linéaire simple lorsqu'il y a des répliques, l'échantillonnage d'ensembles ordonnés ainsi que l'échantillonnage par double ensembles ordonnés ont été utilisés et favorisés pour leurs avantages par rapport aux statistiques d'ordres classiques d'un échantillon aléatoire simple. Dans cette présentation, nous introduisons des plans d'échantillonnage en deux étapes d'ensembles ordonnés et nous obtenons le meilleur estimateur linéaire sans biais du modèle de régression linéaire simple pour différents cas. De plus, à titre d'illustration, nous appliquons les méthodes proposées et rapportons le gain substantiel en efficacité lorsque la distribution sous jacente de l'erreur est soit normale ou Laplace.

### [MS-98]

#### **A multivariate generalization of the Hodges-Lehmann estimator**

#### **Généralisation multivariée de l'estimateur de Hodges-Lehmann**

Shojaeddin CHENOURI, *Carleton University*

The Hodges-Lehmann estimator is a very well-known and well-explored robust and efficient estimate of location in one sample univariate problems. In this talk we shall introduce a multivariate generalization of the Hodges-Lehmann estimator and derive its asymptotic normality using the  $U$ -process approach. We shall show that this multivariate generalization is robust as well.

L'estimateur de Hodges-Lehmann est bien connu comme étant un estimateur de position robuste et efficace, qui a été grandement exploré pour des problèmes à un échantillon univarié. Dans cet exposé nous introduisons une généralisation multivariée de l'estimateur de Hodges-Lehmann et nous dérivons la normalité asymptotique en utilisant l'approche par processus- $U$ . Nous montrons que la généralisation multivariée est également robuste.

**[MS-99]****Estimation of regression parameters: Parallelism restrictions****Estimation des paramètres de régression : restrictions de parallélisme**Bashir KHAN, *Saint Mary's University*, A.K. Md. E SALEH, *Carleton University*

In this paper, we consider two regression models. Our problem is the estimation of slope and intercept vectors of Model I, when it is a priori suspected that slopes of the two models may be equal. We propose five estimators namely; unrestricted, restricted, preliminary test, James-Stein type and the positive version of James-Stein estimators for the estimation of the slope and intercept vectors of Model I. The quadratic bias functions, mean square error matrices and risk functions of the proposed estimators are derived. We study the properties of these estimators and do their comparison with classical estimator through mean square error matrices and quadratic risk functions criteria.

Dans cette présentation, nous considérons deux modèles de régression. Notre problème consiste en l'estimation des vecteurs de pente et d'ordonnée à l'origine pour le modèle I, lorsque l'on suspecte a priori que les pentes des deux modèles peuvent être égales. Nous proposons cinq estimateurs soient : non restreint, restreint, test préliminaire, de type James-Stein et la version positive des estimateurs de James-Stein pour l'estimation du vecteur de pente et d'ordonnée à l'origine du modèle I. Les fonctions de biais quadratiques, les matrices d'erreurs quadratiques moyennes et les fonctions de risque des estimateurs proposés sont dérivées. Nous étudions les propriétés de ces estimateurs et les comparons avec des estimateurs classiques à travers des critères de matrices d'erreur quadratique moyenne et de fonctions de risque quadratique.

**Session 07A****Mardi 14 juin • Tuesday, June 14, 13:30 – 15:00****133****Statistics and technology****Statistique et technologie****[MS-100]****Statistics and technology****Statistique et technologie**Jerry LAWLESS, *University of Waterloo*

Statistics has long been used for technological purposes such as acceptance sampling, process improvement and reliability. More recent examples include fraud detection, risk scoring and biometric profiling. This talk will review the history of statistics in technology and make some general observations based on the premise that technological problems involve learning about systems and processes through experimentation or data collection, and making decisions or predictions that have economic consequences. The relevance of statistical models and inference, computation, and education will be discussed, using illustrations from a variety of areas.

Les statistiques ont longtemps été utilisées pour des fins technologiques tels que l'échantillonnage d'acceptation, l'amélioration des processus et la fiabilité. Des exemples plus récents incluent la détection de fraudes, la gestion du risque et le profilage biométrique. Cette présentation passe en revue l'histoire des statistiques en technologie et présente des observations générales basées sur l'hypothèse que les problèmes technologiques impliquent un apprentissage sur les systèmes et les processus à travers des expérimentations ou la cueillette de données et une prise de décisions ou prédictions qui ont un impact sur l'économie. La validité des modèles statistiques et l'inférence, le calcul informatique et l'éducation seront discutés en utilisant des illustrations provenant de plusieurs domaines.

**Session 07B****Mardi 14 juin • Tuesday, June 14, 13:30 – 15:00****212**

## Limit theorems

## Théorèmes limites

**[MS-101]****Some remarks on the law of large numbers****Quelques remarques sur la loi des grands nombres**Allan Gut, *Uppsala University*

The basic form of the law of large numbers states that the arithmetic mean of a sequence of independent, identically distributed random variables converges almost surely if and only if the mean is finite, in which case the limit is the mean. In this talk we provide a number of results around this law; results concerning convergence rates, complete convergence, convergence along subsequences, weighted averages, random variables with multidimensional indices, with random indices, and so on.

La forme de base de la loi des grands nombre stipule que la moyenne arithmétique d'une séquence de variables aléatoires indépendantes et identiquement distribuées converge presque sûrement si et seulement si la moyenne est finie, et dans ce cas la limite est la moyenne. Dans cet exposé, nous présentons un certain nombre de résultats sur cette loi. Des résultats concernant la vitesse de convergences, la convergence complète, la convergence le long de sous suites, la convergence des moyennes pondérées dans le cas où les variables ont des indices multiples, des indices aléatoires, et ainsi de suite.

**[MS-102]****Some new LIL type results for variables with infinite variance****De nouveaux résultats de type LIL pour des variables avec variance infinie**Uwe EINMAHL, *Vrije Universiteit Brussels*, Deli LI, *Lakehead University*

We present a very general LIL type result for normalized sums of iid random variables. Our result implies the LIL results of Hartman-Wintner and Klass, but can also be applied in many other situations especially for iid variables with a symmetric distribution given by the density function  $p(x) = 1/|x|^3$ ,  $|x| > 1$ . This example has been considered by several authors, but there seems to be no concrete norming sequence known leading to LIL behavior.

Nous présentons un résultat de type LIL très général pour des sommes normalisées de variables aléatoires iid. Notre résultat implique le résultat LIL de Hartman-Wintner et Klass, mais peut également être appliqué dans plusieurs autres situations, spécialement pour des variables iid avec une distribution symétrique donnée par la fonction de densité  $p(x) = 1/|x|^3$ ,  $|x| > 1$ . Cet exemple a été considéré par plusieurs auteurs, mais il ne semble pas avoir de normalisation concrète connue menant vers un comportement LIL.

Beside this example we look at the question when one has LIL behavior for normalized sums  $S_n / \sqrt{nh(n)}$  where  $h$  is a slowly varying function. Under an additional mild assumption on the function  $h$  we obtain a necessary and sufficient condition for this to happen, thereby showing that the classical Hartman-Wintner LIL is a special case of a much more general result.

À part cet exemple, nous regardons la situation lorsqu'on a un comportement LIL pour des sommes normalisées  $S_n / \sqrt{nh(n)}$ , où  $h$  est une fonction qui varie lentement. Sous une faible hypothèse additionnelle sur la fonction  $h$ , nous obtenons une condition nécessaire et suffisante pour obtenir ce comportement, montrant ainsi que la LIL Hartman-Wintner est un cas spécial d'un résultat beaucoup plus général.

**[MS-103]****Tail behaviour of weighted sums of random variables****Comportement des queues d'une somme pondérée de variables aléatoires**Ulrich STADTMUELLER, *University of Ulm*

Consider a sequence of iid random variables  $(X_k)$ , a sequence or an infinite matrix of weights  $(p_{n,k})$  and the related partial sums  $T_n = \sum_k p_{n,k} X_k$ . We are interested in the tail behaviour of these sums. In case exponential moments exist, we find a more or less explicit large deviation principle, in case of weaker moment assumptions various versions of Baum-Katz-type theorems are appropriate. We will present versions of such results for  $(T_n)$  and we will study in particular weights coming from summability.

Considérons une suite de variables aléatoires iid  $(X_k)$ , une suite ou une matrice infinie de poids  $(p_{n,k})$  et la somme partielle reliée  $T_n = \sum_k p_{n,k} X_k$ . Nous sommes intéressé au comportement des queues de ces sommes. Dans le cas où les moments exponentiels existent, nous trouvons un principe de grandes déviations plus ou moins explicite et dans le cas d'hypothèses plus faibles sur les moments, plusieurs versions de théorèmes de type Baum-Katz sont appropriées. Nous présentons des versions de tels résultats pour  $(T_n)$  et nous étudions en particulier les poids provenant de la sommabilité.

**Session 07C****Mardi 14 juin • Tuesday, June 14, 13:30 – 15:00****241****Statistical methodology in clinical trial design****Méthodologie statistique dans la planification d'essais cliniques****[MS-104]****A sequential procedure that stops clinical trials without later reversal of the early conclusion****Une procédure séquentielle qui arrête un essai clinique sans renverser une conclusion précoce**Xiaoping XIONG, *St. Jude Children's Research Hospital*, Ming TAN, *University of Maryland Greenebaum Cancer Center*, James BOYETT, *St. Jude Children's Research Hospital*

Investigators may concern about the conclusion of test from interim analysis in clinical trials, as the conclusion was drawn with smaller sample size and might be changed to contrary if the trial were not stopped but continued to the end of trial at which the conclusion were facilitated with full sample size. To deal with this concern, the sequential conditional probability ratio test (SCPRT) is equipped with a property that a conclusion made at early stopping is unlikely to be reversed if the trial were not stopped but continued to the planned end. It provides good reason to stop a trial early: i.e., if the trial were not stopped as it should, then adding more data and continuing the trial by the planned end would not change the conclusion. We will illustrate the SCPRT procedure based on Brownian motion with applications to clinical trials with different data endpoints.

Les investigateurs peuvent être préoccupés par les conclusions de tests intérimaires dans des essais cliniques parce qu'elles sont faites à partir d'un échantillon de plus petite taille et qu'elles pourraient changer si l'étude n'avait pas été arrêtée et que les conclusions avaient été tirées à partir de l'échantillon complet. Pour surpasser cette préoccupation, le test progressif du rapport des probabilités conditionnelles (TPRPC) possède la propriété que lorsqu'une conclusion est faite avant la fin de l'étude, il est peu probable que cette dernière soit renversée si l'étude est continuée pour la durée planifiée au départ. Le test donne de bonnes raisons pour arrêter une étude avant la date prévue, c'est à dire que si l'étude est poursuivie à terme, l'ajout de données de changerait pas les conclusions. Nous allons illustrer la procédure du TPRPC basée sur le mouvement Brownien avec des applications aux essais cliniques avec différents temps d'arrêts.

**[MS-105]****Confirmatory non-inferiority studies for binary outcomes  
Études confirmatoires de non infériorité pour des résultats binaires**Zhengqing LI, *Pfizer*

In clinical trials comparing the efficacy of a new antibiotic versus a standard treatment, the primary efficacy endpoint is typically the cure rate at the test of cure visit. In such a study, one needs to perform a non-inferiority test in order to establish the efficacy of the new treatment. In this presentation, we will discuss statistical approaches that are commonly used in designing and analyzing such studies. In addition, we will compare two methods for constructing confidence intervals in non-inferiority analyses using simulations. Some practical guide to individuals responsible for the analysis of data from such trials will be provided.

Dans les essais cliniques qui tentent de comparer l'efficacité d'un nouvel antibiotique versus un traitement standard, le point final d'efficacité est typiquement le taux de guérison au moment de la visite où l'on vérifie cette guérison. Dans une telle étude, on doit effectuer un test de non infériorité pour pouvoir établir l'efficacité du nouveau traitement. Dans cette présentation, nous discutons des approches statistiques qui sont habituellement utilisées pour planifier et analyser de telles études. De plus, nous comparons deux méthodes pour construire des intervalles de confiance pour des analyses de non infériorité à l'aide de simulations. Nous donnons également quelques guides pratiques pour les individus responsables de l'analyse des données de telles études.

**[MS-106]****Sample size determination for testing the mean of skewed distribution  
Détermination de la taille de l'échantillon pour tester la moyenne de distributions asymétriques**Jianrong WU, *St. Jude Children's Research Hospital*

If the primary outcome of a trial is a skewed log-normal response and the primary objective of the trial is to test the mean of the response, one face to determine the required sample size for the trial. However, the sample size calculation for testing the mean of a log-normal variable is not still available in the literature. This is partly because of absence of a satisfied test statistic. The primary goal of this research is to develop some test statistics which are nearly exact tests and can be severed as for the sample size determination.

Si le résultat principal d'un essai est une réponse log-normale asymétrique et si l'objectif principal est de tester la moyenne des réponses, alors nous faisons face à un problème de détermination de la taille de l'échantillon. Cependant, le détail du calcul de la taille de l'échantillon pour tester la moyenne d'une variable log-normale n'est présentement pas répertorié dans la littérature. Cela est dû en partie à l'absence d'une statistique de test. Le but premier de cette recherche est de développer quelques statistiques de tests qui sont presque des tests exacts et qui peuvent servir à la détermination de la taille de l'échantillon.

**Session 07D                      Mardi 14 juin • Tuesday, June 14, 13:30 – 15:00                      200****Applications of mixture models in genetic linkage analysis  
Application des mélanges de modèles en analyse de liaison génétique****[MS-107]****Genetic analysis of ordinal traits  
Analyse génétique de traits ordinaux**Heping ZHANG & Rui FENG, *Yale University*

Genetic mechanisms underlying many human diseases and conditions. The existing genetic analysis methods require, however, that the diseases or con-

Les mécanismes génétiques sont sous jacents à plusieurs maladies et conditions humaines. Les méthodes d'analyses génétiques existantes requièrent que les maladies ou les condi-



ditions must either be dichotomized or measured by a quantitative trait such as blood pressure for hypertension. In the latter case, normality is generally assumed for the trait. On the other hand, many diseases and conditions are rated on ordinal scales such as cancer and mental and behavioral conditions. Our objective is to establish a framework to conduct genetic analysis for ordinal traits. We proposed and exploited a latent variable, proportional odds logistic model that relates inheritance patterns to the distribution of the ordinal trait. I will present simulation studies and real examples to demonstrate that the power of our proposed model to detect genetic effects was substantially higher than other methods based on binary traits. I will also present statistical challenges for understanding the asymptotic distributions of some test statistics and for computing the statistics.

tions soient dichotomiques ou mesurées par un trait quantitatif, tel que la pression artérielle pour l'hypertension par exemple. Dans ce dernier cas, la normalité est généralement supposée pour le trait. D'un autre côté, plusieurs maladies et conditions sont notées sur une échelle ordinale tel que le cancer et les conditions mentales et comportementales. Notre objectif est d'établir un cadre pour conduire des analyses génétiques pour des traits ordinaux. Nous proposons et utilisons une variable latente, un modèle logistique de cote proportionnelle qui relie des patterns d'héritage à la distribution du trait ordinal. Je vais présenter des simulations et des exemples réels pour démontrer que la puissance du modèle que nous proposons pour détecter des effets génétiques est substantiellement plus élevée que pour d'autres modèles basés sur des traits binaires. Je vais également présenter les défis statistiques pour comprendre les distributions asymptotiques de quelques statistiques de tests et également pour calculer ces statistiques.

#### [MS-108]

##### **Likelihood ratio tests for mixture models with applications to genetic linkage analysis**

##### **Tests du ratio de vraisemblance pour des modèles de mélanges avec des applications aux analyses de liaisons génétiques**

Yongzhao SHAO, Jing HAN & Mengling LIU, *New York University*, Wenbin LU, *North Carolina State University*

Genetic linkage and association analyses are crucial steps toward identifying genetic markers that are associated with the disease. Detecting genetic linkage via likelihood analysis of mixture models has received enormous attention due to several reasons. One reason is identifying disease genes often must overcome many practical challenges, including genetic heterogeneities for which the most natural statistical tools are finite mixture models. Another reason is the likelihood ratio test (LRT) has served as a benchmark of classical linkage analysis; however, the LRT for mixture models has long been known as non-standard and complicated. In this talk we will present some recent advances concerning likelihood based linkage and association analysis.

Les analyses de liaisons et d'associations génétiques sont des étapes cruciales pour identifier les marqueurs qui sont associés à une maladie. La détection des liaisons génétiques à partir d'analyses de vraisemblance de modèles de mélanges a reçu beaucoup d'attention pour plusieurs raisons. Une des raisons est que l'identification de gènes de maladies doit souvent surpasser plusieurs problèmes pratiques, incluant l'hétérogénéité génétique pour laquelle les outils statistiques les plus naturels sont les modèles de mélanges finis. Une autre raison est que le test de ratio de vraisemblance (TRV) a servi comme un seuil de référence pour les analyses de liaisons classiques, cependant le TRV pour des mélanges de modèles a longtemps été vu comme non standard et compliqué. Dans cette présentation, nous allons présenter quelques développements récents concernant les analyses de liaisons et d'associations basées sur la vraisemblance.

#### [MS-109]

##### **On using quadratic risk to select high dimensional models**

##### **Utilisation du risque quadratique pour sélectionner des modèles à haute dimension**

Surajit RAY, *University of North Carolina at Chapel Hill and SAMSI*, Bruce LINDSAY, *Pennsylvania State University*

This paper summarizes the results on using the quadratic distance as the loss function of the risk estimate. This risk is then used for selecting the

Cet exposé résume des résultats sur l'utilisation de la distance quadratique comme fonction de perte pour l'estimation du risque. Le risque est alors utilisé pour sélectionner le

most parsimonious model among the competing models. Although this tool can be applied to any model selection scenario, in this paper we use risk functions to select the number of components in a multivariate mixture model. Moreover, using different tuning parameters for the quadratic kernels we can analyze a dataset at several levels of smoothness. Further, we show that the quadratic risk can be decomposed into model-lack-of fit and parameter estimation cost. Finally, using a similar idea as the Mallows's  $C_p$  in the context of variable selection, we select the model which achieves either the lowest estimated risk or achieves nearly zero model lack of fit. The risk-based tool was applied to find the number of clusters in a gene-expression data (Chitikala et al; Molecular Cell; 2002), which came from an experiment that was designed to investigate the genome-wide interplay of several TBP (TATA binding protein)-interacting inhibitors in the yeast genome.

modèle le plus parcimonieux parmi les modèles concurrents. Même si cet outil peut être appliqué à n'importe quel scénario de sélection de modèles, nous regardons le cas où les fonctions de risque sont utilisées pour sélectionner le nombre de composantes dans un mélange de modèles multivariés. De plus, en utilisant différents paramètres d'ajustements pour le noyau quadratique, nous pouvons analyser un jeu de données à plusieurs niveaux de lissage. Nous montrons également que le risque quadratique peut être décomposé en coûts du manque d'ajustement du modèle et d'estimation des paramètres. Finalement, en utilisant une idée similaire au  $C_p$  de Mallows dans le contexte de la sélection de variable, nous sélectionnons le modèle qui a soit le plus petit risque estimé ou qui a presque aucun manque d'ajustement au modèle. L'outil basé sur le risque est appliqué au problème de trouver le nombre de grappes pour des données d'expression de gènes (Chitikala et al ; Molecular Cell ; 2002) provenant d'une expérience conçue pour analyser les échanges au niveau du génome complet de la levure pour plusieurs inhibiteurs TBP (TATA binding protein) interagissant.

## Session 07E

Mardi 14 juin • Tuesday, June 14, 13:30 – 14:40

214

### Case Study 1: Antidepressant Medication and Stability in Seniors

#### Analyse de cas 1 : Sur les antidépresseurs et la stabilité chez les personnes âgées

[MS-110]

#### Presentation of the data set

#### Présentation du jeu de données

Amy PETERS, *Centre for Studies in Aging at Sunnybrook and Women's College Health Sciences Centre*

Falls and hip fracture are important sources of morbidity and mortality in the geriatric population. Several studies have identified antidepressant medications as a risk factor. In a recent study, healthy older adults were assigned to one of four treatment groups (one of three antidepressant medications or placebo) and were assessed at three time points for their static and dynamic balance.

Teams of students will present their analyses of the resulting data. Statistical issues that may be addressed include a small sample size (17 subjects), evidence of non-compliance, missing values, and the large number (more than 1500) of potential outcome variables.

Les chutes et les fractures des hanches sont des sources considérables de morbidité et de mortalité dans la population gériatrique. Plusieurs études ont identifié les médicaments antidépresseurs comme facteur de risque. Dans une étude récente, des adultes âgés en santé ont été assignés dans quatre groupes (trois groupes avec médicaments antidépresseurs et un groupe placebo) et ils ont été évalués à trois moments différents leur équilibre statique et dynamique.

Des équipes d'étudiants présenteront leur analyse des données résultantes. Certains des problèmes statistiques qui peuvent être discutés sont la petite taille de l'échantillon (17 sujets), certaines évidences de non conformité, les valeurs manquantes et le grand nombre de variables d'analyses potentielles (plus de 1500).

**Session 07F****Mardi 14 juin • Tuesday, June 14, 13:30 – 14:45****211**Applied survey methods  
Méthodes d'enquête appliquées**[MS-111]****Application of two-phase sampling in education survey in Tehran****Application de l'échantillonnage en deux phases dans le sondage sur l'éducation à Téhéran**Taraneh ABARIN, *University of Manitoba*

Two-phase sampling is based on sub sampling method. This method can be also used in mailing surveys to reduce non response error. Existence of bias in mailing surveys is unavoidable. Two-phase sampling can solve this problem. Firstly, questionnaires are mailed out to units. Based on response rate and unit costs in each phase, second phase sample size is obtained. These two samples totally give a reasonable estimator for desired attribute with a high precision and low cost. Like others, this method needs precise and updated framework. Other wise, response rate and precision of estimators are reduced. The mentioned method is applicable in many surveys, such as: education, manufacturing, mining, health and social survey, and etc. These are used in determining national and regional accounts. This paper presents application of two-phase sampling in "Education Survey" in Tehran. Results shows that two-phase has less variance for value added than ordinary stratification sampling. A very important point is that this method has a remarkable reduction in cost.

L'échantillonnage en deux phases est basé sur la méthode de sous échantillonnage. Cette technique peut également être utilisée dans des sondages utilisant une collecte par courrier postal pour réduire l'erreur due à la non réponse. L'existence de biais dans les sondages utilisant une collecte par courrier postal est inévitable et l'échantillonnage en deux phases peut résoudre ce problème. Premièrement, le questionnaire est posté aux unités. Les tailles des échantillons pour la deuxième phase sont obtenues en se basant sur le taux de réponse et le coût des unités associé à chacune des phases. Ces deux échantillons donne un estimateur raisonnable pour les attributs désirés avec une grande précision et des faibles coûts. Comme certaines autres méthodes, cette dernière nécessite une base de sondage à jour et précise sinon le taux de réponse et la précision de l'estimateur sont réduits. La méthode mentionnée ici est applicable dans plusieurs types de sondages couvrant soit l'éducation, la fabrication, l'exploitation minière, la santé, les sondages sociaux et bien d'autres. Ils sont utilisés pour déterminer les comptes nationaux et régionaux. Cet article présente l'application de l'échantillonnage en deux phases au sondage sur l'éducation à Téhéran. Les résultats montrent que l'échantillonnage en deux phases donne des estimateurs avec une variance plus faible que l'échantillonnage par stratification ordinaire. Un point important est que cette méthode présente une réduction importante des coûts.

**[MS-112]****Sampling issues in national HIV behavioral surveillance of injecting drug users****Problèmes d'échantillonnage dans le programme national de surveillance du VIH pour les utilisateurs de drogues injectives**MYRON KATZOFF, Lillian LIN & Michael MONSOUR, *National Center for Health Statistics/CDC*, Steven THOMPSON, *Simon Fraser University*

The Centers for Disease Control and Prevention (CDC) is about to begin National HIV Behavioral Surveillance (NBS) of the injecting drug user (IDU) population at 25 sites. Sampling of this hard-to-locate population can be facilitated through the

Le Centers for Disease Control and Prevention (CDC) est sur le point de lancer un programme de surveillance nationale du comportement du VIH (National HIV Behavioral Surveillance (NBS) ) pour la population des utilisateurs de drogues injectives (UDI) dans 25 sites. L'échantillonnage de

use of adaptive sampling procedures such as respondent driven sampling or snowball sampling. This paper will describe an adaptive web design as a modification to those designs which are of the strict random walk or snowball type. The adaptive web design has the advantage of enabling the direct use of the information that will be derived from the formative assessments to be conducted at the sites and selection of interview subjects based on the characteristics of seeds (or, to use a graph-theoretic term, originating nodes); it also will provide for effective constraints on sample size. The discussion of sampling issues will include consideration of: (1) ways to avoid getting “locked” into a single network of IDUs or a very small number of such strongly connected components of the IDU population; and (2) what provisions will be made to enable not only the estimation of HIV prevalence but also to perform other analyses of drug use practices and utilization of programs and services available to the IDU population.

cette population difficile à localiser peut être facilité par l'utilisation de procédures d'échantillonnage adaptatives telles que l'échantillonnage mené par le répondant ou l'échantillonnage boule de neige. Cet article décrit un plan adaptatif en ligne comme modification à ces plans d'échantillonnage qui sont de types strictement marche aléatoire ou boule de neige. Le plan d'échantillonnage adaptatif en ligne a l'avantage de permettre l'utilisation directe de l'information dérivée des évaluations formatrices qui seront menées aux sites et de sélectionner les sujets à interviewer basé sur les caractéristiques de provenance (ou pour utiliser un terme de la théorie des graphes, le nœuds de provenance). Il pourra également donner des contraintes efficaces sur la taille de l'échantillon. La discussion sur les problèmes d'échantillonnage va inclure des considérations sur : 1) les manières d'éviter de rester pris dans le même réseau d'UDI et 2) quelles mesures seront prises pour non seulement pouvoir estimer la prévalence du VIH mais également pouvoir effectuer d'autres analyses sur la pratique des drogues et l'utilisation des services et des programmes d'aides disponibles à la population des UDI.

#### [MS-113]

#### **Dual frame strategy in the survey of financial security: Do we need a high income sample?**

#### **L'utilisation de bases multiples dans l'Enquête sur la sécurité financière : Avons-nous besoin d'un échantillon de ménages à hauts revenus ?**

Richard LAROCHE & Barbara ARMSTRONG, *Statistique Canada*

The Survey of Financial Security (SFS) provides a picture of the net worth of Canadians. Run last in 1999, it is to be run again in 2005. To ensure adequate sample representation of the wealthiest families, the survey uses a dual frame approach. The area frame covers Canada's ten provinces. A list frame, made up of households in geographic areas in which there is a high concentration of households with high incomes, is employed to increase the sample take of high income households. In 1999 the SFS employed a sample of 23,000 households. The 2005 sample will be reduced substantially to 9,000 households.

The paper will present results from the 1999 survey. Based on these results, an evaluation will be made of the performance of the 1999 list sample. Impact on survey estimates and CVs will be discussed. Such an evaluation is required to determine whether, in the face of the sizeable reduction in sample size in 2005, use of a list frame sample is still warranted.

L'Enquête sur la sécurité financière (ESF) dresse un portrait de l'avoir net total des ménages canadiens. Réalisée pour la dernière fois en 1999, elle sera répétée de nouveau en 2005. Pour assurer une représentation adéquate des familles les plus riches, l'enquête utilise une approche à bases multiples. La base aréolaire couvre les dix provinces du Canada. Une base liste, composée de ménages vivant dans des régions géographiques ayant une forte concentration de ménages à hauts revenus, est également utilisée. En 1999, l'ESF avait une taille d'échantillon de 23 000 ménages. En 2005, l'échantillon ne compte que 9000 ménages.

La présentation fournira les résultats de l'enquête de 1999. Suite à ces résultats, une évaluation de la performance de l'échantillon de la base liste sera faite. L'impact sur les estimations et les coefficients de variation sera examiné. Une telle évaluation est nécessaire pour déterminer si, suite à la réduction importante de l'échantillon en 2005, l'utilisation d'une base liste de ménages à hauts revenus est toujours essentielle.

**[MS-114]****Improving the quality of the estimates of “special crops”****L'amélioration de la qualité des estimations des « cultures spéciales »**Laurie REEDMAN, *Statistics Canada*

The Crops Series of surveys collect information from Canadian crop producers six times each year. These data are used to estimate the acreage and yield of all major crops, such as wheat, corn, barley and oats. Also, estimates of many “special” crops are produced such as lentils, dry peas and beans, and canary seed. These “special” crops are only grown by a small percentage of Canadian producers, but they play a significant role in the agriculture industry. This paper explores the use of a composite estimator to improve the accuracy and precision of estimates of certain “special” crops while not increasing the regular sample size.

La série d'enquêtes sur les cultures recueille de l'information auprès des exploitants agricoles canadiens six fois par année. Ces données sont utilisées pour l'estimation de la superficie et du rendement de tous les grandes cultures, comme le blé, le maïs, l'orge et l'avoine. On produit également des estimations pour plusieurs « cultures spéciales » dont les lentilles, les pois secs, les haricots secs, et les graines de l'alpiste des canaries. Ce sont certes là des cultures menées par un petit pourcentage des exploitants agricoles canadiens, mais elles jouent néanmoins un rôle significatif dans l'industrie de l'agriculture. Dans cet article, on examine la possibilité de recourir à un estimateur composite pour améliorer l'exactitude et la précision des estimations des certaines « cultures spéciales » tout en conservant la même taille d'échantillon.

**[MS-115]****Weighting strategy of the Survey of Labour and Income Dynamics (SLID)****Pondération de l'Enquête sur la Dynamique du travail et du revenu (EDTR)**Caroline CAUCHON & Sylvie LAROCHE, *Statistique Canada*

L'Enquête sur la Dynamique du travail et du Revenu (EDTR) est une enquête longitudinale par panel menée auprès des individus. Elle est la principale source d'information sur le revenu des particuliers, des familles et des ménages, incluant l'incidence des familles à faible revenu. L'échantillon de l'EDTR est composé de deux panels qui se chevauchent. À chaque année de référence, l'EDTR produit deux ensembles de poids : un ensemble de poids représentatif de la population initiale (l'aspect longitudinal) et un autre qui est représentatif de la population actuelle (l'aspect transversal). Cette présentation décrira la méthodologie employée à l'EDTR pour pondérer les échantillons longitudinaux et transversaux.

**Session 07G****Mardi 14 juin • Tuesday, June 14, 13:30 – 15:00****217****Advanced Regression Methods  
Méthodes de regression avancées****[MS-116]****Use of auxiliary data in semiparametric regression with nonignorable missing responses  
Utilisation de données auxiliaires en régression semi paramétriques avec des données manquantes non ignorables**MARCO GERACI & MATTEO BOTTAL, *University of South Carolina*

We propose a method for reducing the error of the prediction of a quantity of interest, when the outcome has missing values that are suspected to be nonignorable and the data are correlated in space. We develop a maximum likelihood approach for the parameter estimation of semiparametric regressions in a mixed model framework. We apply the proposed method to phytoplankton data. A simulation study is also performed. The availability of a variable correlated to the response allows us to achieve a substantial reduction of the error of the prediction, without having to specify a nonignorable model.

Nous proposons une méthode pour réduire l'erreur de prévision d'une quantité d'intérêt lorsque la variable dépendante contient des valeurs manquantes que l'on suspecte d'être non ignorables et que les données sont corrélées dans l'espace. Nous développons une approche par le maximum de vraisemblance pour l'estimation des paramètres d'une régression semi paramétrique dans un cadre de modèles mixtes. Nous appliquons la méthode proposée à des données sur le phytoplancton. Une étude de simulation est également effectuée. La disponibilité d'une variable corrélée à la variable dépendante nous permet d'obtenir une réduction substantielle de l'erreur de prévision, sans avoir à spécifier un modèle de non ignorabilité.

**[MS-117]****Prediction of binary outcomes from longitudinal data  
Prévision d'un résultat binaire pour des données longitudinales**Julie HORROCKS, *University of Guelph*

In this talk I investigate some models for prediction of a (time-fixed) binary outcome using longitudinal information. The time-varying longitudinal covariates are often measured with error, which can lead to bias in the estimated coefficients. As an application, data on changes in the adhesion of certain blood lymphocytes in the days prior to ovulation are used to predict whether a woman will respond to in-vitro fertilization treatment for infertility. Adhesion is subject to substantial measurement error. The extent of bias due to measurement error in the longitudinal data is examined through simulation.

Dans cet exposé j'étudie certains modèles pour prédire un résultat binaire (fixe dans le temps) en utilisant de l'information longitudinale. Les covariables longitudinales, variables dans le temps, sont souvent mesurées avec erreur, ce qui peut mener à un biais dans l'estimation des coefficients. À titre d'application, nous utilisons des données sur les changements dans l'adhésion de certains lymphocytes sanguins dans les jours précédents l'ovulation pour prédire si la femme va répondre positivement à un traitement de fertilisation in vitro. L'adhésion est sujette à des erreurs de mesures substantielles. L'ampleur du biais dû à l'erreur de mesure dans les données longitudinales est examinée avec des simulations.

**[MS-118]****Time series models for the relationship of foodborne diseases and temperature****Modèles de séries chronologiques pour la relation entre la température et les maladies alimentaires**

Manon FLEURY, *Public Health Agency of Canada*, John HOLT, *University of Guelph*, Dominique CHARRON, *Public Health Agency of Canada*, Brian ALLEN, *University of Guelph*, Abdel MAAROUF, *Environment Canada*

This paper explores the relationship between temperature and enteric disease using generalized additive models and generalized linear models with regression splines. These methods are widely used for environmental time series analysis because they permit flexible adjustments for non-linear confounding effects of time trends, seasonality and temperature variables. There are, however, certain issues related to the application of these methods to time series data: there are no formal tests for the choice of smoothing parameters, and autocorrelation needs to be controlled in the model. Other methods explored include Fourier transformations, transitional regression models and threshold models.

Cet article explore la relation entre la température et les maladies entériques en utilisant des modèles additifs généralisés et des modèles linéaires généralisés avec des splines de régression. Ces méthodes sont largement utilisées pour l'analyse de séries chronologiques environnementales, puisqu'elles permettent un ajustement flexible pour des effets confondants non linéaires de la tendance temporelle, la saisonnalité et des variables sur la température. Il y a cependant certains problèmes liés à l'application de ces méthodes à des données de séries chronologiques : il n'y a pas de tests formels pour le choix du paramètre de lissage et les autocorrélations doivent être contrôlées dans le modèle. D'autres méthodes explorées incluent les transformations de Fourier, les modèles de régression transitionnels et les modèles à seuil.

**[MS-119]****The time-varying autoregressive model with covariates for analyzing longitudinal studies****Le modèle autorégressif à temps variable avec covariables pour analyser des études longitudinales**

Tulay KORU-SENGUL, *University of Saskatchewan*, David S. STOFFER, *University of Pittsburgh*

The time-varying autoregressive model with covariates can be thought of as a version of the transitional general linear models for analyzing longitudinal data. Variations of the model have been used previously in a great deal of applications except for longitudinal studies. In this content longitudinal outcome is modeled as a linear regression model for each time point by taking into account the correlation between the repeated measurements over time. The major advantages of the model are identifying the time points where the outcome changes, handling of unequally spaced data, and being able to use time-varying covariates. An epidemiological longitudinal data from Saskatchewan will be used to demonstrate the model. Research supported in part by the grant from National Science Foundation in the USA (DMS-0102511).

Le modèle autorégressif à temps variable avec covariables peut être vu comme une version du modèle linéaire général transitionnel pour analyser des données longitudinales. Des variations du modèle ont été utilisées dans le passé pour un grand nombre d'applications excepté pour des études longitudinales. À ce titre, la réponse longitudinale est modélisée comme un modèle de régression linéaire pour chaque période temps en tenant compte de la corrélation entre les répétitions de mesures dans le temps. Les avantages majeurs du modèle sont l'identification des temps où le résultat change, de tenir des données espacées inégalement et d'être capable d'utiliser des covariables qui changent dans le temps. Des données épidémiologiques longitudinales provenant de la Saskatchewan seront utilisées pour illustrer le modèle. Cette recherche est supportée en partie par la bourse du National Science Foundation aux Etats-Unis (DMS-0102511).

**[MS-120]****Reparameterization of the boundary of the admissible region for ARMA models****Reparamétrisation de la frontière de la région d'admissibilité pour des modèles ARMA**Ying ZHANG, *Acadia University*, Ian McLEOD, *University of Western Ontario*

The theorem of Barndorff-Neilsen and Schou (1973) is extended to the boundary of the admissible region of autoregressive and moving average models as well as the subset autoregressive models and it is noted that the rigorous proof of this extension is non-obvious. A new maximum likelihood algorithm for the moving average model is described and its use illustrated in a simulation example. It is shown how the result can be used to construct new diagrams of the admissible regions for autoregressions and subset autoregressions. In the subset  $AR(p)$  model, it is shown that the admissible parameter space is spanned by the corresponding partial autocorrelations.

Le théorème de Barndorff-Neilsen et Schou (1973) est étendu à la frontière de la région admissible des modèles autorégressifs moyennes mobiles ainsi qu'aux modèles autorégressifs emboîtés. Il est utile de remarquer qu'une preuve rigoureuse de cette extension n'est pas évidente. Un nouvel algorithme du maximum de vraisemblance pour le modèle moyenne mobile est décrit, et son utilisation est illustrée dans une étude de simulations. Nous montrons comment les résultats peuvent être utilisés pour construire de nouveaux diagrammes des régions admissibles pour l'autorégression et l'autorégression emboîtée. Dans le modèle  $AR(p)$  emboîtée, nous montrons que l'espace des paramètres admissibles est engendré par les autocorrélations partielles correspondantes.

**[MS-121]****An alternative algorithm to the accelerated failure time model****Un algorithme alternatif pour le modèle de temps de vie accéléré**Jiajia ZHANG & Yingwei PENG, *Memorial University of Newfoundland*

Frailty models based on the proportional hazards model are useful to analyse correlated or clustered failure time data. However, there are some recent interests in a frailty model based on the famous accelerated failure model for correlated failure time data. In this paper we review recent advances on this model in the literature. We propose a general estimation method based on M-estimators and the EM algorithm to estimate the parameters in the model. To evaluate the performance of the proposed method, we carry out a simulation study. The results of the simulation study show that the proposed method performs well in comparison with existing estimation methods. As an illustration, we also apply the model and the proposed method to a dataset of failure times from kidney patients.

Les modèles de fragilité basés sur le modèle de panne proportionnel sont utiles pour analyser des données de temps de défaillance corrélées ou groupées. Cependant, il y a certains intérêts récents dans un modèle de fragilité basé sur le modèle de défaillance accélérée pour des données de temps de défaillance corrélées. Dans cet article, nous passons en revue la littérature sur les avancements récents de ce modèle. Nous proposons une méthode d'estimation générale basée sur les M-estimateurs et l'algorithme EM pour les paramètres du modèle. Pour évaluer la performance de la méthodologie proposée, nous effectuons une étude de simulation. Les résultats de l'étude de simulation montrent que la méthode proposée performe bien comparativement à d'autres méthodes d'estimation existantes. À titre d'illustration, nous appliquons le modèle et la méthodologie proposée à un jeu de données de temps de défaillances provenant de patients avec des problèmes de reins.



<b>Session 09A</b>	<b>Mercredi 15 juin • Wednesday, June 15, 09:00 – 09:45</b>	<b>241</b>
Pierre Robillard Award Address		
Allocution du prix Pierre Robillard		

**Session 09B      Mercredi 15 juin • Wednesday, June 15, 09:45 – 10:30      241**

Canadian Journal of Statistics Award Address

Allocution du prix de La Revue canadienne de statistique

**Session 10A      Mercredi 15 juin • Wednesday, June 15, 11:00 – 12:30      214**

## Imputation

## Imputation

[MS-122]

**COM server for multiple imputation****Serveur COM pour l'imputation multiple**Joseph SCHAFER, *Pennsylvania State University*

Multiple imputation (MI) is a simulation-based technique for handling missing data which replaces each unknown datum by a set of  $m > 1$  simulated values drawn from a predictive distribution. Algorithms for generating MI's in multivariate settings typically involve Markov chain Monte Carlo. This presentation will introduce and demonstrate new routines for MI that have been packaged as a server compliant with the Microsoft Component Object Model (COM). A COM server is an object-oriented software module which, once it is installed and registered with the Windows operating system, can interact with a wide variety of applications (S-PLUS, R, Excel, SAS, SPSS and MATLAB, to name a few), producing consistent results across these platforms.

L'imputation multiple (IM) est une technique basée sur des simulations pour tenir compte des données manquantes qui remplace chaque référence inconnue par un ensemble de  $m > 1$  valeurs simulées et tirées d'une distribution prédictive. Les algorithmes pour générer des IM dans un cadre multivarié impliquent typiquement les techniques de chaînes de Markov Monte Carlo. Cette présentation introduit et démontre de nouvelles routines pour l'IM qui sont fournies avec le Microsoft Component Object Model (COM) conforme aux serveurs. Un serveur COM est un module de programme orienté objet qui, une fois installé et enregistré avec le système d'opération Windows, peut interagir avec une grande variété d'applications (S-PLUS, R, Excel, SAS, SPSS et MATLAB pour en nommer quelques uns), produisant des résultats consistants à travers ces plates-formes.

[MS-123]

**Vector imputation at Westat****Imputation de vecteurs chez Westat**Thomas KRENZKE, David JUDKINS & Zizhong FAN, *Westat*

Westat uses primarily hot-deck and nearest-neighbor procedures for imputation. In-house software was standardized a decade ago that would preserve marginal distributions of variables. Recently, we have developed new software that makes it easy for users to preserve marginal distributions, skip patterns, and covariances among ordinal and continuous items. We report on the algorithm, which uses existing procedures, while incorporating greedy parametric modeling of relationships. A simulation study will include a feature that contrasts it with Bayesian procedures, in that it can preserve semi-continuous distributions, such as household income, that are generally continuous but display heaping at round numbers.

Westat utilise principalement les procédures hot-deck et du plus proche voisin pour faire de l'imputation. Un logiciel maison qui préserve les distributions marginales des variables à été standardisé il y a une dizaine d'années. Récemment, nous avons développé de nouveaux logiciels qui simplifient la tâche de l'utilisateur pour préserver les distributions marginales, omettre des patrons et des covariances entre des items ordinaux et continus. Nous faisons un rapport sur l'algorithme qui utilise des procédures existantes et incorpore une modélisation paramétrique des relations. Une étude de simulations va inclure une caractéristique qui démarque l'algorithme par rapport aux procédures bayésiennes, dans le sens où il peut préserver les distributions semi-continues comme le revenu familial, qui est généralement continu mais qui se concentre aux valeurs arrondies.

**[MS-124]****The Banff system for automated editing and imputation  
Le système Banff pour l'éditage et l'imputation automatique**Robert KOZAK, *Statistics Canada*

Banff is a generalized system recently developed at Statistics Canada for the automated editing and imputation of quantitative survey data. It evolved from the Generalized Edit and Imputation System (GEIS), which has been used at Statistics Canada since the late 1980s. The system is a collection of nine independent, specialized SAS procedures that provide functionality similar to GEIS. However, Banff is much more flexible with respect to the operating environment and ease-of-use. This paper briefly reviews the initial development of Banff, and describes each of the system functions in detail. An overview of the methodology behind the functions is presented. Finally, there is a discussion about the current and future development of the system which is targeted at making Banff even more versatile, including the development of new methodologies and a graphical interface.

Banff est un système généralisé développé récemment à Statistique Canada pour la vérification et l'imputation automatique de données de sondages quantitatives. Il représente une évolution par rapport au Système généralisé de vérification et d'imputation (SGVI) qui a été utilisé à Statistique Canada depuis la fin des années 80. Le système est une collection de neuf procédures SAS spécifiques et indépendantes qui permettent des fonctionnalités similaires à SGVI. Cependant, Banff est beaucoup plus flexible par rapport à l'environnement d'opération et est aussi plus convivial. Cet article passe brièvement en revue le développement initial de Banff et décrit en détail les fonctionnalités du système. Un survol de la méthodologie associée à chacune des fonctions est présentée. Finalement, il y a une discussion sur les développements présents et futurs du système qui doivent rendre Banff encore plus versatile, incluant le développement de nouvelles méthodologies et d'une interface graphique.

**Session 10B      Mercredi 15 juin • Wednesday, June 15, 11:00 – 12:30      200****Recent Advances in Lifetime Data  
Progrès récents pour les données de durée****[MS-125]****Inference methods based on the EM algorithm for the competing risk model with masked causes of failure  
Méthodes d'inférence basées sur l'algorithme EM pour le modèle des risques concurrents avec causes de pannes masquées**Thierry DUCHESNE, *Université Laval*, Radu CRAIU, *University of Toronto*, Isabelle MICHAUD, *Université Laval*

In some applications of the competing risk model, only partial information about the causes of death is available; when a cause of death is only known to be in a subset of all possible causes of death, we say that the cause of death is masked.

In this talk, we consider the methods based on the EM algorithm of Craiu & Duchesne (2004). In addition to the published results, we present findings from extensive simulation studies on the robustness of the methods as well as an investigation of the behavior of the methods with no second stage data.

Dans certaines applications du modèle des risques concurrents, seule une information partielle sur les causes de décès est disponible. Lorsque tout ce que nous savons est que le décès est dû à une des causes d'un sous-ensemble des causes possibles, nous disons que la cause de décès est masquée.

Dans cet exposé, nous considérons les méthodes basées sur l'algorithme EM proposées par Craiu et Duchesne (2004). En plus des résultats publiés, nous présentons les conclusions d'études par simulation exhaustives sur la robustesse des méthodes et sur le comportement de ces méthodes en l'absence de données de deuxième étape.

**[MS-126]****A sequential stratification method to estimate the effect of a time-dependent treatment in the analysis of recurrent event data****Une méthode de stratification séquentielle pour estimer l'effet d'un traitement temporellement dépendant pour l'analyse de données d'événements récurrents**

Douglas SCHAUBEL, *University of Michigan*

In biomedical studies, the event of interest is often recurrent. We propose a semiparametric method for estimating the effect on the conditional event rate of a time-dependent treatment. When applied to observational studies, the proposed method is intended to yield a treatment effect estimator with more of a causal interpretation than that resulting from the typical time-dependent analysis. Asymptotic properties of the regression parameter estimator are derived and evaluated in finite samples through simulation. The proposed methods are applied to quantify the morbidity benefit of transplantation using observational data from a national organ failure registry.

Dans les études biomédicales, l'événement d'intérêt est souvent récurrent. Nous proposons une méthode semi paramétrique pour estimer l'effet d'un traitement temporellement dépendant sur le taux conditionnel de l'événement. Lorsqu'elle est appliquée à des études observationnelles, la méthode proposée est supposé donner un estimateur de l'effet du traitement avec une meilleur interprétation causale que ceux résultants des analyses typique de dépendance temporelle. Les propriétés asymptotiques de l'estimateur du paramètre de régression sont dérivées et évaluées pour des échantillons finis à partir de simulations. Les méthodes proposées sont appliquées pour quantifier les avantages d'une transplantation sur la morbidité en utilisant des données d'observations d'un registre national sur les problèmes d'organes.

**[MS-127]****The analysis of multivariate point processes with partially missing event types****L'analyse de processus ponctuels multivariés avec des types d'événements partiellement manquants**

Bingshu CHEN, *NCI/NIH*, Richard COOK, *University of Waterloo*

In many clinical studies, subjects are at risk of experiencing more than one type of recurrent event. Sometimes however, only the occurrence of an event is observed and the specific type is not determined. We develop methods for analysing this sort of incomplete multi-type recurrent event data. A Monte-Carlo EM algorithm is proposed to handle the missing data problem in the context of likelihood analyses. Robust arginal methods based on estimating functions are also developed. Both approaches are studied via simulation.

Dans plusieurs études cliniques, les individus sont sujets à expérimenter plus d'un type d'événements récurrents. Cependant, des fois seulement l'occurrence d'un événement est observée et le type spécifique n'est pas déterminé. Nous développons une méthode pour analyser ce type de données d'événements récurrents incomplets multi types. Un algorithme EM Monte-Carlo est proposé pour résoudre le problème de données manquantes dans le contexte d'analyse de vraisemblance. Des méthodes marginales robustes basées sur des fonctions d'estimation sont également développées. Les deux approches sont analysées à l'aide de simulations.

## Session 10C      Mercredi 15 juin • Wednesday, June 15, 11:00 – 12:30      133

### Estimation in constrained parameter spaces

### Estimation dans des espaces de paramètres avec contraintes

#### [MS-128]

#### **A unified approach for minimax estimation of a bounded parameter**

#### **Approche unifiée pour l'estimation minimax d'un paramètre borné**

Wolfgang BISCHOFF, *Catholic University Eichstaett-Ingolstadt*

In practice the unknown parameter of an experiment is often bounded. Therefore there is practical interest to include such additional knowledge into statistical procedures. For a sufficiently small parameter interval we prove under certain assumptions that a minimax estimator exists which is Bayes with respect to a two-point prior with mass at the boundaries when the loss is strictly convex. This unifies the results of a sequence of papers. If the loss is not strictly convex, however, we show that the above mentioned result is not true.

En pratique, le paramètre inconnu dans une expérience a souvent un domaine borné. Ainsi en pratique il est intéressant d'inclure une telle information dans les procédures statistiques. Pour un domaine relativement petit, nous prouvons sous certaines hypothèses qu'un estimateur minimax existe, et ce dernier est un estimateur de Bayes par rapport à une densité a priori bipolaire qui porte sa masse aux deux bornes, lorsque la fonction de perte est strictement convexe. Cependant, si la fonction de perte n'est pas strictement convexe, nous montrons que le résultat mentionné ci-dessus ne tient plus.

#### [MS-129]

#### **Shape restricted estimation in the search for dark matter**

#### **Estimation restreinte sur la forme dans la recherche de la matière noire**

Michael WOODROOFE, *University of Michigan*

In addition to visible matter, like stars, there is matter that astronomers cannot see. They know that it is there from gravitational effects. Its nature and even location are not well known at present, but there is a wide belief that the vast majority of matter in the Universe is dark. The talk centers on statistical issues that arise in the non-parametric estimation of dark matter in the dwarf spheroidal galaxies that are satellites of the Milky Way. Statistically, this is an inverse problem with missing data. The distribution of (total) mass together with the law of gravity determines the positions and velocities of stars, and the statistical problem is to recover the distribution of mass from observed positions and velocities. Complete positions and velocities cannot be observed here—only the radial velocity and the projection of position on the plane orthogonal to the line of sight. The distribution of projected position can be estimated quite precisely from large (~ 150,000 stars) photometric studies. Velocities require some time to measure, however, and previous studies have been too small

En plus de la matière visible comme les étoiles, il existe de la matière que les astronomes ne peuvent pas voir. Cependant, ils savent qu'elle existe à partir des effets gravitationnels. Sa nature et la localisation ne sont pas bien connues présentement, mais il y a une grande croyance que la grande majorité de la matière dans l'univers est noire. La présentation met un emphase sur les problèmes statistiques qui surviennent dans l'estimation non paramétrique de la matière noire dans les petites galaxies sphéroïdales satellites à la Voie Lactée. Statistiquement, ceci est un problème inverse avec données manquantes. La distribution de la masse totale avec la loi de la gravité détermine la position et la vitesse des étoiles. Le problème statistique est de retrouver la distribution et la masse à partir des positions et des vitesses. Les positions et vitesses complètes ne peuvent pas être observées ici, seulement la vitesse radiale et la projection de la position sur la plan orthogonal de la ligne d'horizon. La distribution de la position projetée peut être estimée assez précisément par d'importantes études photométriques (~ 150000 étoiles). Les vitesses requièrent du temps pour être mesurées et les études antérieures étaient trop petites (~ 30 étoiles) pour permettre une estimation en toute généralité.

( $\sim 30$  stars) to allow estimation in any generality. Instead, astronomers have been forced to adopt highly structured parametric models. New instrumentation will soon make much larger samples ( $\sim 1500$  stars) available and non-parametric methods feasible. Assuming spherical symmetry, let  $M(r)$  denote the total mass within  $r$  of the center of a dwarf spheroidal galaxy. Then Jeans' Equation (from classical potential theory) shows that  $M(r) = c\Psi''(r)/f(r)$ , where  $c$  is a known constant,  $f$  is a (nearly) known non-negative function, and  $\Psi$  is a function for which an unbiased,  $\sqrt{n}$ -consistent estimator exists. From Jeans' Equation,  $\Psi$  is a convex function, and  $\Psi''(r)/f(r)$  is non-decreasing in  $r$ . The unbiased estimator,  $\Psi^\#$  say, does not satisfy these shape restrictions and can be improved by imposing them. The talk centers on methods for accomplishing this, and properties of the resulting estimators. This is joint work with Xiao Wang (UM Statistics) and Mario Mateo and Matt Walker (UM Astronomy).

#### [MS-130]

**Estimation of the proportion  $p$  for a Binomial( $n, p$ ) distribution when  $p$  is near  $1/2$**

**Estimation d'une proportion  $p$  pour une loi Binomiale( $n, p$ ) lorsque  $p$  est près de  $1/2$**

Francois PERRON, *Université de Montréal*

We consider the problem of estimating the parameter  $p$  for a Binomial( $n, p$ ) distribution when  $p$  lies in the symmetric interval  $[(1 - m/\sqrt{n})/2, (1 + m/\sqrt{n})/2]$ ,  $m < \sqrt{n}$ . We are interested on improving the MLE and finding minimax estimators under the quadratic loss. We will study some of the estimators that have been proposed by other authors. We will use a bayesian approach as well. We will give simple sufficient conditions for the problem of dominating the MLE. We will show that when  $m$  is small any Bayes estimator with respect to a symmetric prior dominates the MLE. In particular, we will use the prior which put all of the mass to the boundary of the parameter space. This prior being the least favourable prior when  $m$  is small. We will give asymptotic results.

De nouveaux instruments vont bientôt permettre de plus gros échantillons ( $\sim 1500$  étoiles) et ainsi rendent possibles les méthodes non paramétriques. Supposons une symétrie sphérique, soit  $M(r)$  la masse totale autour d'un rayon  $r$  du centre de la petite galaxie. L'équation de Jean (de la théorie potentielle classique) montre que  $M(r) = c\Psi''(r)/f(r)$ , où  $c$  est une constante connue,  $f$  est une fonction non négative quasiment connue et  $\Psi$  est une fonction pour laquelle un estimateur sans biais  $\sqrt{n}$ -consistant existe. Par l'équation de Jean,  $\Psi$  est une fonction convexe et  $\Psi''(r)/f(r)$  est non décroissant en  $r$ . L'estimateur sans biais, disons  $\Psi^\#$ , ne satisfait pas ces restrictions sur la forme et peut être amélioré en imposant ces restrictions. Cette présentation met l'emphase sur des méthodes pour accomplir cela et sur les propriétés des estimateurs résultants. Ceci est un travail conjoint avec Xiao Wang (UM Statistics), Mario Mateo and Matt Walker (UM Astronomy).

Nous considérons le problème d'estimer  $p$  pour une Binomiale( $n, p$ ) lorsque  $p$  se trouve dans l'intervalle symétrique  $[(1 - m/\sqrt{n})/2, (1 + m/\sqrt{n})/2]$ ,  $m < \sqrt{n}$ . On s'intéresse à améliorer l'EQM et à trouver des estimateurs minimax pour une perte quadratique. Nous allons étudier quelques estimateurs proposés par d'autres chercheurs. On utilisera aussi une approche bayésienne. Nous proposons des conditions simples pour s'assurer qu'un nouvel estimateur domine l'EQM. Nous démontrons qu'il suffit que  $m$  soit assez petit pour qu'un estimateur de Bayes domine l'EQM lorsque la loi a priori est symétrique. En particulier, nous allons nous intéresser à la loi a priori concentrée sur la frontière de l'espace paramétrique et montrer qu'elle devient la loi a priori la moins favorable lorsque  $m$  est petit. Finalement, nous allons donner des résultats asymptotiques.

**Session 10D      Mercredi 15 juin • Wednesday, June 15, 11:00 – 12:30      217****Graphical Gaussian models  
Modèles gaussiens graphiques****[MS-131]****Binary models for marginal independence  
Modèles binaires pour l'indépendance marginale**Thomas RICHARDSON, *University of Washington*, Mathias DRTON, *University of California Berkeley*

Marginal independence hypotheses cannot be accommodated in general by traditional log-linear models. For example, it is not possible to formulate a model for four variables  $(A, B, X, Y)$  such that  $A$  is independent of  $B$ , and  $X$  is independent of  $Y$  (and no other restrictions are imposed). Focusing on binary variables I will present a new model class that provides a framework for addressing this problem. The approach is graphical and based on bi-directed graphs, which are in the tradition of path diagrams. In many respects the resulting models and associated fitting algorithms are dual to graphical log-linear models.

Les hypothèses d'indépendance marginale ne peuvent généralement pas être traitées par des modèles log linéaires. Par exemple, il n'est pas possible de formuler un modèle pour quatre variables  $(A, B, X, Y)$  tel que  $A$  soit indépendant de  $B$  et  $X$  soit indépendant de  $Y$  (sans aucune autre restriction). En mettant l'accent sur les variables binaires, je présente une nouvelle classe de modèles qui fournit un cadre pour adresser ce problème. L'approche est graphique et basée sur des graphes bi-dirigés qui entrent dans la tradition des diagrammes de trajectoires. À plusieurs égards, les modèles résultants et les algorithmes d'ajustement associés présentent une certaine dualité par rapport aux modèles log linéaires graphiques.

**[MS-132]****A glance at graphical modelling in statistics and artificial intelligence  
Un bref regard sur la modélisation graphique en statistique et en intelligence artificielle**Milan STUDENY, *Academy of Sciences of the Czech Republic*

The goal of this elementary talk will be to inform general statistical audience briefly and superficially about graphical modelling of conditional independence structures by means of so-called chain graphs. This will be done by presenting three artificial toy examples whose aim is to motivate and illustrate the possible use of graphical methods in the following three areas: in contingency tables processing, in multivariate statistical analysis and in probabilistic decision making. The relevant mathematical definitions will be given throughout the talk.

Le but de cette présentation élémentaire est d'informer brièvement et superficiellement le public statistique sur la modélisation graphique de structures d'indépendance conditionnelle par le biais des graphes en chaîne. Nous présenterons trois exemples jouets artificiels dont le but est de motiver et illustrer l'utilisation possible des méthodes graphiques dans les trois domaines suivant : le traitement des tableaux de contingence, l'analyse statistique multivariée et la prise de décisions probabiliste. Les définitions mathématiques utiles sont fournies tout au long de l'exposé.

**[MS-133]****A new family of conjugate priors for decomposable graphical Gaussian models  
Une nouvelle famille de densités a priori conjuguées pour des modèles gaussiens graphiques décomposables**Hélène MASSAM, *York University*, Gérard LETAC, *Université Paul Sabatier*

We construct a new family of conjugate prior distributions for the covariance parameter of graphical Gaussian models Markov with respect to a decom-

Nous construisons une nouvelle famille de densités a priori conjuguées pour le paramètre de covariance de modèles graphiques gaussiens markoviens par rapport à un graphe



posable graph. The Diaconis-Ylvisaker conjugate prior which is the hyper inverse Wishart as given by Dawid and Lauritzen (1993) is a particular case of our new family. Our family has the advantage of flexibility since its shape parameter is of dimension at least equal to the number of cliques plus one. Also, like the hyper inverse Wishart, it is strong hyper Markov thus allowing inference to be made locally, clique by clique, in a Bayesian framework. This property is most important for the analysis of complex high-dimensional data.

décomposable. La densité a priori conjuguée de Diaconis-Ylvisaker qui est l'hyper inverse Wishart donnée par Dawid and Lauritzen (1993) est un cas particulier de notre nouvelle famille. Notre famille a l'avantage d'être flexible puisque son paramètre de forme est de dimension au moins égale au nombre de cliques plus un. De plus, comme l'hyper inverse Wishart, elle est hyper-markovienne forte, permettant ainsi d'inférer localement clique par clique, dans un cadre bayésien. Cette propriété est très importante pour l'analyse de données complexes à haute dimension.

## **Session 10E      Mercredi 15 juin • Wednesday, June 15, 11:00 – 12:30      241**

Teaching statistics in the health and life sciences

L'enseignement de la statistique dans les sciences de la santé et de la vie

[MS-134]

**Comparing apples to oranges - issues in confounding**

**Comparer des pommes et des oranges : problème d'effets confondants**

Paul COREY, *University of Toronto*

Harmful effects of occupational exposures or benefits from drug therapy are measured by comparing the effects in two groups, one exposed to the pollutant or drug and the other not exposed. The ideal involves groups identical in all respects except the exposure. The real world compromise involves two groups similar, on average, using randomization in the drug trial and statistical adjustment in the occupational study. Some believe that statistical adjustment using ANACOVA is never appropriate in randomized studies, others that it is only appropriate in randomized studies. Examples appropriate for a class of graduate health scientists will be presented.

Les effets néfastes d'une exposition à un polluant au travail ou les bénéfices d'un traitement par un certain médicament sont mesurés en comparant les effets dans deux groupes, un premier exposé au polluant ou au médicament et l'autre non exposé. Idéalement nous voulons avoir des groupes identiques à tout égards à l'exception de l'exposition. Dans la vraie vie, le compromis est d'avoir des groupes similaires en moyennes, en utilisant la randomisation pour l'essai clinique sur le médicament et l'ajustement statistique pour l'étude sur l'exposition au travail. Certains croient que l'ajustement statistique par l'ANACOVA n'est jamais approprié pour les études randomisées, et d'autres croient que cela l'est seulement pour les études randomisées. Des exemples appropriés pour un groupe de scientifiques gradués en santé seront présentés.

[MS-135]

**Teaching statistics in the presence of heteroscedasticity**

**Enseigner les statistiques en présence d'hétéroscédasticité**

Penny BRASHER, *University of Calgary*

The students enrolled in the graduate-level course, Biostatistics I, at the UofC are required to have an introductory statistics course prior to enrolment. However, this may be as recent as "last term" or "twenty years ago". In addition, the students come from different programs of study and diverse back-

Les étudiants dans le cour gradué biostatistique 1 à l'UofC doivent avoir comme préalable un cour d'introduction à la statistique. Cependant, ils peuvent avoir suivi le cours la session précédente ou bien 20 ans auparavant. De plus, les étudiants viennent de plusieurs programmes d'études différents. Cet exposé reflète les défis à motiver un audience aussi

grounds. This talk will reflect on the challenge of engaging such a mixed audience, discuss things that worked and things that didn't, and the rewards at the end of the day.

mixte, discute des choses qui fonctionnent et de celles qui ne fonctionnent pas et finalement de la récompense que l'on a à la fin d'une journée.

**[MS-136]**

**An online system for teaching the design (and analysis) of experiments  
Un système en ligne pour enseigner la planification et l'analyse d'expériences**

Carl SCHWARZ, *Simon Fraser University*

Most books on experimental design have many exercises at the end of chapters that give students practice in the *ANALYSIS* of completed experiments, but students often receive little experience in the *DESIGN* of experiments. This paper presents a virtual construction machine that generates a virtual laboratory. This is web-accessible, easily used by students, and gives different responses for every student. A key advantage of this approach is that students can quickly try many different designs on the same experiment and learn to distinguish among designs based on randomization choices.

La majorité des manuels sur la planification d'expériences contiennent plusieurs exercices à la fin des chapitres qui permettent à l'étudiant de pratiquer l'analyse d'expériences déjà complétées, mais l'étudiant a rarement de pratique dans la planification des expériences. Cet article présente un environnement de construction virtuel qui génère un laboratoire virtuel. Ce dernier est accessible via Internet, facilement utilisable par l'étudiant et donne des réponses différentes pour chacun des élèves. Un avantage important de cette approche est que les étudiants peuvent rapidement essayer plusieurs plans d'analyses différents pour la même expérience et apprendre à distinguer les plans basés sur des choix de randomisation.

**Session 10F      Mercredi 15 juin • Wednesday, June 15, 11:00 – 12:30      212**

**Model Selection and Robust Modelling  
Sélection des modèles et la modélisation robuste**

**[MS-137]**

**Analysis of Longitudinal Bivariate Binary Data Under a Mixed Effects Model  
Analyse de données longitudinales binaires bivariées pour un modèle à effet mixte**

J. N. K. RAO, *Carleton University*, Kalyan DAS, *Calcutta University, India* & Qunshu REN, *Carleton University*

Ten Have and Morabia (1999) proposed a mixed effects model for bivariate binary data that consists of two marginal logistic linear mixed models and a log-odds ratio linear mixed model for the bivariate association. Under normality of random effects in the component models and using a binomial approximation, they obtained MLE of model parameters and associated standard errors. We propose a general MCMCEM method for arbitrary distributions of random effects. In particular, we use Huber's least favourable distribution to obtain robust estimators and associated standard errors. Empirical results are presented to demonstrate the advantage of robust estimators.

Ten Have et Morabia (1999) ont proposé un modèle à effets mixtes pour des données binaires bivariées qui consiste en deux modèles linéaires mixtes à marges logistiques et un modèle linéaire mixte du rapport log-cote pour la dépendance bivariée. Sous la normalité de l'effet aléatoire dans les modèles de composantes et en utilisant une approximation binomiale, ils obtiennent les EVM des paramètres du modèle et les écarts-types associés. Nous proposons une méthode MCMCEM générale pour des distributions arbitraires de l'effet aléatoire. En particulier, nous utilisons la distribution la moins favorable de Huber pour obtenir des estimateurs robustes et leur variance. Des résultats empiriques sont présentés pour démontrer les avantages des estimateurs robustes.

**[MS-138]****Maximum likelihood estimation and inference in non-normal random effects models****Estimation du maximum de vraisemblance et inférence dans des modèles à effet aléatoire non normal**Peng ZHANG & Peter SONG, *University of Waterloo*, Annie QU, *Oregon State University*

We propose maximum likelihood estimation and inference in linear mixed models when random effects might be non-normal. We assume that the distribution of random effects follows a family of parametric distributions which include the normal as a special case. This approach provides an extension of a typical normal random effects model, and also allows heavy tail  $t$ -distributed or positively skewed random effects. We develop a general fixed point algorithm to obtain the maximum likelihood estimation for both fixed effects and variance component parameters. We also propose a lack-of-fit test to assess the departure of the normal random effects from a given non-normal random effects distribution. To illustrate the proposed methods, we investigate two distribution families for random effects, the multivariate  $t$ -distribution and the semi-nonparametric (SNP) distribution families. We provide both simulation studies and data analysis examples, and compare our approach to the existing EM-algorithm.

Nous proposons une estimation du maximum de vraisemblance et de l'inférence dans des modèles linéaires mixtes quand l'effet aléatoire peut ne pas être normal. Nous supposons que la distribution des effets aléatoires suit une famille de distributions paramétriques qui admet la loi normale comme cas spécial. Cette approche donne une extension du modèle à effet aléatoire normal habituel et permet également des effets aléatoires à queues épaisses de loi  $t$  ou positivement asymétriques. Nous développons un algorithme ponctuel fixe général pour obtenir l'estimateur du maximum de vraisemblance pour les paramètres d'effets fixes et des composantes de la variance. Nous proposons également un test d'ajustement pour évaluer la distorsion par rapport à l'effet aléatoire normal d'une distribution non normale d'effet aléatoire. Pour illustrer les méthodes proposées, nous analysons deux familles de distribution pour l'effet aléatoire, la distribution  $t$  multivariée et les familles de distributions semi non-paramétriques. Nous montrons des résultats d'études de simulations et des exemples d'analyse de données, et comparons notre approche à l'algorithme EM existant.

**[MS-139]****The Analysis of Multivariate Repeated Measures Data with Structured Covariances****Analyse de données de mesures répétées multivariées avec covariance structurée**Lisa LIX, *University of Manitoba*

Multivariate repeated measures (RM) data, which arise when responses are collected across multiple occasions for two or more correlated outcomes, are common in studies of health-related quality of life. Choosing the correct covariance structure is critical for obtaining unbiased estimates of RM effects in a joint general linear mixed model. Likelihood ratio tests of Kronecker product and spherical covariance structures are highly sensitive to sample size and non-normal distributions. Simulations show that the Aikake and Schwarz information criteria frequently select the wrong covariance structure. The effect of covariance misspecification on tests of RM effects is also examined.

Les données de mesures répétées (MR) multivariées surviennent lorsque les réponses sont recueillies à plusieurs occasions pour deux résultats corrélés ou plus. Elles sont fréquentes dans les études sur la qualité de vie reliées à la santé. Le choix de la bonne structure de covariance est critique pour obtenir des estimations sans biais de l'effet des MR dans un modèle général linéaire mixte conjoint. Les tests de rapport de vraisemblance du produit de Kronecker et les structures de covariances sphériques sont très sensibles à la taille de l'échantillon et à la non normalité. Les simulations montrent que les critères d'information de Akaike et Schwarz sélectionnent fréquemment la mauvaise structure de covariance. Nous examinons également l'effet de la mauvaise spécification de la covariance sur les tests de l'effet de MR.

**[MS-140]****Clustering countries according to socioeconomic and behavioural measures of disease risk  
Classification de pays selon des mesures socio-économiques et comportementales du risque de maladies**Melanie BUITENDYK, Nicole FERKO, Michele KOHLI & Nancy RYAN, *Innovus Research Inc.*

Health-economic models can be used to predict disease outcomes. Country-specific analyses are often required, which are not always feasible because of modeling time restrictions. We will discuss how we applied principal components analysis followed by cluster analysis (CA) to reduce the number of measures and group countries. Thirteen socio-economic and behavioural measures were reduced using a combination of factoring criteria and manageable number of factors to include in the CA. Two clustering methods were performed on 170 countries and the most appropriate chosen. The optimal cluster number was determined using consensus amongst 4 statistical criteria.

Des modèles sur la santé et l'économie peuvent être utilisés pour prédire le résultat d'une maladie. Des analyses spécifiques aux pays sont souvent requises, cependant pas toujours possibles à cause des limites de temps pour effectuer les analyses. Nous discutons comment nous appliquons l'analyse en composantes principales suivie d'une analyse de clusters pour réduire le nombre de mesures et le nombre de groupes de pays. Treize mesures socio-économiques et comportementales ont été réduites en utilisant une combinaison de facteurs de factorisations et un nombre gérable de facteurs à inclure dans l'analyse de clusters. Deux méthodes de clusters ont été appliquées sur 170 pays et la plus appropriée fut sélectionnée. Le nombre optimal de groupe à été déterminé selon un accord entre quatre critères statistiques.

**[MS-141]****Model selection for hierarchical linear regression models for longitudinal data  
Sélection de modèles de régression linéaire hiérarchique pour données longitudinales**Geneviève LEFEBVRE, Alain C. VANDAL & Russell STEELE, *McGill University*

Model selection in a Bayesian framework is not straightforward, particularly when dealing with hierarchical models. Among the most common tools to compare models are the Bayesian Information Criterion (BIC), the Marginal Likelihood (ML) via posterior model probabilities, and the Deviance Information Criterion (DIC). The aim of the project was to do model selection on five hierarchical linear regression models fitted on longitudinal neurological data. Regressions were performed using Gibbs sampling, and for each model the effect of two or three covariates was allowed to vary between subjects. The implementation and results of model selection are discussed and presented for the BIC, ML and DIC.

La sélection de modèle dans le contexte bayésien présente des difficultés, particulièrement lorsque l'on considère des modèles hiérarchiques. Parmi les outils les plus utilisés afin de comparer différents modèles sont le Critère d'Information Bayésien (CIB), la Vraisemblance Marginale (VM) via les probabilités a posteriori des modèles, ainsi que le Critère d'Information de la Déviance (CID). L'objectif du projet était d'effectuer une sélection de modèles sur cinq régressions linéaires hiérarchiques ajustées à des données neurologiques longitudinales. Les régressions furent obtenues en utilisant l'échantillonneur de Gibbs, et pour chacun des modèles l'effet de deux ou trois covariables pouvait varier entre les sujets. L'implémentation et les résultats de la sélection de modèles sont discutés et présentés pour le CIB, la VM et le CID.

[MS-142]

**Kuhnian perspectives applied to the increasing popularity of Bayesian approaches in Health Care research: Is it a paradigm shift?**

**Perspectives de Kuhn appliquées à la popularité croissante de l'approche bayésienne en recherche sur la santé : est-ce que ce changement est un paradigme ?**

Afisi ISMAILA, *McMaster University*

Two statistical approaches, Bayesian and the frequentist approach, are fundamental to inference and decision-making in health care research. Both frequentist and Bayesian approaches address learning under uncertainty but they differ in a variety of ways. The frequentist approaches have been the dominant one until recently when the advances in computational methods have made the Bayesian methods a handy tool for overcoming some of the difficulties encountered by the use of frequentist methods. This talk uses the Kuhn's perspectives to examine whether Bayesian analysis is really a paradigm shift in medical research.

Les approches statistiques bayésiennes et fréquentistes sont fondamentales pour l'inférence et la prise de décision en recherche sur la santé. Ces deux approches traitent de l'apprentissage dans l'incertitude, mais elles diffèrent sur plusieurs aspects. L'approche fréquentiste a dominé jusqu'à récemment, mais les développements récents dans les méthodes numériques ont rendu l'approche bayésienne très pratique pour surpasser certaines difficultés rencontrées avec les méthodes classiques. Cette présentation utilise les perspectives de Kuhn pour examiner si le changement vers les méthodes bayésiennes est réellement un paradigme en recherche médicale.

**Session 10G****Mercredi 15 juin • Wednesday, June 15, 11:00 – 12:30****211**

Sampling and Design Issues

Questions de sondage et de planification

[MS-143]

**Two-treatment repeated measurement designs for a model with self and mixed carryover effects**

**Designs de mesures répétées pour deux traitements pour un modèle avec des effets résiduels uniques et mixtes**

YuanYuan LIANG & Keumhee Cho CARRIÈRE, *University of Alberta*

We consider the problem of finding  $p$ -period repeated measurements designs for two-treatment that minimize the variance of the best linear unbiased estimator (BLUE) of the treatment contrast based on the self and mixed carryover effects model proposed by Afsarnejad and Hedayat (2002, JSPI). For  $p = 2$ , contrasts in direct treatment effects cannot be estimated unbiasedly under the model being considered. For  $p > 2$ , baseline measurements improve efficiency only slightly for  $p$  odd, and are of no help for  $p$  even. Therefore, use of the baseline measurement is not recommended when constructing optimal designs under the self and mixed carryover effects model.

Nous considérons le problème de trouver des designs de mesures répétées à  $p$  périodes pour deux traitements qui minimisent la variance du meilleur estimateur linéaire sans biais (MELSB) pour le contraste de traitements basé sur le modèle d'effets résiduels uniques et mixtes de Afsarnejad et Hedayat (2002, JSPI). Pour  $p = 2$ , les contrastes dans les effets directs du traitement ne peuvent pas être estimés sans biais pour le modèle considéré. Pour  $p > 2$ , des mesures de référence améliorent légèrement l'efficacité lorsque  $p$  est impair et n'aident pas du tout lorsque  $p$  est pair. Ainsi, l'utilisation d'une mesure de référence n'est pas recommandée pour construire des plans optimaux sous le modèle d'effets résiduels uniques et mixtes.

**[MS-144]****Modelling left censored durations in analyzing repeated durations data****Modélisation de durées censurées à gauche dans l'analyse de données de durées répétées**

Ali Reza FOTOUHI, *University College of the Fraser Valley*

Two types of sampling can be considered in renewal processes. In the first type of sampling each individual is under observation within a fixed observational window. In the second type of sampling each individual is under observation for a fixed number of completed durations, from the start of sampling. Simulation results and an analysis of real data demonstrate the substantial parameters estimation biases which can occur in analysing repeated durations data if proper account is not taken of interrupted, left-censored, durations at the beginning of the data window. Conditional or joint likelihood methods overcome this problem.

Deux types d'échantillonnage peuvent être considérés pour des processus de renouvellement. Dans le premier type d'échantillonnage, chaque individu est sous observation pour une fenêtre de suivi fixe. Dans le second type, chacun des individus est sous observation pour nombre fixe de durées complétées à partir du début de l'échantillonnage. Des résultats de simulation et une analyse de données réelles démontrent le biais substantiel dans l'estimation des paramètres lorsqu'on analyse des données de durées répétées, si l'on ne tient pas compte des durées interrompues censurées à gauche au début de la fenêtre d'observation. Les méthodes de vraisemblances conditionnelles ou conjointes peuvent régler ce problème.

**[MS-145]****Sampling schemes for studying growth of children****Plans d'échantillonnage pour étudier la croissance des enfants**

Ram YADAV & Padam SINGH, *Indian Council of Medical Research*

The studies relating to growth children generally cover the age from 0 to 18 years. In these studies, the parameters of interest are the norm (i.e. the average of the characteristic under study), growth (i.e. change in average) and velocity (i.e. change in the growth) at different ages during the period. For this, generally longitudinal schemes are used which take long time, pose a number of organizational problems and involve more expenses. In these situational problems and involve more expenses. In these situations, linked cross sectional scheme (LCS) considered by Rao & Rao (1966) is more feasible, less expensive and permits a more carefully planned collection of data. Singh and Yadav (1991) studied linked cross sectional scheme for studying growth of children. An attempt has been done in this paper to compare the efficiency of the Linked cross sectional schemes for estimation of norm, Growth and velocity during the period of growth of children. Generalized estimation procedures have been developed and expressions for the optimum estimators along with their variances have been derived. It has been observed that the linked cross sectional scheme provides more efficient estimators as compared to pure cross-sectional and pure longitudinal schemes.

Les études reliées à la croissance des enfants couvrent généralement les âges de 0 à 18 ans. Dans ces études, les paramètres d'intérêts sont la norme (i.e. la moyenne des caractéristiques à l'étude), la croissance (i.e. le changement dans la moyenne) et la vélocité (i.e. le changement de croissance) à différents âges durant la période de croissance. Pour cela, des plans longitudinaux sont généralement utilisés ce qui prend du temps, pose un certains nombre de problèmes organisationnels et implique plus de dépenses. Dans ces situations, le plan lié trans-sectionnel (LTS) considérés par Rao et Rao (1966) et plus plausible, moins coûteux et permet une collecte de données plus délicate. Dingh et Yadav (1991) ont étudié le plan lié trans-sectionnel pour analyser la croissance des enfants. Cet article tente de comparer l'efficacité des plans liés trans-sectionnels pour l'estimation de la norme, la croissance et la vélocité durant la période de croissance des enfants. Des procédures d'estimations généralisées ont été développées et des expressions pour les estimateurs optimaux ainsi que leur variance ont été dérivées. Il a été observé que le plan lié trans-sectionnel donne de meilleurs estimateurs comparativement aux plans purement trans-sectionnels et purement longitudinaux.

**[MS-146]****Asthma Prevalence and its relationship with age, sex, and location: A National Population Health Survey study: 1994-2001****La prévalence de l'asthme et sa relation avec l'âge, le sexe et la localisation : une étude sur l'Enquête nationale de santé publique 1994-2001**Sunita GHOSH & Punam PAHWA, *University Of Saskatchewan*

Asthma prevalence is on the rise, and is more prevalent in adult females compared to males. The purpose of the study was to examine the relationship of asthma prevalence with sex, age group [0-4, 5-14 (reference category), 15-34, 35-64 and 65+ years], and location [rural vs. urban]. National Population Health Survey, questionnaire data was examined at four time periods. Logistic regression analysis was conducted for each of NPHS cycles to study the above relationship. In order to adjust for clustering and estimating standard error of regression estimates in such complex survey data bootstrap weights were used. The results showed that asthma was more prevalent in females, urban location and in the age group 15-34 years.

La prévalence de l'asthme est en hausse et beaucoup plus fréquente chez les femmes adultes que chez les hommes. Le but de l'étude est d'examiner la relation entre la prévalence de l'asthme et le sexe, le groupe d'âge [0-4, 5-14 (catégorie de référence), 15-34, 35-64 et 65 ans+] et la localisation [rural vs urbain]. Dans l'Enquête nationale de santé publique (ESNP), des données provenant de questionnaires ont été examinées à quatre périodes de temps. Une analyse de régression logistique a été effectuée pour chaque cycle de l'ENSP pour étudier les relations mentionnées ci-dessus. Pour ajuster pour l'erreur de classification et d'estimation des estimés de régression pour de telles données de sondage complexe, des poids bootstrap ont été utilisés. Les résultats montrent que l'asthme est plus prévalente chez les femmes vivant dans une région urbaine et provenant du groupe d'âge 15-34 ans.

**Session 11A      Mercredi 15 juin • Wednesday, June 15, 13:30 – 15:00****133****Special Invited Session of the Survey Section****Séance spéciale de l'invité du groupe d'enquête****[MS-147]****Bootstrapping in complex surveys****Bootstrap pour des sondages complexes**Randy SITTER, *Simon Fraser University*

An overview of the developments in bootstrapping methods in complex surveys. Discussion will include the bootstrap, balanced repeated replications, balanced bootstrap, random grouping and related topics, partially balanced repeated replications, and relationships and comparisons between them. Notions of reduced numbers of replicate weights, publicly released data, and data disclosure issues will be touched upon. Some things have been accomplished, some are still being considered and some challenges still remain.

Nous effectuerons un survol des développements des méthodes bootstrap dans les sondages complexes. Les discussions porteront sur le bootstrap, les répliques répétées équilibrées, le bootstrap équilibré, la classification aléatoire et ses sujets connexes, les répliques répétées partiellement équilibrées ainsi que les relations et les comparaisons entre ces techniques. Nous aborderons également les notions de réduction du nombre de poids de réplique, des données rendues publiques et des problèmes de publications. Certaines difficultés ont été résolues, d'autres sont présentement considérées et plusieurs défis demeurent.

**Session 11B**      **Mercredi 15 juin • Wednesday, June 15, 13:30 – 15:00**      **214**

Case Study II: The National Longitudinal Study on Children and Youth (NLSCY)

Étude de cas II : L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ)

**[MS-148]****Presentation of the Data Set****Présentation du jeu de données**Peggy NG, *York University*

A sub-sample of the synthetic data file from the National Longitudinal Study on Children and Youth (NLSCY) is used to introduce the problems related to estimation and variance calculation in longitudinal surveys. Many issues relating to children's behaviour and care are important concerns for Canadians and so the data set was created with the objective that the student confronts a realistic situation in a longitudinal study and assesses the best way in which to analyze and present the data to answer the research questions.

Un sous échantillon du jeux de données synthétique de L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) est utilisé pour introduire les problèmes reliés à l'estimation et au calcul de la variance pour des sondages longitudinaux. Plusieurs problèmes reliés aux comportements et aux soins des enfants sont des préoccupations importantes des canadiens, ainsi le jeu de données fut créé avec l'objectif que l'étudiant confronte une situation réaliste dans une étude longitudinale et évalue la meilleur approche pour analyser et présenter des données pour répondre aux questions de recherche.

**Session 11C**      **Mercredi 15 juin • Wednesday, June 15, 13:30 – 15:00**      **241**

Analysis of gene expression data

L'analyse des données d'expression génétique.

**[MS-149]****Statistical methods for gene clustering from high-throughput data****Méthodes statistiques pour la classification de gènes à partir de données à haute résolution**Jennifer BRYAN, *University of British Columbia*

I use the term "high-throughput data" to encompass a large variety of current assays in which a response is measured across a range of conditions or subjects for a large number genes (often for practically an entire genome). This certainly includes transcriptional profiling via microarrays, as well as, for example, highly parallel phenotypic studies in the yeast deletion set. A common use of such data is to cluster genes, with the hope that apparent gene clusters will have substantial overlap with biological gene groups, such as pathways, protein complexes, or regulons. I will present statistical methods for interpreting the significance of gene clusters, statistical arguments for choosing divisive and partitioning algorithms over agglomerative, and an interesting use of survival analysis to compute a

J'utilise le terme " données à haute résolution " pour encadrer une grande variété d'études dans lesquelles une réponse est mesurée à travers un ensemble de conditions ou sujets pour un grand nombre de gènes (souvent un génome au complet). Ceci inclus entre autres le profilage transcriptionnel par les microréseaux, ainsi que les études phénotypiques fortement parallèles dans l'ensemble de délétion de la levure. Une utilisation habituelle pour de telles données est la classification de gènes, où l'on espère que les groupes apparents de gènes chevaucheront suffisamment les groupes biologiques de gènes tels que les voies, les complexes de protéines ou les régulons. Je présente des méthodes statistiques pour interpréter la signification des groupes de gènes et des arguments statistiques permettant de sélectionner des algorithmes de division et de partition pour l'agglomération. Je présente également une utilisation intéressante de l'ana-



p-value for hierarchical clusters. These methods and results will be illustrated in real datasets from ongoing collaborations, including HTP phenotypic studies of the yeast deletion set and time courses in both fermenting wine yeast and differentiating murine stem cells.

lyse de survie pour calculer une valeur-p pour des groupes hiérarchiques. Ces méthodes sont illustrées à l'aide de jeux de données réels tirés de collaborations en cours, incluant des études phénotypiques HTP de la délétion de la levure et du cours du temps dans la fermentation du vin et la différenciation des cellules souches de murin.

### [MS-150]

#### **Weighted likelihood estimation with application to micro-array data**

#### **Estimation par vraisemblance pondérée avec applications pour les données de microréseaux**

Steven WANG, *York University*

The (relevance) weighted likelihood is introduced to formally embrace a variety of statistical procedures that trade bias for precision. That likelihood, unlike the classical likelihood, combines all relevant information while inheriting many of its desirable features including good asymptotic properties. We propose to choose the adaptive weights by cross-validation. We show that the weighted likelihood estimator (WLE) with adaptive weights to be weakly consistent and asymptotically normal. Results of simulation studies have shown that WLE out-performs the traditional maximum likelihood estimator for small and moderate samples. An application to micro-array data is also demonstrated.

La vraisemblance pondérée est introduite pour inclure formellement une variété de procédures statistiques qui échangent biais contre précision. Contrairement à la version classique, cette vraisemblance combine toute l'information utile tout en héritant de plusieurs de ses bonnes propriétés asymptotiques. Nous proposons de sélectionner les poids adaptatifs par validation croisée et montrons que l'estimateur de vraisemblance pondérée (EVP) avec les poids adaptatifs est faiblement consistant et asymptotiquement normal. Des résultats d'études de simulations ont montrés que l'EVP performe mieux que l'estimateur du maximum de vraisemblance traditionnel pour des échantillons de petite taille petite ou de taille modérée. Une application aux données de microréseaux est également présentée.

### [MS-151]

#### **Gene selection and prediction of survival information under the Cox proportional hazards model**

#### **Sélection de gènes et prévision de l'information de survie sous le modèle de taux de panne proportionnel de Cox**

Wenqing HE, *University of Western Ontario*

There is extensive research on the prediction of various clinical phenotypes using gene expression profiles. Success has been demonstrated in molecular classification of different cancer types. However, relatively less attention has been paid to study the connection of gene expressions to times to events of patients such as time to metastasis and time to death due to cancer. It is of scientific interest to select informative genes based on survival information in order to make prediction for cancer patients. In this talk, I will discuss a method to select survival relevant genes under the Cox proportional hazards model, and to build a predictive model using those selected genes. Numerical studies will be discussed to assess the proposed method.

Il y a beaucoup de recherche portant sur la prévision de différents phénotypes cliniques qui utilisent les profils d'expression des gènes. On a démontré beaucoup de succès en classification moléculaire de différents types de cancer. Cependant, beaucoup moins d'attention est portée à l'étude des connections entre l'expression des gènes et les temps jusqu'à un événement chez le patient, comme par exemple le temps jusqu'à l'arrivée d'une métastase ou le temps jusqu'à la mort dû à un cancer. Il est d'intérêt scientifique de sélectionner des gènes informatifs basé sur l'information de survie, pour effectuer des prévisions pour des patients atteints du cancer. Dans cette présentation, je discute d'une méthode de sélection de gènes qui ont un intérêt pour la survie sous le modèle de taux de panne proportionnel de Cox. Je discute également de la construction d'un modèle de prévision en utilisant les gènes sélectionnés auparavant. Je montre des

résultats d'études numériques pour évaluer la méthode proposée.

## **Session 11D**      **Mercredi 15 juin • Wednesday, June 15, 13:30 – 14:30**      **217**

Business and industrial statistics methods and applications

Méthodologie et applications de la statistique industrielle et de gestion

[MS-152]

### **Mining functional data**

#### **Exploration de données fonctionnelles**

Sofia MOSESOVA, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, Jock MACKEY & Stefan STEINER, *University of Waterloo*

In this talk we explore a number of different techniques for analyzing functional data, with a particular emphasis on curve-clustering. Following Ramsay & Silverman (1997), smoothing splines are used to examine differences between the curves using a functional version of analysis of variance. The coefficients of these smoothing splines are also utilized as low-dimensional summaries of the data, which are then used to cluster the curves. The proposed curve-clustering model is a generalized version of a model-based technique, which allows both random and fixed functional effects to be estimated simultaneously. These techniques are motivated and illustrated with a high-dimensional data set taken from the automotive industry.

Dans cette présentation, nous explorons un certain nombre de techniques pour analyser des données fonctionnelles avec une attention particulière sur la classification de courbes. En suivant la démarche de Ramsay et Silverman (1997), nous utilisons des splines de lissage pour analyser les différences entre les courbes avec une version fonctionnelle de l'analyse de variance. Les coefficients de ces splines de lissage sont également utilisés comme résumé à faible dimension pour les données et sont ensuite utilisés pour classer les courbes. Le modèle de classification de courbes proposé est une version généralisée d'une technique basée sur les modèles qui permet aux effets fonctionnels fixes et aléatoires d'être estimés simultanément. Ces techniques sont motivées et illustrées avec un jeu de données à grande dimension provenant de l'industrie automobile.

[MS-153]

### **A new mixture discriminant model for drug discovery data**

#### **Un nouveau modèle discriminant de mélanges pour des données de recherche de nouveaux médicaments**

Xu (Sunny) WANG, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, William WELCH, *University of British Columbia*

In drug discovery, statistical models are a powerful tool for predicting activity of compounds against biological targets. In this supervised learning problem, descriptors of molecular structure (e.g. atomic weight, types of bonds, many other exotic characteristics) are used to predict activity. The features of drug discovery data include the rareness of active compounds, multiple mechanisms, and high dimensional descriptor spaces. Conventional mixture discriminant methods have difficulty finding the best model for the drug discovery data due to the complication of data sets and the number of parameters. It is believed that the biological activity

En recherche de nouveaux médicaments, les modèles statistiques sont des outils puissants pour prédire l'activité de composés contre des cibles biologiques. Dans ce problème d'apprentissage supervisé, les descripteurs des structures moléculaires (ex. le poids atomique, les types d'amalgames et plusieurs autres caractéristiques exotiques) sont utilisés pour prédire une activité. Les caractéristiques des données de recherche de nouveaux médicaments incluent la rareté des composés actifs, les mécanismes multiples et la grande dimension de l'espace des descripteurs. Les méthodes discriminantes de mélanges conventionnelles ont de la difficulté à trouver le meilleur modèle pour les données dû à la complexité du jeu de données et au nombre de paramètres. On

of compounds only depends on several descriptors, so we introduce a new mixture model, which has fewer parameters, and seeks to predict using multiple subspaces (ie multiple mechanisms). The EM algorithm is used to estimate parameter, in conjunction with carefully chosen initial values.

croit que l'activité biologique de composés dépend seulement de quelques descripteurs, ce qui nous mène à présenter un nouveau modèle de mélange qui possède moins de paramètres et qui vise à prévoir en utilisant plusieurs sous-espaces (i.e. plusieurs mécanismes). L'algorithme EM est utilisé pour estimer les paramètres, en conjonction avec des valeurs initiales soigneusement choisies.

#### [MS-154]

##### **Affine futures and forward prices**

##### **Contrat à termes et évaluation postérieures affines**

Cody HYNDMAN, *University of Waterloo*

For a class of factor models for the price of a risky asset paying stochastic dividends in an economy with stochastic interest rates we show that when the factors process is defined by a Gaussian or square-root affine diffusion the forward price and futures price of the risky asset are exponential affine functions. The futures price and forward price are obtained by using the risk neutral measure for the reinvested futures price as numeraire and the risk neutral measure for the forward price reinvested in the zero-coupon bond as numeraire respectively. Calibration to market data is briefly discussed.

Pour une classe de modèles de facteurs du prix d'un titre risqué payant des dividendes stochastiques dans une économie avec des taux d'intérêts stochastiques, nous montrons que lorsque le processus des facteurs est défini par une diffusion gaussienne ou racine carré affine, les contrat à termes et les évaluation postérieures du titre risqué sont des fonctions exponentielles affines. Les prix des contrats à termes et des évaluation postérieures sont obtenus en utilisant respectivement la mesure neutre au risque pour les prix contrats à termes réinvestis comme numéraire et la mesure neutre au risque pour les prix contrats à termes réinvestis dans des obligations à coupons zéro comme numéraire. Nous discutons brièvement de la calibration pour des données de marché.

#### [MS-155]

##### **Statistical distributions for service times**

##### **Distributions statistiques pour des temps d'appels de service**

Adebolanle ADEDIGBA, *University of Saskatchewan*

A (customer) contact center is a collection of resources providing an interface between a service provider and its customers. Management of contact centers is complicated because it is essential to serve many different call types requiring different service skills in a timely fashion; thus there is need to maintain a reasonable service level. Quantifying this service level requires modeling of service time. Traditionally, in the context of queuing theory, exponential distribution has been used in modeling service times due to its markov property, thus making the analysis of the queue convenient (M|M|S|N). But unfortunately, real life data seldom follow exponential distribution. In this presentation, lognormal distribution as a statistical model, is discussed and fitted to overall service time of a help desk and to four major service types.

Un centre de service à la clientèle constitue un groupe de ressources qui fourni une interface entre un fournisseur de service et sa clientèle. La gestion des centres de service est compliquée parce qu'il est essentiel de répondre à plusieurs types d'appels nécessitant différentes qualifications, le tout dans un délai respectable tout en conservant un niveau de qualité de service raisonnable. La quantification de ce niveau de qualité de service requiert la modélisation du temps d'appel.

Traditionnellement, dans le contexte de la théorie des files d'attente, la distribution exponentielle a été utilisée pour modéliser les temps d'appels à cause de sa propriété markovienne, rendant commode l'analyse du modèle (M|M|S|N). Cependant, les données réelles suivent rarement des distributions exponentielles. Dans cette présentation nous discutons de la distribution log normale comme modèle statistique et

Phase-type distribution is introduced and fitted to the same data.

nous ajustons ce dernier au temps de service global d'un centre de service à la clientèle pour quatre types de services. Une distribution de type phase est introduite et ajustée aux mêmes données.

## **Session 11E      Mercredi 15 juin • Wednesday, June 15, 13:30 – 15:00      200**

### **Clinical Trials and Related Issues Essais cliniques et questions associées**

#### **[MS-156]**

#### **Efficient estimation of semi-parametric models for censored cost-effectiveness data**

#### **Estimation efficace de modèles semi paramétriques pour des données censurées coût-efficacité**

Eleanor PINTO, *University of Toronto & Hospital for Sick Children*, Andrew WILLAN,

Cost-effectiveness analysis of clinical trial data usually focuses on the differences in mean cost and effectiveness between two treatment groups. These differences can be estimated through semi-parametric models that express a patient's expected cost and expected effectiveness as linear functions of baseline covariates and treatment groups. Appropriate handling of censored data is vital if the estimators for the regression coefficients are to be consistent. Although inverse-probability weighting achieves this, the resulting estimators may not be efficient. By applying existing results on semi-parametric efficiency, we suggest new estimators, and evaluate their efficiency relative to the inverse-probability estimators through a simulation study.

L'analyse coût-efficacité de données d'essais cliniques focalise habituellement sur les différences entre les coûts moyens et l'efficacité entre deux groupes de traitement. Ces différences peuvent être estimées par des modèles semi paramétriques qui exprime les coûts et l'efficacité anticipés d'un patient comme une fonction linéaire de covariables et de groupes traitement sous-jacents. La manipulation appropriée de données censurées est vitale pour la consistance de l'estimateur des coefficients de régression. Même si la pondération de probabilité inverse réussie cela, l'estimateur résultant n'est pas nécessairement efficace. En appliquant des résultats existant sur l'efficacité semi paramétrique, nous suggérons des nouveaux estimateurs et évaluons leur efficacité par rapport aux estimateurs de probabilité inverse à l'aide d'une étude de simulation.

#### **[MS-157]**

#### **Competing risks methodology enhances interpretation of the defibrillator in acute myocardial infarction trial (DINAMIT)**

#### **La méthodologie de risques concurrents améliore l'interprétation du défibrillateur lors d'essais sur les infarctus du myocarde aigus (DINAMIT)**

Kevin THORPE, *University of Toronto*, Robin ROBERTS, *McMaster University*, Stefan HOHNLOSER, *J.W. Goethe University*, Paul DORIAN, *St. Michael's Hospital*, Stuart CONNOLLY & Michael GENT, *McMaster University*

Many clinical trials aimed at preventing death anticipate that the study intervention targets a specific cause of death. This assumption leads to a classic competing-risks problem, yet few trials consider it in their analysis and publication.

DINAMIT (NEJM 2004) is a randomized trial of optimal drug therapy with or without an Implantable Cardioverter-Defibrillator (ICD) in recent myocardial infarction survivors. A significant reduction in arrhythmic death in favour of ICD was off-

Plusieurs essais cliniques visant à prévenir la mort anticipent que l'intervention à l'étude cible une cause spécifique de mortalité. Cette supposition mène vers un problème classique de risques concurrents. Cependant, peu d'essais considèrent ce problème dans l'analyse et la publication des résultats.

DINAMIT (NEJM 2004) est un essai randomisé de thérapie par médication optimale avec ou sans un défibrillateur cardiovasculaire implantable (DCI) pour des survivants récents d'un infarctus du myocarde. Une réduction significa-

set by an increase on non-arrhythmic death for a net effect of no difference.

Competing-risks modelling suggest that ICD patients saved from arrhythmic death in DINAMIT simply died another way.

tive de la mortalité arythmique en faveur du DCI était effacée par une augmentation de la mortalité non arythmique résultant ainsi à aucun effet net.

La modélisation des risques concurrents suggère que les patients avec DCI dans l'essai DINAMIT qui ne sont pas morts d'un problème cardiaque sont tout simplement morts d'une autre cause.

#### [MS-158]

##### **The semiparametric mixture model for competing risks data**

##### **Modèle semi-paramétrique de mélanges pour des données de risques concurrents**

Gabriel ESCARELA, *Universidad Autonoma Metropolitana, Mexico City*

We study a model for survival analysis in samples which are subject to multiple modes of failure. We deal with a semiparametric specification of a mixture model which allows us to adjust for concomitant variables and to assess their effects on the probabilities of eventual causes of failure through a generalized logistic model, and of the corresponding conditional hazard functions by employing the Cox proportional hazards model. Estimation of the coefficients are obtained by using an EM algorithm based on a profile likelihood construction. The methods proposed are demonstrated on a prostate cancer data set.

Nous étudions un modèle d'analyse de survie pour des échantillons qui sont sujets à plusieurs types d'échecs. Nous utilisons une spécification semi-paramétrique d'un modèle de mélanges qui nous permet de tenir compte de variables concomitantes tout en évaluant leur effet sur la probabilité d'éventuelles causes d'échec à travers un modèle logistique généralisé. Les fonctions de pannes conditionnelles correspondantes, obtenues à l'aide d'un modèle de taux de panne proportionnel de Cox, sont aussi étudiées. Nous estimons les coefficients en utilisant un algorithme EM basé sur une construction profilée de la fonction de vraisemblance. Les méthodes proposées sont illustrées à l'aide d'un jeu de données sur le cancer de la prostate.

#### [MS-159]

##### **A copula model for the Markov regression of time series count data**

##### **Un modèle de copules pour la régression markovienne de données de comptage de séries chronologiques**

Gabriel ESCARELA, *Universidad Autonoma Metropolitana*

A general parametric class of transitional models for the analysis of serially correlated counts is presented. The first-order Markov chain model is defined by the transition probabilities, which are constructed from a bivariate copula, and the stationary distribution. Because overdispersion in the data is not attributable solely to the autocorrelation, we propose marginal models for overdispersion based on the Double Poisson family formulation. The proposed transition model is used to extend a class of mixture transition distribution models for the regression analysis of high-order Markov chains. Some relevant properties of the Markov chain are discussed.

Une classe paramétrique générale de modèles transitionnels pour l'analyse de dénombrements sériellement corrélés est présentée. Le modèle de chaînes de Markov du premier ordre est défini par les probabilités de transitions qui sont construites à partir d'une copule bivariable et de la distribution stationnaire. Puisque la surdispersion dans les données n'est pas attribuable uniquement à l'autocorrélation, nous proposons des modèles marginaux basés sur la famille Double Poisson. Le modèle de transition proposé est utilisé pour étendre une classe de modèles de mélanges de distributions de transitions pour l'analyse de régression de chaînes de Markov d'ordre élevé. Quelques propriétés intéressantes des chaînes de Markov seront discutées.

**[MS-160]**

**Bayesian meta-analysis of trials of chemotherapy with radiotherapy in the management of patients with newly diagnosed locally advanced squamous cell or undifferentiated nasopharyngeal cancer**  
**Méta-analyse bayésienne d'essais de chimiothérapie avec radiothérapie pour la gestion de patients nouvellement diagnostiqués de cellules squameuses localement avancées ou d'un cancer rhino-pharyngien non différencié**

Xiaohui (Helen) GUO, Lehana THABANE, Kullathorn THEPHAMONGKHOL & George BROWMAN, *McMaster University*, Ian HODSON, *McMaster Univeristy*

We discussed the results of Bayesian meta-analysis of the effect adding chemotherapy to radiotherapy treatment in patients with newly diagnosed locally advanced squamous cell or undifferentiated nasopharyngeal cancer. Two major outcomes are two-year disease-free survival and two-year overall survival. Bayesian meta-regression models were used to determine the impact of timing and regimen of chemotherapy on survival. We also compared the Bayesian results with those obtained with the classical random effects model.

Nous discutons des résultats de la méta-analyse bayésienne de l'effet d'ajouter la chimiothérapie à un traitement de radiothérapie pour des patients nouvellement diagnostiqués de cellules squameuses localement avancées ou d'un cancer rhino-pharyngien non différencié. Les deux principales variables d'intérêt sont la survie sans maladie après deux ans et la survie globale après deux ans. Des modèles de méta-régression bayésiens ont été utilisés pour déterminer l'impact de la synchronisation et du régime de la chimiothérapie sur la survie. Nous avons également comparé les résultats bayésiens à ceux obtenus à partir des modèles à effet aléatoires classiques.

**[MS-161]**

**Assessment of the quality of reporting of obesity trials. A systematic review.**  
**Évaluation de la qualité de diffusion d'essais sur l'obésité, une revue systématique**

Rong CHU, James DOUKETIS, Lehana THABANE & Kurl CUDDY, *McMaster University*

This study is a systematic review of the literature to assess the quality of reporting of final publications of randomized controlled trials that investigated weight loss interventions in patients with obesity. The revised CONSORT checklist was adapted to evaluate reporting as well as other general questions on clinical assessments of obesity. The Pearson and partial correlation coefficients were calculated to describe the relationships between quality of reporting score and other variables that include publication year. Overall, the quality of reporting is very low and varied substantially between different obesity therapies.

Cette étude est une revue systématique de la littérature pour évaluer la qualité de la diffusion des publications finales des essais randomisés contrôlés qui analysent les interventions de perte de poids chez des patients obèses. La liste de vérification révisée CONSORT a été adaptée pour évaluer la diffusion ainsi que d'autres questions d'ordre générale sur l'évaluation clinique de l'obésité. Les coefficients de corrélation de Pearson et de corrélation partielle ont été calculés pour décrire la relation entre un score de qualité de la diffusion et d'autres variables qui incluent l'année de publication. De manière générale, la qualité de diffusion est très faible et varie substantiellement entre différentes thérapies pour l'obésité.

**Session 11F      Mercredi 15 juin • Wednesday, June 15, 13:30 – 14:30      212****Correlated Discrete Data  
Données discrètes corrélées****[MS-162]****Use of Markov transition model to analyze longitudinal smoking (WSSP3) data****Utilisation du modèle de transition markovien pour analyser des données longitudinales sur le tabagisme (WSSP3)**Rashid AHMED, Steve BROWN & Paul McDONALD, *University of Waterloo*

We examined a progressive three-state model for longitudinal data from a community trial that incorporates a random effect distribution. Our model allows the distributions of the transition times between states to depend on: covariates, duration of time in the previous state, and correlation structure for the individuals who remain in the same class or community. We tracked the progression of students annually through three states: non-smoker, regular smoker, and quitter. Exact timing of transition through states is unknown.

The EM algorithm was used to estimate the model parameters and the MCMC method was used to test different random effect distributions.

Nous examinons un modèle progressif à trois états qui inclut une distribution de l'erreur aléatoire, pour des données longitudinales d'une étude communautaire. Notre modèle permet une dépendance des distributions des temps de transitions sur les covariables, sur le temps passé dans l'état précédent et sur la structure de corrélation pour les individus qui restent dans la même classe ou communauté.

Nous suivons annuellement la progression d'étudiants à travers trois états : non fumeur, fumeur régulier et arrêt de fumer. Le temps exact de la transition entre les états est inconnu.

L'algorithme EM est utilisé pour estimer les paramètres du modèle et la méthode MCMC pour tester différentes distributions d'effets aléatoires.

**[MS-163]****Mixed Markov models for spatiotemporal multi-state processes****Modèles markoviens mixtes pour des processus spatio-temporels à plusieurs états**Farouk NATHOO & Charmaine DEAN, *Simon Fraser University*

Studies of recurring infection or chronic disease often collect longitudinal data on the disease status of subjects. Multi-state transitional models are often employed for analysis in such studies where at any point in time, individuals may be said to occupy one of a discrete set of states and interest centers on the transition process between states. In this work we discuss the use of Mixed Markov models for the analysis of such longitudinal data when the processes corresponding to different subjects may be correlated spatially over a region. Markov models incorporating spatially correlated random effects are introduced in the context of two applications. The first is a study of recurrent weevil infestation in British Columbia forests and the second a study of revascularization intervention in Quebec.

Les études d'infections récurrentes ou de maladies chroniques recueillent souvent des données longitudinales sur le statut de la maladie des sujets. Les modèles transitionnels multi-états sont souvent utilisés dans de telles études, où à n'importe quel moment les individus occupent un des ensembles d'états, et l'emphase est mise sur le processus de transition entre les états. Dans cette présentation, nous discutons de l'utilisation des modèles markoviens mixtes pour l'analyse de telles données longitudinales lorsque les processus correspondants à différents sujets peuvent être corrélés spatialement sur une région. Les modèles de Markov qui incorporent des effets aléatoires de corrélation spatiale sont introduits dans le contexte de deux applications. La première est l'étude d'une infestation récurrente de charançons dans les forêts de la Colombie-Britannique et la deuxième une étude sur les interventions de revascularisation au Québec.

**[MS-164]****Panel event history data with incompletely observed variables****Données d'événements historiques de panel avec des variables observées incomplètes**Bethany GIDDINGS, Richard COOK & Grace YI, *University of Waterloo*

Measurement error can have a complex effect on estimates arising from interval censored event history data. This talk will focus on the impact of covariate measurement error in progressive multi-state models with data arising from panel (intermittent) observation. This type of problem can arise frequently in clinical settings (e.g. when disease progression is of interest and patient information is collected during irregularly spaced physician visits). Empirical biases in covariate effects will be studied via simulation and ways of addressing measurement error in analysis will be discussed.

Les erreurs de mesures peuvent avoir un effet complexe sur les estimations provenant de données d'événements historiques censurées par intervalle. Cette présentation met l'accent sur l'impact des erreurs de mesures de covariables dans des modèles progressifs multi-états avec des données provenant d'observations de panel (intermittentes). Ce type de problèmes peut survenir fréquemment dans un cadre clinique (par exemple lorsqu'on veut suivre la progression d'une maladie et que l'information sur le patient est collectée sporadiquement lors de visite chez le médecin). Nous étudions les biais empiriques dans l'effet des covariables à l'aide de simulations et nous discutons des manières de considérer les erreurs de mesures dans les analyses.

**[MS-165]****Calibrating prediction intervals for recurrent events****Calibrage d'intervalle de prédiction d'événements récurrents**Marc FREDETTE, *HEC Montréal*

In many applications there are types of events occurring repeatedly over time. It is often of interest to predict the future number of occurrences of these events. Usually, these predictive statements are made using plug-in prediction intervals. However, it is well-known that this approach can provide intervals with a coverage probability substantially below the desired level.

Considering cases where these recurrent events can be modeled using non-homogeneous Poisson processes with random effects, I will present methods to obtain intervals with coverage probabilities closer to the desired level. This will be done using calibration techniques and Bayesian-like methods.

Il est fréquent de rencontrer des situations où des événements se produisent plusieurs fois dans le temps. Il est alors intéressant de prédire le nombre d'occurrences de ces événements dans le futur. Habituellement, ces prédictions sont effectuées à l'aide d'intervalles de prédiction de type « plug-in ». Malheureusement, cette approche nous procure souvent des intervalles ayant des probabilités de couverture sous les seuils désirés.

En supposant que l'occurrence de ces événements peut être modélisée avec des processus de Poisson non-homogènes avec effets aléatoires, je présenterai des méthodes nous permettant d'avoir des intervalles ayant des probabilités de couverture plus près des seuils voulus. Nous atteindrons ce but en utilisant des techniques de calibrage et des méthodes semblables aux méthodes bayésiennes.



<b>Session 12A</b>	<b>Mercredi 15 juin • Wednesday, June 15, 15:30 – 17:00</b>	<b>241</b>
CRM-SSC Award Address		
Allocution du lauréat du Prix CRM-SSC 2005		

## Session 12B Mercredi 15 juin • Wednesday, June 15, 15:30 – 17:00 214

### Design of experiments Planification d'expériences

#### [MS-166]

#### Designing experiments for multi-stage processes Planification d'expériences pour des processus multi-niveaux

Derek BINGHAM, *Simon Fraser University*, Elizabeth KELLY & David OLIVAS, *Los Alamos National Laboratories*,  
Randy SITTE, *Simon Fraser University*

Many industrial processes require several steps of production. To help learn about the system dynamics a designed experiment, using a fractional factorial plan, can be performed. The usual designs require independent observations, which can require too many operations of each stage of experimentation. Instead, experiment units can be grouped at each stage, allowing several units to be processed simultaneously. In this talk, a grouping scheme is introduced that maintains the fractional factorial structure of the design and also facilitates simple and efficient analysis. The selection of an optimal design in this setting is also discussed. The grouping scheme is demonstrated on a real experiment conducted at Los Alamos National Lab.

Plusieurs processus industriels requièrent plusieurs étapes de production. Pour aider à comprendre la dynamique du système, un plan d'expérience utilisant un plan factoriel fractionnel peut être utilisé. Les plans habituels requièrent des observations indépendantes, pouvant ainsi demander trop d'opérations à chacune des étapes de l'expérience. Par ailleurs, les unités expérimentales peuvent être groupées à chaque étapes, permettant ainsi à plusieurs unités d'être traitées simultanément. Dans cette présentation, nous introduisons une méthode de regroupement qui maintient la structure factorielle fractionnelle du plan d'expérience et qui facilite les analyses simples et efficaces. La sélection du plan optimal dans cette situation est également discutée. La méthode de regroupement est illustrée pour une expérience réelle conduite au Los Alamos National Lab.

#### [MS-167]

#### Designing efficient simulations for exploring features of response surfaces Conception de simulations efficaces pour explorer les caractéristiques de surfaces de réponses

George MICHAILIDIS, *University of Michigan*, Derek BINGHAM & Pritam RANJAN, *Simon Fraser University*

In many engineering applications, one is interested in identifying the values of the inputs in computer experiments that lead to a response above a pre-specified threshold. In this talk we introduce statistical methodology that identifies the desired contour in the input space. The proposed approach has three main components. Firstly, a stochastic model is used to approximate the global response surface. The model is used as a surrogate for the underlying computer model and provides an estimate of the contour together with a measure of uncertainty, given the current set of computer trials. Then, a strategy for choosing subsequent computer experiments that improve the estimation of the contour is outlined. Finally, we discuss how the contour is extracted and represented. The methodology is illustrated with an example from a multiclass queueing system.

Dans plusieurs applications en ingénierie, on est intéressé à identifier les valeurs des intrants qui mènent à une réponse supérieure à un seuil prédéfini dans des expériences informatiques. Dans cet exposé, nous introduisons une méthodologie statistique qui identifie les contours désirés dans l'espace des intrants. L'approche proposée possède trois composantes principales. Premièrement, un modèle stochastique est utilisé pour approximer la surface de réponse globale. Le modèle est utilisé comme un substitut au modèle informatique sous jacent et donne un estimé du contour et une mesure d'incertitude basée sur l'ensemble actuel des essais informatiques. Ensuite, une stratégie pour sélectionner les expérimentations informatiques subséquentes pour améliorer l'estimation du contour est illustrée. Finalement, nous discutons comment le contour est extrait et représenté. La méthode est illustrée par un exemple d'un système multiclass de file d'attente.

[MS-168]

**Expected number of follow-up runs and cost for two-level fractional**

**Nombre de séquences subséquentes espéré et coûts pour un plan fractionnel à deux niveaux**

Crystal LINKLETTER, Randy SITTER & Boxin TANG, *Simon Fraser University*

Two-level fractional factorial (FF) designs are convenient for screening experiments because they can be used to search through a large number of candidate factors in only a fraction of the trials required by a full factorial design. To achieve this reduction in run-size, however, some effects are confounded, or aliased, and so the results of the analysis may be ambiguous. It is often suggested that screening should be viewed as a two-stage process: First, an experiment is chosen, performed, and analyzed. Given these results, at the second stage, supplemental runs are conducted to resolve any ambiguities and isolate the active factors. With this sequential approach in mind, clearly it would be advantageous to be able to anticipate at the outset how many additional runs will likely be needed at the second stage given the original design. This is a practical consideration in terms of choosing a good design, estimating and reducing costs, and allocating resources at the beginning of experimentation. It has been noted in passing that the number of follow-up runs that will be needed to identify the true underlying model depends on how the active effects are distributed over the alias structure of the initial design (e.g. Cheng, Steinberg and Sun, 1999; Miller and Sitter, 2001). In this article, we present specific techniques for obtaining the expected number of follow-up runs for a design, evaluate it for several designs, and use it as a new way to compare minimum aberration (MA) and maximum clear effect (CE) designs under differing assumptions of a priori knowledge.

Les plans fractionnels factoriels (FF) à deux niveaux sont utiles pour filtrer des expériences parce qu'ils peuvent être utilisés pour chercher à travers un grand nombre de facteurs potentiels avec seulement une fraction du nombre d'essais requis par un plan complètement factoriel. Cependant, pour obtenir cette réduction du nombre de séquences, certains effets sont confondus et ainsi les résultats de l'analyse peuvent être ambigus. Il est souvent suggéré que le filtrage devrait être vu comme un processus à deux niveaux. Premièrement, une expérience est choisie, mise en application et analysée. Selon les résultats obtenus à l'étape précédente, d'autres séquences expérimentales sont effectuées pour résoudre toute ambiguïté et isoler les facteurs actifs. Avec cette approche séquentielle en tête, il serait clairement avantageux de pouvoir anticiper le nombre de séquences expérimentales nécessaire au deuxième niveau sachant le plan original. Ceci est une considération pratique de sélection d'un bon plan d'expérience, d'estimation, de réduction des coûts et d'attribution des ressources au début de l'expérience. Il est à noter que le nombre de séquences subséquentes requises pour identifier le vrai modèle sous jacent dépend de la distribution des effets actifs sur les structure confondues du plan initial (Cheng, Steinberg and Sun, 1999 ; Miller and Sitter, 2001). Dans cette exposé, nous présentons des techniques spécifiques pour obtenir le nombre espéré de séquences subséquentes pour un plan, nous les évaluons pour plusieurs plans et nous les utilisons comme une nouvelle approche pour comparer les plans du nombre minimal d'aberrations (MA) et du maximum d'effets clairs (CE) sous différentes hypothèses de connaissance a priori.

**Session 12C      Mercredi 15 juin • Wednesday, June 15, 15:30 – 17:00      217**

Large deviations in statistics  
Déviations importantes en statistique

**[MS-169]**

**The complete convergence rates and the law of logarithm for the dependent bootstrap of the mean  
Taux de convergence complet et loi du logarithme pour la moyenne bootstrap dépendante**

S. Ejaz AHMED, *University of Windsor*, A.I. VOLODIN, *University of Regina*

For a sequence of random variables the convergence rate (that is Baum-Katz/Hsu-Robbins/Spitzer complete convergence type result) and the law of logarithm is obtained for the dependent bootstrapped means. In this investigation, the sequences of random variables need not necessarily be independent or identically distributed. Further, no assumptions are made concerning either the marginal or joint distributions of the random variables.

Pour une séquence de variables aléatoires, le taux de convergence (c'est-à-dire un résultat du type convergence complète de Baum-Katz/Hsu-Robbins/Spitzer) et la loi du logarithme sont obtenus pour la moyenne bootstrap dépendante. Dans cette étude, les variables aléatoires de la séquence ne doivent pas nécessairement être indépendantes et identiquement distribuées. De plus, aucune hypothèse n'est faite sur la distribution marginale ou conjointe des variables aléatoires.

**[MS-170]**

**Poisson limits for empirical point processes  
Limites pour des processus ponctuels empiriques**

Andre DABROWSKI, *University of Ottawa*, Rafal KULIK, *University of Ottawa & Wroclaw University*, Gail IVANOFF, *University of Ottawa*

For a sequence of independent and identically distributed random vectors, one may define the  $n$ th empirical point process as the random point measure with masses at the first  $n$  observations, suitably scaled. For appropriate scaling constants, we obtain the weak limit of these point processes through a novel use of a method based on convergence of the compensators of multiparameter-indexed martingales. The method extends previous results on empirical point processes of random variables to the case of random vectors, and limits at interior quantiles. Applications are provided to nearest-neighbour density estimation in high dimensions, and to multivariate extremes.

Pour une suite de vecteurs aléatoires indépendants et identiquement distribués, on peut définir le  $n$ ème processus ponctuel empirique comme la mesure ponctuelle aléatoire avec des masses aux  $n$  premières observations, convenablement proportionnés. Pour des constantes appropriées, nous obtenons la limite faible de ces processus ponctuels par une utilisation novatrice d'une méthode basée sur la convergence des compensateurs des martingales à indice multidimensionnelle. La méthode étend des résultats précédents des processus ponctuels empiriques des variables aléatoires au cas des vecteurs aléatoires, et aux quantiles intérieurs. Des applications sont fournies à l'estimation de densité de plus proches voisins dans des dimensions élevées, et aux extrêmes multidimensionnels.

[MS-171]

**Asymptotic results for an  $m$  out of  $n$  Bayesian bootstrap****Résultats asymptotiques pour un bootstrap bayésien de  $m$  éléments dans  $n$** 

Mahmoud ZAREPOUR, *University of Ottawa*, Hemant ISHWARAN, *Cleveland Clinic Foundation*, L.F. JAMES, *Hong Kong University of Science and Technology*

It is a well known fact that the Efron's usual bootstrap fails for the mean of heavy tail distributions. We demonstrate that this same type of failure also occurs in many types of weighted bootstraps, including the Bayesian bootstrap. Only through a careful selection for the weights we can produce a consistent weighted bootstrap procedure that will work. Examples illustrating the application of these methods in cases where the bootstrap is known to fail are given and further generalizations are suggested.

C'est un fait connu que le bootstrap usuel développé par Efron échoue pour la moyenne de distributions à queues épaisses. Nous démontrons que cette même difficulté survient également dans plusieurs type de bootstrap pondéré, incluant le bootstrap bayésien. C'est seulement par une sélection minutieuse des poids que nous pouvons construire une procédure de bootstrap pondéré convergente. Nous fournissons des exemples illustrant ces méthodes dans les cas où nous savons que le bootstrap échoue et nous suggérons des généralisations futures.

**Session 12D****Mercredi 15 juin • Wednesday, June 15, 15:30 – 17:00****133****Model and design based inference for sample surveys****Inférence fondée sur le modèle et sur le plan pour les sondages**

[MS-172]

**How important is the informativeness of the sample design?****Quelle est l'importance de "l'informativité" du plan d'échantillonnage ?**

Milorad KOVACEVIC, David BINDER & Georgia ROBERTS, *Statistics Canada*

Typical complex sample designs lead to informative samples, that is, the distribution of a variable in the sample is different from its distribution in the population. To determine and measure the impact of informativeness we compare design-based and model-based variances of estimated parameters (as well as the estimated parameters themselves) in a logistic model under the assumption that the postulated model is true. For an appropriate modelling of data from informative samples we consider two possibilities: to use the design-based inference about the model parameters or to use the model-based inference. In order to address the additional complexity of the impact of informativeness on power, we propose a new approximation for a linear combination of non-central chi-square distributions, using generalized design effects. A large simulation study, based on generating a population under the postulated model, using parameter estimates derived from the NPHS, allows us to detect and to measure the informativeness, and to compare the robustness of studied approaches.

Les plans d'échantillonnage complexes typiques mènent à des échantillons informatifs, c'est à dire que la distribution d'une variable dans l'échantillon est différente de celle dans la population. Pour déterminer et mesurer l'impact de "l'informativité", nous comparons les variances associées au plan de sondage et celles basées sur des modèles de paramètres estimés (ainsi que l'estimation des paramètres) pour un modèle logistique sous l'hypothèse que le modèle est adéquat. Pour une modélisation appropriée des données à partir d'un échantillon informatif, nous considérons deux possibilités pour évaluer les paramètres du modèle : la première consiste à utiliser l'inférence dans le contexte de l'utilisation d'un plan de sondage et la deuxième l'approche modéliste. Pour adresser la complexité additionnelle de l'impact de "l'informativité" sur la puissance, nous proposons une nouvelle approximation pour une combinaison linéaire de distributions khi-deux non centrées et utilisant les effets de plan généralisés. Une vaste étude de simulation où l'on génère une population à partir du modèle postulé en utilisant les paramètres estimés à partir de NPHS nous permet de détecter et de mesurer "l'informativité", et également de comparer la robustesse des approchées étudiées.

**[MS-173]****Model robust inference in survey sampling****Inférence robuste basée sur des modèles en échantillonnage**Narasimha PRASAD & Subhash LELE, *University of Alberta*

Model based and model assisted approaches are known to perform better than the purely design-based approach, provided the assumed model is appropriate. We consider a convex combination of a parametric and a non-parametric model where the weight for the parametric model is determined based on the ‘adequacy’ of it. This compromise between parametric and non-parametric models provides bias-variance trade-off. This approach provides robust inference against model misspecification. We evaluate the performance of this method for stratified random sampling and PPS with replacement sampling designs using simulations. We illustrate the method using family expenditure survey data.

Les approches modélistes et assistées par des modèles sont reconnues pour mieux performer que les approches purement randomisées, en autant que le modèle proposé soit approprié. Nous considérons une combinaison convexe d’un modèle paramétrique et non paramétriques où le poids pour le modèle paramétrique est basé sur la validité de ce dernier. Ce compromis entre le modèle paramétrique et non paramétrique permet un compromis entre le biais et la variance. Cette approche est robuste pour la mauvaise spécification du modèle. Nous évaluons la performance de cette méthode par des simulations pour des plans d’échantillonnages stratifiés et PPS avec remplacement. Nous illustrons la méthode en utilisant des données de sondage sur les dépenses familiales.

**[MS-174]****IRT modeling for latent variable in complex surveys****La modélisation par la TRI de variables latente pour les enquêtes à plan complexe**Andre CYR, Alexander DAVIES & Claude GIRARD, *Statistics Canada*

Item Response Theory (IRT) is a popular modeling tool for measuring latent traits. IRT methods have a long history in the field of psychometrics, and provide a model-based method for characterising both test items and predictions of individual traits. Statistics Canada’s National Longitudinal Survey of Children and Youth (NLSCY) is one of many surveys that use IRT to derive latent variable measurement. The presentation will outline the many steps taken by Statistics Canada to apply psychometric techniques to derive appropriately measured model parameters in the complex design setting and to derive relevant ability scores for children to support longitudinal analysis.

La Théorie de la Réponse aux Items (TRI) est un outil populaire pour modéliser les traits latents. Les méthodes de la TRI a un historique établi en psychométrie et permet la caractérisation par modèle des items d’un test ainsi que de la prédiction des traits des individus. L’Enquête Longitudinale Nationale des Enfants et des Jeunes (ELNEJ) est une des enquêtes de Statistique Canada qui utilise la TRI afin de dériver des mesures pour des variables latentes. La présentation sera un survol des plusieurs étapes prises par Statistique Canada pour dériver les paramètres de ces modèles psychométriques pour des enquêtes à plan complexe, afin de dériver les mesures pertinentes de capacité pour les enfants pour des fins d’analyse longitudinale.

**Session 12E      Mercredi 15 juin • Wednesday, June 15, 15:30 – 16:45      211****Theoretical survey methods  
Méthodes d'enquête théoriques****[MS-175]****The outlier detection and modeling method proposed for the gst replacement project of the monthly wholesale and retail trade survey****Détection de valeurs aberrantes et méthode de modélisation pour le projet de remplacement du sondage mensuel sur les ventes aux détails et en gros par des données de la TPS**Robert PHILIPS, *Statistics Canada*

Statistics Canada undertakes the Monthly Wholesale and Retail Trade Survey (MWRTS) which produces estimates based on monthly data collected for sales and inventories at various province and industry levels. These estimates are a substantial portion of the estimates for the Gross Domestic Product (GDP) and the sales trend is an important economic indicator. In response to budgetary concerns and the need to reduce respondent burden coupled with the availability of good quality data obtained from Canada Revenue Agency's Goods and Services Tax (GST) Program Statistics Canada has embarked on a process whereby the use of this GST data would be integrated into sub-annual business surveys. This paper will outline the outlier detection and modeling method proposed in relating the sales of enterprises to GST data. The method is derived from a Bayesian hierarchical model for outliers in the univariate linear model. Results based on MWRTS data will be presented.

Statistique Canada effectue le sondage mensuel sur les ventes aux détails et en gros (EMCGD) qui donne des estimations basées sur des données mensuelles sur les ventes et les inventaires collectées dans différentes provinces et différents niveaux d'industries. Ces estimations constituent une portion substantielle de l'estimation du produit intérieur brute (PIB) et la tendance des ventes est un indicateur économique important. En raison de contraintes budgétaires, pour réduire le fardeau des répondants, et suite à la disponibilité de données fiables obtenues de l'Agence de Revenu Canada et du programme de la Taxe sur les produits et les services (TPS), Statistique Canada à entamer un projet dans lequel l'utilisation des données sur la TPS seraient intégrées à un sondage infra-annuel auprès des entreprises. Cette présentation met en évidence la détection des valeurs aberrantes et la méthode de modélisation proposée pour relier les ventes des entreprises aux données sur la TPS. La méthode est basée sur un modèle bayésien hiérarchique pour des valeurs aberrantes dans un modèle linéaire simple. Les résultats basés sur les données de l'EMCGD seront présentés.

**[MS-176]****Estimating proportions from survey data by using estimating equations with random recentering****L'estimation des Proportions avec des Plans de Sondage Complexes, en Utilisant des Équations Estimantes Aléatoirement Recentrées**Avinash SINGH & Catalin DOCHITOU, *Statistics Canada*

The problem of constructing point estimates, SE and CI for proportions is revisited from an estimating function (EF) perspective. It is known that the usual frequentist Normality-based CI performs poorly for small  $p$  or  $n$ , while Bayesian methods have problems in accounting for complex designs. Use of frequentist EF gives rise to Wilson type CI, with good coverage properties, but the fact that point estimates and SE's don't change suggests a theoretical incoherence.

Le problème de trouver des estimations ponctuelles et leurs écarts types et intervalles de confiance (IC) est revisité du point de vue des fonctions estimantes (FE). On sait que les intervalles de confiance fréquentistes, basées sur des suppositions de Normalité, ont une performance réduite, pendant que les méthodes Bayésiennes ont des problèmes quand on utilise des plans de sondage complexes. L'utilisation des fonctions estimantes produit des IC de type Wilson, avec des bons couvertures, mais le fait que l'estimation des paramètres et écart types ne change pas suggère une incohérence

Under a semiparametric Bayesian approach with approximate Normal distribution for the EF, along with a uniform prior, an empirical approximation to the posterior distribution is obtained, by solving randomly recentered estimating equations. Different methods are compared using a simulation study. Use of Edgeworth approximations as an alternative to normality is also considered for both frequentist and Bayesian settings.

théorique.

Utilisant une approche semi-paramétrique Bayésienne, avec une distribution approximativement Normale pour les FE et une distribution a priori uniforme, on obtient une approximation de la distribution a posteriori en résolvant des équations estimantes aléatoirement re-centrées. On compare des méthodes différentes à l'aide des simulations. L'utilisation des approximations Edgeworth comme une alternative à la Normale est aussi considérée dans un cadre fréquentiste ainsi que Bayésien.

#### [MS-177]

##### **A refinement of the regression composite estimator in the labour force survey**

##### **Une alternative à l'estimateur par la régression composite à l'enquête sur la population active**

Cynthia BOCCI & Jean-François BEAUMONT, *Statistics Canada*

The Labour Force Survey is a stratified multi-stage monthly survey with a rotating design. Its main objective is to provide estimates of employment and unemployment rates. Estimates are obtained by composite regression which uses covariates defined with previous month variables of interest. For the new entrants in the survey, these covariates are missing. This problem is currently dealt with by imputing the missing values and by modifying the values of non-missing units in order to maintain the design-consistency property of the resulting estimator. We propose a new way of dealing with missing values in the covariates using calibrated imputation.

L'enquête sur la population active est une enquête mensuelle, stratifiée à plusieurs degrés avec renouvellement de l'échantillon. L'objectif principal de l'enquête est de produire des estimations de taux d'emploi et de chômage. Les estimations sont obtenues par la méthode de la régression composite qui utilise des variables auxiliaires définies à partir de variables d'intérêt du mois précédent. Pour le groupe de renouvellement naissant, ces variables auxiliaires sont manquantes. Présentement, ce problème est réglé en imputant les valeurs manquantes et en modifiant les valeurs des unités non-manquantes de telle sorte que l'estimateur qui en résulte maintienne la propriété de convergence par rapport au plan. On propose une nouvelle façon de traiter les valeurs manquantes dans les variables auxiliaires en utilisant l'imputation calée.

#### [MS-178]

##### **Mean squared error estimation for small area means using survey weights**

##### **Estimation par l'erreur quadratique moyenne pour la moyenne sur un petit domaine en utilisant les poids d'échantillonnage**

Mahmoud TORABI & J. N. K. RAO, *Carleton University*

You and Rao (2002) proposed pseudo-EBLUP estimators of small area totals under nested error linear regression models, using survey weights to preserve design consistency as the area sample size increases. In addition, the estimators have a nice benchmarking property in the sense of adding up to a reliable direct estimator at a large area level. We have obtained second order approximations to the MSE of pseudo-EBLUP and its estimator that are approximately unbiased. Empirical results on the performance of the proposed approximations

You et Rao (2002) ont proposé un estimateur pseudo-EBLUP d'un total sur un petit domaine sous des modèles de régressions linéaires à erreurs emboîtées en utilisant les poids d'échantillonnage pour préserver l'efficacité du plan d'échantillonnage à mesure que la taille de l'échantillon du domaine augmente. De plus, les estimateurs ont une propriété intéressante associée aux estimateurs par calage dans le sens qu'ils se somment vers un estimateur direct fiable au niveau d'un domaine plus large. Nous avons obtenu des approximations du deuxième ordre pour l'EQM du pseudo-EBLUP et son estimateur qui sont approximativement sans



are also presented.

biais. Des résultats empiriques sur la performance des approximations proposées sont également présentés.

**[MS-179]**

**Use of generalized design effects for variance function modeling in small area estimation from survey data  
L'utilisation des effets de design généralisés pour la modélisation des fonctions de variance dans l'estimation sur de petits domaines à partir de données de sondage**

Harold MANTEL, Avi SINGH & Yong YOU, *Statistics Canada*

In small area estimation using survey data we typically combine a direct, design-based estimator with a synthetic estimator, with the weight of the direct estimator depending on the ratio of its variance to the model variance of the synthetic part. However, the direct variance estimate is usually unstable, and is often replaced by a smoothed estimate using generalized variance functions (GVF). In this paper we introduce a new approach to GVF modeling based on generalized design effects. The approach will be illustrated by estimation of unemployment rates for Census Metropolitan Areas and Urban Centres using Canadian Labour Force Survey data.

Habituellement, dans l'estimation sur de petits domaines avec des données de sondages, nous combinons un estimateur direct basé sur la randomisation avec un estimateur synthétique. Le poids donné à l'estimateur direct dépend alors du ratio de sa variance par rapport à la variance modélisée de la portion synthétique. Cependant, l'estimateur direct de la variance est habituellement instable et souvent remplacé par un estimateur lissé en utilisant des fonctions de variance généralisées (FVG). Dans cette présentation, nous introduisons une nouvelle approche à la modélisation des FVG basée sur les effets de design généralisés. L'approche est illustrée par l'estimation du taux de chômage pour les régions métropolitaines de recensement et les centres urbains à partir des données de l'Enquête canadienne sur le travail.

**Session 12F      Mercredi 15 juin • Wednesday, June 15, 15:30 – 16:45      212**

**Statistical modeling and applications  
Modélisation statistique et ses applications**

**[MS-180]**

**Fitting differential equations to data  
Ajuster des équations différentielles à des données**

Jiguo CAO, Gregor FUSSMANN & Jim RAMSAY, *McGill University*

Differential equations (DEs) are used to model the rate of change of a process defined on time, space, or some other continuum. In this paper, a method which uses profiling to reduce the number of parameters is given for estimating DEs from noisy data, building on methods for smoothing data with a roughness penalty defined by an arbitrary differential operator, developed by Jim Ramsay. The paper applies this method to real data from biology, using a nonlinear system of equations for describing a predator-prey interaction, published in *Science* by Gregor Fussmann.

Les équations différentielles (ED) sont utilisées pour modéliser le taux de changement d'un processus défini dans le temps, dans l'espace ou tout autre continuum. Dans cet article, une méthode permettant d'estimer des ED à partir de données bruitées et utilisant le profilage, afin de réduire la dimensionnalité du problème, est présentée. Cette méthode est basée sur les méthodes de lissage de données développées par Jim Ramsay, où une pénalité est définie à l'aide d'un opérateur différentiel arbitraire. L'article applique cette méthode à des données réelles tirées de biologie en utilisant un système d'équations non-linéaire pour décrire une interaction prédateur-proie publiée dans *Science* par Gregor Fussmann.

**[MS-181]****Modelling extremal dependence - two actuarial problems****Modélisation de la dépendance de valeurs extrêmes - deux problèmes actuariels**Bruce JONES, *University of Western Ontario*, Debbie DUPUIS, *HEC Montréal*

We present two actuarial problems for which accurate characterization of extremal dependence is important. The first involves the calculation of general liability reinsurance premiums that reflect loss amounts and loss adjustment expenses, the latter being dependent. The second involves the calculation of liabilities for joint and last survivor annuities for which the lifetimes are dependent. In both cases, we make use of tools for the analysis of multivariate extreme values in order to understand and model the dependence that is evident from the data.

Nous présentons deux problèmes actuariels pour lesquels une caractérisation correcte de la dépendance des valeurs extrêmes est importante. Le premier problème concerne l'évaluation de contrats de réassurance responsabilité civile où on doit comprendre le montant de sinistre et les frais de règlement. Ces variables aléatoires sont dépendantes. Le deuxième problème concerne le calcul des provisions pour les rentes réversibles où les durées de vie sont dépendantes. Dans les deux cas, nous utilisons les méthodes d'analyse des valeurs extrêmes multivariées pour l'étude et la modélisation de la dépendance qui ressort des données.

**[MS-182]****Conduction-based statistics for description of anisotropy in sphere packings****Statistiques basées sur la conduction pour décrire l'anisotropie dans des arrangements de sphères**Jeffrey PICKA & Travis STEWART, *University of New Brunswick*

Random sphere packings formed under gravity are anisotropic. To identify aspects of orientation in their structure, models of electrical conduction can be used to define statistics that identify preferred orientations. These are used to investigate anisotropies found in realizations of classical packing algorithms. The statistics can also be extended to analyze other hard-core point patterns and granular materials.

Les arrangements aléatoires de sphères formés par la gravité sont anisotropes. Pour identifier les aspects d'orientation dans leur structure, des modèles de conduction électrique peuvent être utilisés pour définir des statistiques qui identifient l'orientation de préférence. Ces derniers sont utilisés pour analyser les anisotropies trouvées dans les réalisations d'algorithmes classiques d'arrangements. Les statistiques peuvent également être étendues pour analyser d'autres patrons ponctuels complexes ou matériaux granuleux.

**[MS-183]****Identifying dynamical systems****Identification des systèmes dynamiques**Giles HOOKER & James RAMSAY, *McGill University*

Differential Equations are an increasingly popular and powerful tool in modeling real world phenomena. They have been used extensively in models of population dynamics, in neural processes and in chemical engineering. Increasingly, they are applied to noisy data and in situations where parameters are not known or are poorly estimated.

This talk presents a new approach to the problem of parameter estimation for systems that are described by non-linear differential equations. The methodology is based on using a spline approximation using the differential operator from the equation as a

Les équations différentielles sont des outils de plus en plus populaires et sont très puissantes pour modéliser des phénomènes naturels. Elles ont beaucoup été utilisées dans les modèles de dynamique des populations, dans les processus neuronaux et en génie chimique. Elles commencent à être de plus en plus utilisées pour des données bruitées et dans des situations où les paramètres sont inconnus ou bien sont mal estimés.

Cet exposé présente une nouvelle approche au problème de l'estimation des paramètres pour des systèmes définis par des équations différentielles non linéaires. La méthodolo-

smoothness penalty. Parameters within the penalty are then chosen concurrently with the spline coefficients to provide a good fit to the data. I will discuss some properties of the technique and diagnostic tools for analyzing model misspecification.

gie est basée sur l'approximation par des splines en utilisant l'opérateur différentiel comme pénalité de lissage. Les paramètres à l'intérieur de la pénalité sont ensuite choisis de manière concurrentielle avec les coefficients des splines pour obtenir un bon ajustement aux données. Je discute de quelques propriétés de la technique et des outils de diagnostics pour analyser la mauvaise spécification du modèle.

**[MS-184]**

**Multi-signal wavelet coherence**

**Cohérence d'ondelettes multi-sinaux**

David CAMPBELL & Jim RAMSAY, *McGill University*

Regions of the brain associated with the visual system simultaneously share information in several feedback loops to make the leap from seeing to recognizing objects. The language of communication is electrical impulses and instances of frequency coupling suggest transfer of information. The frequencies at which the coupling occurs change over time and the impulses at one frequency often impact another part of the brain at a different frequency. Wavelet coherence examines the information sharing in a small time frequency window between two areas. Ideas will be presented which extend coherence to impulse signals from multiple areas.

Les régions du cerveau associées au système visuel échangent simultanément de l'information par plusieurs boucles de feedback pour faire le saut entre l'action de voir et celle de reconnaître un objet. La communication se fait par des impulsions électriques et les patrons de couplage de fréquences suggèrent un transfert d'information. Les fréquences auxquelles les couplages surviennent changent dans le temps et les impulsions à une certaine fréquence ont souvent un impact sur une autre région du cerveau à une autre fréquence. L'analyse des patrons de cohérence d'ondelettes examine les échanges d'informations entre deux régions sur une petite fenêtre de fréquence de temps. Des idées seront présentées pour étendre l'analyse de cohérence à des impulsions de signaux provenant de plusieurs régions.

**Session 12G**

**Mercredi 15 juin • Wednesday, June 15, 15:30 – 17:00**

**200**

**Methods with Censored Data**

**Méthodologie pour les données tronquées**

**[MS-185]**

**Wavelet estimation of density and hazard rate for randomly right censored data**

**Estimation par ondelettes de la densité et du taux de panne pour des données censurés à droite aléatoirement**

Jahida GULSHAN & Alwell OYET, *Memorial University of Newfoundland*

In this study, wavelet estimators of probability density functions and hazard rates are constructed under randomly right censored data. Nonparametric approaches are adopted under the assumption that the density and hazard rate has no specific parametric form. We propose a modified wavelet kernel based estimator for probability density functions. Some currently available methods of density and hazard rate estimation using wavelets are compared to the modified approach.

Dans cette étude, des estimateurs par ondelettes de la fonction de densité et du taux de panne sont construits pour des données censurées à droite aléatoirement. Des approches non paramétriques sont adoptées sous l'hypothèse que la densité et le taux de panne n'ont pas de formes paramétriques spécifiques. Nous proposons un estimateur modifié basé sur le noyau d'ondelette pour la fonction de densité. Quelques méthodes existantes d'estimation de la densité et du taux de panne utilisant les ondelettes sont comparées à l'approche modifiée.

**[MS-186]****Smooth estimation of mean residual life under random censoring****Estimation lisse de la moyenne de vie résiduelle avec censure aléatoire**

Yogendra CHAUBEY & Arusharka SEN, *Concordia University*, Pranab K. SEN, *University of North Carolina*

The methodology developed in Chaubey et al (2003), generalizing the idea of Chaubey and Sen (1996), is adapted here for smooth estimation of mean residual life under random censoring. A modified weighting scheme is proposed, as in Chaubey and Sen (1999), which is different from the one based on Hille's theorem in Chaubey and Sen (1996). Asymptotic properties of the resulting estimator are investigated.

Nous adaptons la méthodologie développée dans Chaubey et autres (2003) qui généralise l'idée de Chaubey et Sen (1996) pour l'estimation lisse de la moyenne de vie résiduelle avec censure aléatoire. Un scénario de pondération modifié est proposé comme dans Chaubey et Sen (1999), mais n'est pas basé sur le théorème de Hille. Nous investiguons les propriétés asymptotiques de l'estimateur résultant.

**[MS-187]****A nonparametric trend test for  $k$  samples of doubly censored data****Un test de tendance non paramétrique pour  $K$  échantillons de données doublement censurés**

Hwashin SHIN, Timothy RAMSAY & Daniel KREWSKI, *University of Ottawa*

We propose a nonparametric statistical method to compare tumor incidence rates across multiple samples. In a typical carcinogenicity experiment, animals are randomly assigned to one of  $K$  groups; one is a control group and the others are treated with increasing doses of a chemical agent of interest. Animals dying during the study period are subject to necropsy to check if tumors have occurred in specific body tissues. Animals surviving to the end of the study are sacrificed and also subject to necropsy. For each animal the time of death and the tumor status (present or absent) are recorded.

Our interest is to see whether or not tumor incidence rates are increased as the dose of the test agent increases. Some tests which have been proposed to date are based on certain assumptions about tumor lethality or serial sacrifices. Our method, however, is for the case when such information is unavailable and only the death time and tumor indicator are available.

The interesting feature of this problem is that exact tumor times are never available. Either an animal died with a tumor, the time of whose occurrence is left censored, or else the animal died tumor-free and the tumor time is right censored. Jonckheere in 1954 developed a distribution free  $K$ -sample test against ordered alternatives for uncensored data. Gehan in 1965 proposed a 2-sample test for doubly censored data. In 1970 Breslow generalized the Kruskal-Wallis test for comparing  $K$ -samples sub-

Nous proposons une méthode statistique non paramétrique pour comparer le taux d'incidence d'une tumeur à travers plusieurs échantillons. Dans une étude de cancérogénicité typique, des animaux sont assignés aléatoirement à un des  $K$  groupes, un des groupes étant un groupe contrôle et les autres sont traités avec des doses croissantes d'un agent chimique d'intérêt. Les animaux qui meurent durant la période d'étude sont envoyés en nécropsie pour vérifier si des tumeurs sont apparues dans des tissus spécifiques du corps. Les animaux qui survivent jusqu'à la fin de l'étude sont sacrifiés pour être autopsiés. Pour chaque animal, le moment de la mort et le statut de la tumeur (présent ou absent) sont enregistrés.

Notre intérêt est de voir si les taux d'incidences d'une tumeur augmentent à mesure que la dose de l'agent chimique testé augmente. Certains tests qui ont été proposés jusqu'à ce jour sont basés sur des hypothèses sur la létalité de la tumeur ou les sacrifices sériels. Contrairement à ces derniers, notre méthode traite les cas où une telle information n'est pas disponible, et seulement le moment de la mort et l'indicateur de tumeur sont disponibles.

La caractéristique intéressante de ce problème est que le temps d'apparition de la tumeur n'est jamais disponible. Soit qu'un animal meurt avec une tumeur, le temps d'apparition est alors censuré à gauche, ou bien l'animal est mort sans aucune tumeur et ainsi le temps d'apparition est censuré à droite. En 1954, Jonckheere a développé un test non paramétrique pour  $K$  échantillons pour des données non censurées.

ject to unequal patterns of censorship for only right censored data. Combining ideas from these three methods, my talk will present a k-sample trend test for doubly censored data.

En 1965, Gehan a proposé un test pour deux échantillons de données doublement censurés. En 1970, Breslow a généralisé le test de Kruskal-Wallis pour comparer K échantillons sujets à des patterns inégaux de censure pour des données censurées à droite seulement. En combinant des idées de ces trois méthodes, mon exposé va présenter un test de tendance pour K échantillons de données doublement censurées.

[MS-188]

**Residual plots for accelerated life regression modelling of dependent bivariate time-to-event data**

**Graphiques des résidus pour la modélisation de temps de vie accéléré de données dépendantes et bivariées de temps jusqu'à un événement**

Yun Hee CHOI, David E. MATTHEWS & Grace Y. YI, *University of Waterloo*

We have previously described accelerated life regression models involving shared frailties to fit paired dependent time-to-event data. To assess model fit, we define a residual for each response measurement pair via the bivariate probability integral transformation of univariate residuals derived from the response measurements and the fitted model, and use these to confirm the choice of an appropriate frailty distribution. We also introduce an adjustment to accommodate right-censoring of the original responses. Through simulation studies and graphical displays, we characterize the sampling behaviour of these residuals and demonstrate how well these tools address key questions of model fit.

Nous avons préalablement décrits les modèles de régression de temps de vie accéléré impliquant des fratries partagées pour ajuster des données dépendantes pairées de temps jusqu'à un événement. Pour évaluer l'ajustement du modèle, nous définissons un résidu pour chaque paires de réponses selon la transformation intégrale de probabilité bivariée pour des résidus univariés provenant des mesures de réponses et du modèle ajusté. Ensuite, nous utilisons ces résidus pour confirmer le choix d'une distribution de fratrie appropriée. Nous introduisons également un ajustement pour tenir compte de la censure à droite des réponses originales. À partir d'études de simulations et de sorties graphiques, nous caractérisons le comportement d'échantillonnage de ces résidus et démontrons comment ces outils répondent bien aux questions d'ajustement de modèles.

[MS-189]

**Confidence intervals for quantiles of the lognormal distribution**

**Intervalles de confiances pour les quantile de distributions log-normales**

Steve OVERDUIN & Michael STEPHENS, *Simon Fraser University*

The log-normal distribution is frequently used in medical statistics, particularly in survival analysis. By using a new pivot, we give simple formulas for finding confidence intervals for quantiles, without the necessity of Monte Carlo sampling; such intervals are especially important when the data is right-censored. The intervals have very good coverage probabilities, and associated software is available. Some remarks will be offered on robustness of the intervals to mis-specification of the fitted distribution.

La distribution log-normale est fréquemment utilisée en statistique médicale, particulièrement en analyse de survie. En utilisant un nouveau pivot, nous donnons des formules simples pour trouver des intervalles de confiances pour les quantiles sans utiliser un échantillonnage de Monte Carlo. De tels intervalles sont spécialement importants lorsque les données sont censurées à droite. Les intervalles ont une très bonne probabilité de couverture et des logiciels associés sont disponibles. Quelques remarques seront présentées sur la robustesse des intervalles à la mauvaise spécification de la distribution ajustée.

[MS-190]

**Local linear estimators of conditional variance functions under random censorship**  
**Estimateurs linéaires locaux de fonctions de variance conditionnelle avec censure aléatoire**

Xuewen Lu, *University of Calgary*, Wanrong Liu, *Hunan Normal University*

In this paper, we consider the estimation of conditional variance in a heteroscedastic nonparametric regression model under random censorship. We propose two methods for estimating the conditional variance. One is the synthetic-residual-based estimator and the other is the synthetic-data-based direct estimator. The basic idea is to apply a synthetic-data method and a local linear regression to synthetic squared residuals or synthetic data. We establish the results of asymptotic normality for both estimators. Further, we demonstrate that, using synthetic data, without knowing the regression function and the censoring distribution, the synthetic-residual-based estimator can estimate the conditional variance asymptotically as well as if the regression and the censoring distribution were given. This implies that the synthetic-residual-based estimator is efficient and regression-adaptive for estimating the conditional variance. Numerical examples with both simulated and real data sets are used as illustration.

Dans cet article, nous considérons l'estimation de la variance conditionnelle dans un modèle de régression non paramétrique hétéroscédastique avec censure aléatoire. Nous proposons deux méthodes pour estimer la variance conditionnelle. Le premier estimateur est basé sur les résidus synthétiques et le deuxième est l'estimateur direct basé sur les données synthétiques. L'idée de base est d'appliquer une méthode de données synthétiques et une régression linéaire locale aux résidus carrés synthétiques ou aux données synthétiques. Nous établissons les résultats pour la normalité asymptotique pour les deux estimateurs. De plus, nous démontrons qu'en utilisant les données synthétiques sans connaître la fonction de régression et la distribution de censure, l'estimateur basé sur les résidus synthétiques peut estimer la variance conditionnelle asymptotiquement de même que si les distributions de régression et de censures étaient données. Ceci implique que l'estimateur basé sur les résidus synthétiques est efficace et régression-adaptatif pour estimer la variance conditionnelle. Des exemples numériques sur des données simulées et réelles sont utilisées comme illustration.

The translation team did its best. Please forgive us for our misinterpretations, we are neither professional translators nor specialists in all areas of statistics.

*L'équipe de traduction a fait de son mieux. Excusez-nous pour les erreurs de traduction, nous ne sommes ni des traducteurs professionnels ni des spécialistes dans tous les domaines de la statistique.*

## Author Index • Index des auteurs

ABARIN, Taraneh .....	109	CAMPOS-BONILLA, M. Begoña .....	89
ADEDIGBA, Adebolanle .....	133	CANTY, Angelo .....	88, 89
ADEWALE, Adeniyi .....	130	CAO, Jiguo .....	147
AHMED, Rashid .....	137	CARILLO, Ivan .....	130
AHMED, S. Ejaz .....	142	CARLIN, Brad .....	84
ALLEN, Brian .....	113	CARRIÈRE, Keumhee Cho .....	127
ALTMAN, Rachel .....	60	CARROLL, Raymond .....	81
ALVAREZ, Juan .....	68	CASANOVA-GURRERA, Amparo .....	89
ANDREEV, Kirill .....	88	CAUCHON, Caroline .....	111
ANGERS, Jean-François .....	76	CHANG, Edward .....	90
ARMSTRONG, Barbara .....	110	CHAPMAN, Judy-Anne .....	54
AUGER, Sylvie .....	74	CHARLEBOIS, Joanne .....	130
BAILEY, Carmen .....	90	CHARRON, Dominique .....	87, 113
BALAN, Raluca .....	62	CHAUBEY, Yogendra .....	150
BANERJEE, Sudipto .....	84	CHEN, Bingshu .....	119
BASINGER, Jim .....	59	CHEN, Gemai .....	71
BEAUMONT, Jean-François .....	146	CHEN, Jiahua .....	72, 73
BELLHOUSE, David .....	62	CHENOURI, Shojaeddin .....	102
BÉRARD-CHAGNON, Julien .....	130	CHIPMAN, Hugh .....	132
BERTHELOT, Jean-Marie .....	65	CHOI, Yun Hee .....	151
BIERNACKA, Joanna .....	66	CHU, Christina Chor-Yan .....	130
BINDER, David .....	143	CHU, Ka Lok .....	69
BINGHAM, Derek .....	140	CHU, Rong .....	136
BISCHOFF, Wolfgang .....	120	COLEMAN, Kimberley .....	108
BLACKWELL, Debra L. ....	64	CONNOLLY, Stuart .....	134
BLUMBERG, Stephen .....	65	COOK, Andrea .....	83
BOCCI, Cynthia .....	146	COOK, Richard .....	83, 119, 138
BOTTAI, Matteo .....	112	COONER, Freda .....	84
BOUHADDIUI, Chafik .....	86	COOPER, Michael .....	102
BOURBEAU, Robert .....	88	COPE, Eric .....	96
BOYETT, James .....	105	COREY, Paul .....	123
BRASHER, Penny .....	123	CRAIU, Radu .....	118
BROWMAN, George .....	136	CUDDY, Kurl .....	136
BROWN, Elizabeth .....	60	CYR, Andre .....	144
BROWN, Steve .....	137	DABROWSKI, Andre .....	142
BRUNSCHWIG, Hadassa .....	130	DAS, Kalyan .....	124
BRYAN, Jennifer .....	130	DAVIES, Alexander .....	144
BUITENDYK, Melanie .....	126	DAVIS, Karen .....	65
BULL, Shelley .....	66	DAVIS, Richard .....	93
CAMPBELL, David .....	149	DEAN, Charmaine .....	137
		DESGAGNÉ, Alain .....	76

- DICESARE, Joe ..... 100  
 DILLON, Jo-Anne ..... 59  
 DINU, Irina ..... 130  
 DOCHITOIU, Catalin ..... 145  
 DOLOVICH, Lisa ..... 88  
 DORIAN, Paul ..... 134  
 DOUKETIS, James ..... 136  
 DRTON, Mathias ..... 122  
 DUCHESNE, Thierry ..... 118  
 DUFOUR, Jean-Marie ..... 86  
 DUPUIS, Debbie ..... 148  
  
 EARL, Lawrence ..... 97  
 EDWARDS, Betty ..... 56  
 EINMAHL, Uwe ..... 104  
 EL-SAIDI, Mohammed ..... 102  
 ESCARELA, Gabriel ..... 135  
 EVANS, Michael ..... 76  
  
 FAN, Zizhong ..... 117  
 FANG, Miao ..... 59  
 FENG, Rui ..... 106  
 FERKO, Nicole ..... 126  
 FLEURY, Manon ..... 113  
 FOTOUHI, Ali Reza ..... 128  
 FRASER, Don ..... 63, 64  
 FREDETTE, Marc ..... 138  
 FU, Rao ..... 130  
 FUSSMANN, Gregor ..... 147  
  
 GAMITO, Eduard ..... 55  
 GENT, Michael ..... 134  
 GENTLEMAN, Jane ..... 64  
 GERACI, Marco ..... 112  
 GHOSH, Sunita ..... 129  
 GIDDINGS, Bethany ..... 138  
 GILLIS, Daniel ..... 87  
 GIRARD, Claude ..... 144  
 GRUBISIC, Maja ..... 108  
 GUAY, Mireille ..... 56  
 GULSHAN, Jahida ..... 149  
 GUO, Meijie ..... 108  
 GUO, Pengfei ..... 55  
 GUO, Xiaohui (Helen) ..... 136  
 GUSTAFSON, Paul ..... 71, 92  
 GUT, Allan ..... 104  
  
 HAN, Jing ..... 107  
 HASHEMI, Seyed ..... 61  
 HE, Wenqing ..... 131  
  
 HELBING, Caren ..... 90  
 ZHANG, Heping ..... 73  
 HEUNIS, Andrew ..... 61  
 HODSON, Ian ..... 136  
 HOETING, Jennifer ..... 93  
 HOHNLOSER, Stefan ..... 134  
 HOLBROOK, Anne ..... 88  
 HOLT, John ..... 87, 113  
 HOOKER, Giles ..... 148  
 HOOPER, Peter ..... 57  
 HORROCKS, Julie ..... 112  
 HORVATH, Steve ..... 95  
 HOWARD, Allan ..... 78  
 HSIAO, Cheng ..... 72  
 HUANG, Hui ..... 98  
 HUANG, Jian ..... 99  
 HYNDMAN, Cody ..... 133  
  
 ILHAMTO, Gatot ..... 108  
 ISHWARAN, Hemant ..... 143  
 ISMAILA, Afisi ..... 88, 127  
 ISOTALO, Jarkko ..... 69  
 IVANOFF, Gail ..... 142  
  
 JAMES, L.F. .... 143  
 JEMIAI, Yannis ..... 82  
 JIANG, Hedy ..... 88  
 JOHNSTON, Karissa ..... 90  
 JONES, Bruce ..... 148  
 JONES, Richard ..... 55  
 JUDKINS, David ..... 117  
  
 KARUNAMUNI, Rohana ..... 101  
 KARUNANAYAKE, Chandima ..... 58  
 KATZOFF, Myron ..... 109  
 KAWCZAK, Janusz ..... 68  
 KELLY, Elizabeth ..... 140  
 KELLY, I. W. .... 58  
 KEYFITZ, Barbara ..... 59  
 KHALILI, Abbas ..... 73  
 KHAN, Bashir ..... 103  
 KHARE, Meena ..... 65  
 KOHLI, Michele ..... 126  
 KORU-SENGUL, Tulay ..... 113  
 KOUIDER, Elies ..... 102  
 KOVACEVIC, Milorad ..... 94, 143  
 KOWGIER, Matthew ..... 130  
 KOZAK, Robert ..... 118  
 KRENZKE, Thomas ..... 117  
 KREWSKI, Daniel ..... 150



- KULIK, Rafal ..... 67, 142  
 KUSTRA, Rafal ..... 96  
 KWON, Jaimyoung ..... 97  
  
 LAROCHE, Richard ..... 110  
 LAROCHE, Sylvie ..... 75, 111  
 LATOUCHE, Michel ..... 74, 75  
 LAVERTY, W. H. .... 58  
 LAVERTY, William H. .... 58  
 LAWLESS, Jerry ..... 103  
 LE, Lisa Wen ..... 130  
 LE, Nhu ..... 87  
 LEBLANC, Alexandre ..... 101  
 LEBLANC, Nancy ..... 130  
 LEE, Sandra ..... 98  
 LEFEBVRE, Geneviève ..... 126  
 LELE, Subhash ..... 144  
 LESOSKY, Maia ..... 108  
 LESPERANCE, Mary ..... 90  
 LETAC, Gérard ..... 122  
 LI, Deli ..... 104  
 LI, Yi ..... 83  
 LI, Zhengqing ..... 106  
 LIANG, Yuanyuan ..... 127, 130  
 LIN, Jing ..... 59  
 LIN, Lillian ..... 109  
 LIN, Shili ..... 66  
 LIN, Xihong ..... 81  
 LINDSAY, Bruce ..... 107  
 LINKLETTER, Crystal ..... 141  
 LISAWADI, Supranee ..... 67  
 LIU, Jixin ..... 130  
 LIU, Mengling ..... 107  
 LIU, Wanrong ..... 152  
 LIU, Zhang ..... 130  
 LIU, Zhenqiu ..... 66  
 LIU, Zhong ..... 87  
 LIX, Lisa ..... 125  
 LOHR, Sharon ..... 80  
 LOU, Wendy ..... 89  
 LU, Irene ..... 93  
 LU, Pingguo ..... 86  
 LU, Wenbin ..... 107  
 LU, Xuewen ..... 152  
  
 MA, Renjun ..... 84  
 MA, Shuangge ..... 99  
 MAAROUF, Abdel ..... 87, 113  
 MACKAY, Jock ..... 70, 132  
 MACKINNON, Peter ..... 59  
  
 MACNEILL, Ian ..... 85  
 MANTEL, Harold ..... 147  
 MARCHAND, Éric ..... 76  
 MARTINEZ, Michael ..... 65  
 MARTINEZ, Michael E. .... 64  
 MASSAM, Hélène ..... 122  
 MATTHEWS, David E. .... 151  
 McDONALD, Paul ..... 137  
 McLEISH, Don ..... 100  
 McLEOD, Ian ..... 114  
 MELNICHOUK, Olga ..... 108  
 MERTON, Andrew ..... 93  
 MICHALIDIS, George ..... 97, 140  
 MICHAUD, Isabelle ..... 118  
 MIGLIORETTI, Diana ..... 60  
 MIKET, M. .... 58  
 MIREA, Lucia ..... 130  
 MONSOUR, Michael ..... 109  
 MOSESOVA, Sofia ..... 108, 132  
 MOSHONOV, Hadas ..... 76  
 MOUIHA, Aberazzak ..... 85  
 MYLES, Jamie ..... 62  
  
 NAHM, Erika ..... 108  
 NAIR, Kalpana ..... 88  
 NATHOO, Farouk ..... 137  
 NG, Peggy ..... 130  
 NOURIJELYANI, Keramat ..... 54  
  
 O'DONNELL, Colin ..... 55  
 OGDEN, Sheryl ..... 55  
 OH, Hee-Seok ..... 78  
 O'HARA HINES, Jeanette ..... 62  
 OLIVAS, David ..... 140  
 OVERDUIN, Steve ..... 151  
 OYET, Alwell ..... 55, 149  
  
 PAHWA, Punam ..... 129  
 PAKYARI, Reza ..... 101  
 PANTOJA GALICIA, Norberto ..... 94  
 PENG, Yingwei ..... 114  
 PENZO-GIACCA, Wilma ..... 89  
 PERRON, Francois ..... 121  
 PETERS, Amy ..... 108  
 PETKAU, John ..... 62  
 PHILIPS, Robert ..... 145  
 PICKA, Jeffrey ..... 148  
 PINTO, Eleanor ..... 134  
 PRASAD, Narasimha ..... 58, 144  
 PRETZEL, Shannon ..... 55

- PRINGLE-NELSON, C. .... 58  
 PROVOST, Serge ..... 85  
 PUNTANEN, Simo ..... 69  
  
 QU, Annie ..... 125  
  
 RAIS, Nourddine ..... 85  
 RAIZENNE, Mark ..... 56  
 RAMSAY, James ..... 148  
 RAMSAY, Jim ..... 147, 149  
 RAMSAY, Timothy ..... 150  
 RANJAN, Pritam ..... 140  
 RAO, J. N. K. .... 124, 146  
 RAO, J.N.K. .... 80  
 RAO, Jon N.K. .... 81  
 RAY, Surajit ..... 107  
 REEDMAN, Laurie ..... 111  
 REINKER, Stefan ..... 60  
 REN, Qunshu ..... 94, 124  
 RICHARDSON, Thomas ..... 122  
 ROBERTS, Georgia ..... 143  
 ROBERTS, Robin ..... 134  
 ROSS, Nancy ..... 65  
 ROSYCHUK, Rhonda ..... 61  
 ROUSSEAU, Judith ..... 63  
 ROYCE, Melanie ..... 54  
 RYAN, Nancy ..... 126  
  
 SALEH, A.K. Md. E ..... 103  
 SALOPEK, Donna ..... 100  
 SANJEL, Deepak ..... 85  
 SCHAFER, Joseph ..... 117  
 SCHAUBEL, Doulgas ..... 119  
 SCHENNACH, Susanne ..... 71  
 SCHWARZ, Carl ..... 124  
 SEN, Arusharka ..... 150  
 SEN, Pranab K. .... 150  
 SENGUL, Haydar ..... 90  
 SEVERINI, Tom ..... 63  
 SHAO, Yongzhao ..... 107  
 SHEN, Hui ..... 77, 130  
 SHEN, Samuel ..... 78, 79  
 SHIN, Hwashin ..... 150  
 SINGH, Avi ..... 147  
 SINGH, Avinash ..... 80, 145  
 SINGH, Padam ..... 128  
 SITTER, Randy ..... 129, 140, 141  
 SONG, Peter ..... 125  
 SOTEROS, Chris ..... 68  
 SOTEROS, Christine ..... 56  
  
 STADTMUELLER, Ulrich ..... 105  
 STAUDENMAYER, John ..... 60  
 STEELE, Russell ..... 126  
 STEINER, Stefan ..... 70, 132  
 STEPHENS, Michael ..... 151  
 STEWART, Travis ..... 148  
 STOFFER, David S. .... 113  
 STRAWDERMAN, William E. .... 76  
 STUDENY, Milan ..... 122  
 STYAN, George P. H. .... 69  
 SU, Wanhua ..... 130  
 SUN, Jiayang ..... 72  
 SUN, Lei ..... 66  
 SUN, Ning ..... 65  
 SUN, Ye ..... 130  
 SZAFRON, Michael ..... 56  
  
 TAI, Patricia ..... 54  
 TAN, Ming ..... 105  
 TANG, Boxin ..... 141  
 THABANE, Lehana ..... 88, 136  
 THEPHAMONGKHOL, Kullathorn ..... 136  
 THIESSEN, Heather ..... 62  
 THOMPSON, Mary ..... 94  
 THOMPSON, Sandra ..... 93  
 THOMPSON, Steven ..... 109  
 THORPE, Kevin ..... 134  
 TORABI, Mahmoud ..... 146  
 TREMBLAY, Johanne ..... 74, 75  
 TYMES, Nathaniel ..... 102  
  
 UBARTAS, Cindy ..... 130  
  
 VAN BELLE, Gerald ..... 69  
 VANDAL, Alain C. .... 126  
 VANNUCCI, Marina ..... 92  
 VARAIYA, Pravin ..... 97  
 VIJAYAN, Nair ..... 97  
 VOLODIN, Andrei ..... 67, 142  
  
 WALIULLAH, Maria ..... 108  
 WALSH, Kayo ..... 55  
 WALTNER-TOEWS, David ..... 87  
 WANG, Hsiao-Hsuan Alice ..... 130  
 WANG, Liqun ..... 72, 101  
 WANG, Steven ..... 131  
 WANG, Xu ..... 108, 132  
 WELCH, Will ..... 77  
 WELCH, William ..... 132  
 WILLAN, Andrew ..... 134

WOODROOFE, Michael	120
WORSLEY, Keith	91
WRIGHT, Ralph	78
WU, Changbao	81
WU, Jianrong	106
WU, Jingjing	101
WU, Long Yang	66
WU, Shiyong	80
XIE, Xinlei	130
XIONG, Xiaoping	105
YADAV, Ram	128
YAN, Guohua	130
YAN, Ping	99
YE, Xiufang	130
YI, Grace	83, 138
YI, Grace Y.	151
YOU, Yong	147
YU, Edward	54
YU, Hao	86
YU, P.	57
YUAN, Zhushun	108
ZAREPOUR, Mahmoud	143
ZELEN, Marvin	98
ZHANG, Heping	106
ZHANG, Jiajia	114
ZHANG, Julie	130
ZHANG, Lingyun	71
ZHANG, Peng	125
ZHANG, Ying	114
ZHAO, Hongyu	65
ZHOU, Kehu	130
ZHU, Hongtu	73
ZHU, Qiaohao	58, 130
ZHU, Yongtao	108
ZIDEK, James	87

# Author Index (per session)

## Index des auteurs (par session)

ABARIN, Taraneh .....	07F	BULL, Shelley .....	02F
ADEDIGBA, Adebolanle .....	11D	CAMPBELL, David .....	12F
ADEWALE, Adeniyi .....	11B	CAMPOS-BONILLA, M. Begoña .....	04F
AHMED, Rashid .....	11F	CANTY, Angelo .....	04F, 04G
AHMED, S. Ejaz .....	12C	CAO, Jiguo .....	12F
ALLEN, Brian .....	07G	CARILLO, Ivan .....	11B
ALTMAN, Rachel .....	02A	CARLIN, Brad .....	04D
ALVAREZ, Juan .....	02G	CARRIÈRE, Keumhee Cho .....	10G
ANDREEV, Kirill .....	04F	CARROLL, Raymond .....	04C
ANGERS, Jean-François .....	03F	CASANOVA-GURRERA, Amparo .....	04F
ARMSTRONG, Barbara .....	07F	CAUCHON, Caroline .....	07F
AUGER, Sylvie .....	03E	CHANG, Edward .....	04G
BAILEY, Carmen .....	04G	CHAPMAN, Judy-Anne .....	0
BALAN, Raluca .....	02B	CHARLEBOIS, Joanne .....	11B
BANERJEE, Sudipto .....	04D	CHARRON, Dominique .....	04F, 07G
BASINGER, Jim .....	01A	CHAUBEY, Yogendra .....	12G
BEAUMONT, Jean-François .....	12E	CHEN, Bingshu .....	10B
BELLHOUSE, David .....	02C	CHEN, Gemai .....	03B
BÉRARD-CHAGNON, Julien .....	11B	CHEN, Jiahua .....	03D
BERTHELOT, Jean-Marie .....	02E	CHENOURI, Shojaeddin .....	06G
BIERNACKA, Joanna .....	02F	CHIPMAN, Hugh .....	11D
BINDER, David .....	12D	CHOI, Yun Hee .....	12G
BINGHAM, Derek .....	12B	CHU, Christina Chor-Yan .....	11B
BISCHOFF, Wolfgang .....	10C	CHU, Ka Lok .....	02G
BLACKWELL, Debra L. ....	02E	CHU, Rong .....	11E
BLUMBERG, Stephen .....	02E	COLEMAN, Kimberley .....	07E
BOCCI, Cynthia .....	12E	CONNOLLY, Stuart .....	11E
BOTTAI, Matteo .....	07G	COOK, Andrea .....	04D
BOUHADDIoui, Chafik .....	04E	COOK, Richard .....	04C, 10B, 11F
BOURBEAU, Robert .....	04F	COONER, Freda .....	04D
BOYETT, James .....	07C	COOPER, Michael .....	06G
BRASHER, Penny .....	10E	COPE, Eric .....	06D
BROWMAN, George .....	11E	COREY, Paul .....	10E
BROWN, Elizabeth .....	02A	CRAIU, Radu .....	10B
BROWN, Steve .....	11F	CUDDY, Kurl .....	11E
BRUNSWIG, Hadassa .....	11B	CYR, Andre .....	12D
BRYAN, Jennifer .....	11C	DABROWSKI, Andre .....	12C
BUITENDYK, Melanie .....	10F		

- DAS, Kalyan ..... 10F  
 DAVIES, Alexander ..... 12D  
 DAVIS, Karen ..... 02E  
 DAVIS, Richard ..... 06A  
 DEAN, Charmaine ..... 11F  
 DESGAGNÉ, Alain ..... 03F  
 DICESARE, Joe ..... 06F  
 DILLON, Jo-Anne ..... 01A  
 DINU, Irina ..... 11B  
 DOCHITOIU, Catalin ..... 12E  
 DOLOVICH, Lisa ..... 04F  
 DORIAN, Paul ..... 11E  
 DOUKETIS, James ..... 11E  
 DRTON, Mathias ..... 10D  
 DUCHESNE, Thierry ..... 10B  
 DUFOUR, Jean-Marie ..... 04E  
 DUPUIS, Debbie ..... 12F  
  
 EARL, Lawrence ..... 06D  
 EDWARDS, Betty ..... 0  
 EINMAHL, Uwe ..... 07B  
 EL-SAIDI, Mohammed ..... 06G  
 ESCARELA, Gabriel ..... 11E  
 EVANS, Michael ..... 03F  
  
 FAN, Zizhong ..... 10A  
 FANG, Miao ..... 0  
 FENG, Rui ..... 07D  
 FERKO, Nicole ..... 10F  
 FLEURY, Manon ..... 07G  
 FOTOUHI, Ali Reza ..... 10G  
 FRASER, Don ..... 02D  
 FREDETTE, Marc ..... 11F  
 FU, Rao ..... 11B  
 FUSSMANN, Gregor ..... 12F  
  
 GAMITO, Eduard ..... 0  
 GENT, Michael ..... 11E  
 GENTLEMAN, Jane ..... 02E  
 GERACI, Marco ..... 07G  
 GHOSH, Sunita ..... 10G  
 GIDDINGS, Bethany ..... 11F  
 GILLIS, Daniel ..... 04F  
 GIRARD, Claude ..... 12D  
 GNEITING, Tilmann ..... 04E  
 GRUBISIC, Maja ..... 07E  
 GUAY, Mireille ..... 0  
 GULSHAN, Jahida ..... 12G  
 GUO, Meijie ..... 07E  
 GUO, Pengfei ..... 0  
  
 GUO, Xiaohui (Helen) ..... 11E  
 GUSTAFSON, Paul ..... 03C, 06A  
 GUT, Allan ..... 07B  
  
 HAN, Jing ..... 07D  
 HASHEMI, Seyed ..... 02B  
 HE, Wenqing ..... 11C  
 HELBING, Caren ..... 04G  
 HEUNIS, Andrew ..... 02B  
 HODSON, Ian ..... 11E  
 HOETING, Jennifer ..... 06A  
 HOHNLOSER, Stefan ..... 11E  
 HOLBROOK, Anne ..... 04F  
 HOLT, John ..... 04F, 07G  
 HOOKER, Giles ..... 12F  
 HOOPER, Peter ..... 0  
 HORROCKS, Julie ..... 07G  
 HORVATH, Steve ..... 06C  
 HOWARD, Allan ..... 04A  
 HSIAO, Cheng ..... 03C  
 HUANG, Hui ..... 06E  
 HUANG, Jian ..... 06E  
 HYNDMAN, Cody ..... 11D  
  
 ILHAMTO, Gatot ..... 07E  
 ISHWARAN, Hemant ..... 12C  
 ISMAILA, Afisi ..... 04F, 10F  
 ISOTALO, Jarkko ..... 02G  
 IVANOFF, Gail ..... 12C  
  
 JAMES, L.F. ..... 12C  
 JEMIAI, Yannis ..... 04C  
 JIANG, Hedy ..... 04F  
 JOHNSTON, Karissa ..... 04G  
 JONES, Bruce ..... 12F  
 JONES, Richard ..... 0  
 JUDKINS, David ..... 10A  
  
 KARUNAMUNI, Rohana ..... 06G  
 KARUNANAYAKE, Chandima ..... 0  
 KATZOFF, Myron ..... 07F  
 KAWCZAK, Janusz ..... 02G  
 KELLY, Elizabeth ..... 12B  
 KELLY, I.W. ..... 0  
 KEYFITZ, Barbara ..... 01B  
 KHALILI, Abbas ..... 03D  
 KHAN, Bashir ..... 06G  
 KHARE, Meena ..... 02E  
 KOHLI, Michele ..... 10F  
 KORU-SENGUL, Tulay ..... 07G

- KOUIDER, Elies .....06G  
 KOVACEVIC, Milorad ..... 06B, 12D  
 KOWGIER, Matthew ..... 11B  
 KOZAK, Robert .....10A  
 KRENZKE, Thomas ..... 10A  
 KREWSKI, Daniel ..... 12G  
 KULIK, Rafal ..... 02G, 12C  
 KUSTRA, Rafal ..... 06C  
 KWON, Jaimyoung ..... 06D  
  
 LAROCHE, Richard ..... 07F  
 LAROCHE, Sylvie ..... 03E, 07F  
 LATOUCHE, Michel ..... 03E  
 LAVERTY, William H. .... 0  
 LAWLESS, Jerry .....07A  
 LE, Lisa Wen ..... 11B  
 LE, Nhu ..... 04F  
 LEBLANC, Alexandre .....06G  
 LEBLANC, Nancy ..... 11B  
 LEE, Sandra ..... 06E  
 LEFEBVRE, Geneviève ..... 10F  
 LELE, Subhash ..... 12D  
 LESOSKY, Maia .....07E  
 LESPERANCE, Mary ..... 04G  
 LETAC, Gérard ..... 10D  
 LI, Deli ..... 07B  
 LI, Yi ..... 04D  
 LI, Zhengqing ..... 07C  
 LIANG, Yuanyuan ..... 10G, 11B  
 LIN, Jing ..... 0  
 LIN, Lillian ..... 07F  
 LIN, Shili ..... 02F  
 LIN, Xihong ..... 04C  
 LINDSAY, Bruce ..... 07D  
 LINKLETTER, Crystal ..... 12B  
 LISAWADI, Supranee ..... 02G  
 LIU, Jixin ..... 11B  
 LIU, Mengling ..... 07D  
 LIU, Wanrong ..... 12G  
 LIU, Zhang ..... 11B  
 LIU, Zhenqiu ..... 02F  
 LIU, Zhong ..... 04F  
 LIX, Lisa ..... 10F  
 LOHR, Sharon ..... 04B  
 LOU, Wendy ..... 04F  
 LU, Irene ..... 06B  
 LU, Pingguo ..... 04E  
 LU, Wenbin ..... 07D  
 LU, Xuewen ..... 12G  
  
 MA, Renjun ..... 04D  
 MA, Shuangge ..... 06E  
 MAAROUF, Abdel ..... 04F, 07G  
 MACKAY, Jock ..... 03B, 11D  
 MACKINNON, Peter ..... 01A  
 MACNEILL, Ian ..... 04E  
 MANTEL, Harold ..... 06B, 12E  
 MARCHAND, Éric ..... 03F  
 MARTINEZ, Michael ..... 02E  
 MASSAM, Hélène ..... 10D  
 MATTHEWS, David E. .... 12G  
 McDONALD, Paul ..... 11F  
 McLEISH, Don ..... 06F  
 McLEOD, Ian ..... 07G  
 MELNICHOUK, Olga ..... 07E  
 MERTON, Andrew ..... 06A  
 MICHAILEDIS, George ..... 06D, 12B  
 MICHAUD, Isabelle ..... 10B  
 MIGLIORETTI, Diana ..... 02A  
 MIKET, M. .... 0  
 MIREA, Lucia ..... 11B  
 MONSOUR, Michael ..... 07F  
 MOSESOVA, Sofia ..... 07E, 11D  
 MOSHONOV, Hadas ..... 03F  
 MOUIHA, Aberazzak ..... 04E  
 MYLES, Jamie ..... 02C  
  
 NAHM, Erika ..... 07E  
 NAIR, Kalpana ..... 04F  
 NATHOO, Farouk ..... 11F  
 NG, Peggy ..... 11B  
 NOURIJELYANI, Keramat ..... 0  
  
 O'DONNELL, Colin ..... 0  
 OGDEN, Sheryl ..... 0  
 OH, Hee-Seok ..... 04A  
 O'HARA HINES, Jeanette ..... 02C  
 OLIVAS, David ..... 12B  
 OVERDUIN, Steve ..... 12G  
 OYET, Alwell ..... 0, 12G  
  
 PAHWA, Punam ..... 10G  
 PAKYARI, Reza ..... 06G  
 PANTOJA GALICIA, Norberto ..... 06B  
 PENG, Yingwei ..... 07G  
 PENZO-GIACCA, Wilma ..... 04F  
 PERRON, François ..... 10C  
 PETERS, Amy ..... 07E  
 PETKAU, John ..... 02C  
 PHILIPS, Robert ..... 12E

- PICKA, Jeffrey ..... 12F  
 PINTO, Eleanor ..... 11E  
 PRASAD, Narasimha ..... 0, 12D  
 PRETZEL, Shannon ..... 0  
 PRINGLE-NELSON, C. .... 0  
 PROVOST, Serge ..... 04E  
 PUNTANEN, Simo ..... 02G  
  
 QU, Annie ..... 10F  
  
 RAFTERY, Adrian ..... 04E  
 RAIS, Nourddine ..... 04E  
 RAIZENNE, Mark ..... 0  
 RAMSAY, James ..... 12F  
 RAMSAY, Timothy ..... 12G  
 RANJAN, Pritam ..... 12B  
 RAO, J.N.K. .... 04B, 06B, 10F, 12E  
 RAY, Surajit ..... 07D  
 REEDMAN, Laurie ..... 07F  
 REINKER, Stefan ..... 02A  
 REN, Qunshu ..... 06B, 10F  
 RICHARDSON, Thomas ..... 10D  
 ROBERTS, Georgia ..... 06B, 12D  
 ROBERTS, Robin ..... 11E  
 ROSS, Nancy ..... 02E  
 ROSYCHUK, Rhonda ..... 02A  
 ROUSSEAU, Judith ..... 02D  
 ROYCE, Melanie ..... 0  
 RYAN, Nancy ..... 10F  
  
 SALEH, A.K. Md. E ..... 06G  
 SALOPEK, Donna ..... 06F  
 SANJEL, Deepak ..... 04E  
 SCHAFER, Joseph ..... 10A  
 SCHAUBEL, Doulgas ..... 10B  
 SCHENNACH, Susanne ..... 03C  
 SCHWARZ, Carl ..... 10E  
 SEN, Arusharka ..... 12G  
 SEN, Pranab K. .... 12G  
 SENGUL, Haydar ..... 04G  
 SEVERINI, Tom ..... 02D  
 SHAO, Yongzhao ..... 07D  
 SHEN, Hui ..... 03F, 11B  
 SHEN, Samuel ..... 04A  
 SHIN, Hwashin ..... 12G  
 SINGH, Avi ..... 12E  
 SINGH, Avinash ..... 04B, 12E  
 SINGH, Padam ..... 10G  
 SITTER, Randy ..... 11A, 12B  
 SONG, Peter ..... 10F  
  
 SOTEROS, Christine ..... 0, 02G  
 STADTMUELLER, Ulrich ..... 07B  
 STAUDENMAYER, John ..... 02A  
 STEELE, Russell ..... 10F  
 STEINER, Stefan ..... 03B, 11D  
 STEPHENS, Michael ..... 12G  
 STEWART, Travis ..... 12F  
 STOFFER, David S. .... 07G  
 STRAWDERMAN, William E. .... 03F  
 STUDENY, Milan ..... 10D  
 STYAN, George P. H. .... 02G  
 SU, Wanhua ..... 11B  
 SUN, Jiayang ..... 03D  
 SUN, Lei ..... 02F  
 SUN, Ning ..... 02F  
 SUN, Ye ..... 11B  
 SZAFRON, Michael ..... 0  
  
 TAI, Patricia ..... 0  
 TAN, Ming ..... 07C  
 TANG, Boxin ..... 12B  
 THABANE, Lehana ..... 04F, 11E  
 THEPHAMONGKHOL, Kullathorn ..... 11E  
 THIESSEN, Heather ..... 02C  
 THOMPSON, Mary ..... 06B  
 THOMPSON, Sandra ..... 06A  
 THOMPSON, Steven ..... 07F  
 THORPE, Kevin ..... 11E  
 TORABI, Mahmoud ..... 12E  
 TREMBLAY, Johanne ..... 03E  
 TYMES, Nathaniel ..... 06G  
  
 UBARTAS, Cindy ..... 11B  
  
 VAN BELLE, Gerald ..... 03A  
 VANDAL, Alain C. .... 10F  
 VANNUCCI, Marina ..... 06A  
 VARAIYA, Pravin ..... 06D  
 VIJAYAN, Nair ..... 06D  
 VOLODIN, Andrei ..... 02G, 12C  
  
 WALIULLAH, Maria ..... 07E  
 WALSH, Kayo ..... 0  
 WALTNER-TOEWS, David ..... 04F  
 WANG, Hsiao-Hsuan Alice ..... 11B  
 WANG, Liqun ..... 03C, 06G  
 WANG, Steven ..... 11C  
 WANG, Xu ..... 07E, 11D  
 WELCH, Will ..... 03F  
 WELCH, William ..... 11D

WILLAN, Andrew	11E
WOODROOFE, Michael	10C
WORSLEY, Keith	05
WRIGHT, Ralph	04A
WU, Changbao	04B
WU, Jianrong	07C
WU, Jingjing	06G
WU, Long Yang	02F
WU, Shiyong	04B
XIE, Xinlei	11B
XIONG, Xiaoping	07C
YADAV, Ram	10G
YAN, Guohua	11B
YAN, Ping	06E
YE, Xiufang	11B
YI, Grace	04C, 11F, 12G
YOU, Yong	12E
YU, Edward	0
YU, Hao	04E
YU, P.	0
YUAN, Zhushun	07E
ZAREPOUR, Mahmoud	12C
ZELLEN, Marvin	06E
ZHANG, Jiajia	07G
ZHANG, Julie	11B
ZHANG, Lingyun	03B
ZHANG, Peng	10F
ZHANG, Ying	07G
ZHANG, Heping	03D, 07D
ZHAO, Hongyu	02F
ZHOU, Kehu	11B
ZHU, Qiaohao	0
ZHU, Hongtu	03D
ZHU, Qiaohao	11B
ZHU, Yongtao	07E
ZIDEK, James	04F