

Welcome! Bienvenue!

The Inaugural SSC Student Conference / Congrès inaugural étudiant de la SSC

Saturday, May 25th, 2013 / Samedi, 25 mai 2013

University of Alberta, Edmonton

This student-led conference was designed for undergraduate and graduate students to develop their interests in research and careers in statistics. The program includes workshops and contributed sessions on professional skill development, careers in industry and academia, and research presentations.

The keynote speaker, **Dr. Jeffrey Rosenthal** of the Department of Statistics at the University of Toronto, is a leading Canadian researcher, teacher, author, and engaging speaker.

Ce congrès dirigé par des étudiants a été conçu pour les étudiants de premier cycle et des cycles supérieurs pour développer leurs intérêts dans la recherche et les carrières en statistique. Le programme comprend des ateliers et séances de communication sur le développement de compétences professionnelles, les carrières en industrie et en milieu universitaire, et des présentations de recherche.

Dr. Jeffrey Rosenthal du département de statistique de l'Université de Toronto, un chercheur, professeur, auteur et présentateur hors pair au Canada prononcera le discours d'honneur.



Statistical Society of Canada
Société statistique du Canada



Biostatistics



CRC Press
Taylor & Francis Group



Program Overview, May 25th / Horaire du programme, le 25 mai

CAB = Central Academic Building

Time / Heure	Activity / Activité	Chair	Location / Lieu
11:00am – 11:45am	Registration / Inscription		CAB 243
11:45am – 12:15pm	Welcome / Bienvenue <i>Christian Léger, SSC president / président du SSC</i> <i>Brief presentations / Courtes présentations</i>	Mireille Schnitzer Feng He	CAB 243
12:15pm– 1:00pm	Boxed lunch (provided) / Boîte à lunch (fournie par le congrès)		Hallway / Corridor
1:00pm – 2:00pm	Career speakers / Conférence sur l'orientation professionnelle <i>Ori Stitelman, Senior Data Scientist / Scientifique principale</i> <i>Matthias Schonlau, Professor / Professeur, University of Waterloo</i> <i>Kathryn Mills, Manager at Canada Border Services / Gestionnaire à l'Agence des services frontaliers du Canada</i>	Liqun Diao Feng He Min Chen	Ori: CAB 239 Matthias: CAB 273 Kathryn: CAB 281
2:00pm – 3:30pm	Skills session <i>Workshop / Atelier: CV/Resume</i> <i>Graduate Student Panel / Table ronde</i>	Zihui (Amy) Liu	Workshop / Atelier: CAB 239 Panel / Table ronde: CAB 273
3:30pm – 4:15pm	Poster session and snack break / Séance d'affiches et pause collation		CAB 269
4:15pm – 5:50pm	Contributed oral presentations / Présentations orales	Nathaniel Payne Ying Yan Mahdis Azadbakhsh	Session 1: CAB 239 Session 2: CAB 273 Session 3: CAB 281
5:50pm – 6:50pm	Keynote address by Jeffrey Rosenthal, Discours d'honneur	Mireille Schnitzer	CAB 243
6:50pm	Closing remarks and award presentations Discours de clôture et présentation des prix	Mireille Schnitzer Mahdis Azadbakhsh	CAB 243
7:45pm	Social evening (at the participant's expense) / Sortie sociale (souper non inclus dans les frais du congrès)		Hudsons Canadian Tap House, 11113 87 Ave NW

Organizers and Volunteers / Organisateur(e)s et bénévoles

Student Committee / Comité étudiant :

Chair / Présidente:

Mireille Schnitzer (Harvard School of Public Health)

Session organizers and committee members / Organisateur(e)s des sessions et membres du comité:

Mahdis Azadbakhsh (York University), Feng He (University of Waterloo),

Amy Liu (McGill University), Nathaniel Payne (Simon Fraser University),

Ying Yan (University of Waterloo)

Faculty Advisory Committee / Comité consultatif des membres du corps professoral :

Co-chairs / Co-présidentes:

Karen Kopciuk (University of Calgary), Lisa Lix (University of Manitoba)

Committee members / Membres du comité:

Cindy Xin Feng (University of Saskatchewan), Jinko Graham (Simon Fraser University),

Linglong Kong (University of Alberta), Wendy Lou (University of Toronto, ex officio),

Georges Monette (York University)

Other Volunteerism / Autre bénévolat :

Session chairs / Présidents des sessions: Min Chen (University of Waterloo), Liqun Diao (University of Waterloo)

Local support / Soutien local: Qian Shi (University of Alberta), Dengdeng Yu (University of Alberta)

Translations / Traductions: Pierre-Jérôme Bergeron (University of Ottawa), Philippe Chaput (Champlain College), Gary Parker (Simon Fraser University), Dany Plouffe (McGill University), Nathalie Saint-Pierre (Institut de recherches cliniques de Montréal), Joseph François Tagne (Université de Montréal), Denis Talbot (Université Laval), Luc Villandré (McGill University)

Program booklet design / Conception du livret de programme: Tanya Tolokh

Career Session / Conférence sur l'orientation professionnelle

1:00-2:00pm

The career session involves three career speakers working in different sectors (industry/business, academics and government) who will give three parallel in-depth presentations. The three presentations will run concurrently, so *please choose which presentation you would like to attend*.

La séance sur l'orientation professionnelle implique trois conférenciers travaillant dans des différents secteurs (industriel, gouvernemental et académique) qui vont donner trois présentations parallèles approfondies. Les trois présentations vont se dérouler simultanément, donc *SVP choisir quelle présentation à quelle vous voulez assister*.

Ori Stitelman, PhD

Senior Data Scientist, media6degrees / Analyste de données principal à media6degrees

Location: CAB 239

Chair / Présidente: Liqun Diao (University of Waterloo)

From Academic Biostatistician to Applied Data Scientist

After receiving my PhD in Biostatistics at UC Berkeley, developing causal methods for estimating the effects of drug therapies, I was unsure as to the career path I wanted to pursue. One thing was for sure: the skills that I developed while at graduate school were pervasive and could add value anywhere. After looking at many options I decided to take my first job as a Data Scientist at media6degrees, a targeted display advertising technology company. All that really means is that we built machines that determined the best individuals to show banner advertisements to on the internet on behalf of hundreds of clients. Those machines had to make billions of decisions daily in the milliseconds that websites load across the world. Given these constraints it was imperative that we had a massive scale machine learning system to inform those decisions. Just two months ago, after two years at media6degrees, I moved on to manage a statistical team at Wells Fargo. Though I have just started we are trying to answer some interesting questions about the effectiveness of our overall advertising spend. I would love to share with you my experiences at media6degrees and my rather limited experience at Wells Fargo and how I made the transition from developing causal methods to prototyping and building massive machine learning systems that make billions of advertising decisions daily to managing a team of statisticians at a bank.

De biostatisticien en milieu universitaire à analyste de données appliquées

Après avoir reçu mon doctorat en biostatistique à l'Université de la Californie à Berkeley, où je développais des méthodes causales pour l'estimation des effets de médicaments, je n'étais pas certain quel cheminement de carrière je voulais poursuivre. Une chose était sûre : les compétences que j'ai développées durant les cycles supérieurs étaient omniprésentes et pourraient apporter une valeur ajoutée n'importe où. Après avoir examiné plusieurs options, j'ai obtenu mon premier emploi comme analyste de données à media6degrees, une entreprise en technologie d'affichage publicitaire ciblée. Cela veut dire qu'on construisait des machines qui déterminent les meilleurs individus à qui montrer des

bannières de publicités sur l'internet pour le compte de centaines de clients. Ces machines doivent faire des milliards de décisions chaque jour dans les millisecondes qu'il faut pour télécharger un site Web à travers le monde. Étant données les contraintes, il fallait absolument avoir un énorme système d'apprentissage par la machine pour évaluer ces décisions. Il y a seulement deux mois, après deux années passées à media6degrees, je suis devenu gérant d'une équipe statistique à Wells Fargo. Bien que je vienne tout juste de commencer, nous essayons de répondre quelques questions intéressantes sur l'efficacité de nos dépenses publicitaires globales. J'aimerais bien partager mes expériences à media6degrees et mon expérience plutôt limitée à Wells Fargo et comment j'ai fait la transition du développement de méthodes causales au prototypage et à la construction des systèmes d'apprentissage automatique massifs qui font quotidiennement des milliards de décisions en matière de publicité, puis à la gestion d'une équipe de statisticiens dans une banque.

Matthias Schonlau, Ph.D

**Professor / Professeur, Department of Statistics and Actuarial Sciences
University of Waterloo**

Location / lieu : CAB 273

Chair / Président: Feng He (University of Waterloo)

From Industry to Academia: a non-traditional path

After obtaining my Ph.D. in Canada, I spent two years as a postdoctoral fellow at AT&T Labs and the National Institute of Statistical Sciences. I was then recruited to work for the RAND Corporation, a public policy think tank. I spent 6 years in RAND's Los Angeles headquarters followed by another 6 years at RAND's Pittsburgh office. Towards the end my employment with RAND I took a one year leave of absence to work at the German Institute for Economic Analysis in Berlin. Since 2011 I have been a professor at the University of Waterloo. My interests have been strongly affected by my industry career: As a graduate student I thought of survey methodology as boring. After learning about the many practical challenges in fielding and analyzing surveys, survey methodology is now my main area of interest.

De l'industrie au milieu universitaire : un parcours non-traditionnel

Après avoir obtenu mon doctorat au Canada, j'ai passé deux années comme chercheur postdoctoral aux laboratoires AT&T et le *National Institute of Statistical Sciences*. J'ai été ensuite recruté par la RAND Corporation, un *think tank* de politiques publiques. J'ai passé six années au siège de RAND à Los Angeles, suivi par un autre six années au bureau de RAND à Pittsburgh. Vers la fin de mon emploi avec RAND, j'ai pris une année d'absence pour travailler à l'institut allemand pour l'analyse économique à Berlin. Depuis 2011, je suis un professeur à l'Université de Waterloo. Mes intérêts ont été fortement influencés par ma carrière en industrie. Comme étudiant aux cycles supérieurs, je pensais que l'étude des méthodes de sondage était ennuyeuse. Mais après avoir appris plusieurs défis pratiques pour la mise en place et l'analyse des sondages, l'étude des méthodes de sondage est maintenant mon principal domaine d'intérêt.

Kathryn Mills

Manager, Mathematics and Data Exploration at Canada Border Services Agency / Gestionnaire, Section des mathématiques et exploration des données à l'Agence des services frontaliers du Canada

Location / lieu : CAB 281

Chair: Min Chen (University of Waterloo)

From Cryptography to Big Data: A Career Path in Government Now

Starting her full-time government career with a branch of the Department of National Defence, completing an international posting to a US Agency, and then transitioning to the Canada Border Services Agency, Kathryn has combined her Engineering and Statistics backgrounds to support national security and public safety. She currently manages the Mathematics and Data Exploration Section at the Canada Border Services Agency. The group is tasked with providing advanced mathematical and statistical analysis in support of securing our nation's borders. From risk assessment to resource allocation and optimization, the Section handles a wide range of Agency analyses involving a variety of data.

Her career path began back in high school when applying to universities and deciding on Engineering versus Biology, a decision that was actually meant to leave doors open should she decide not to pursue a career in medicine. It was further shaped by her undergraduate studies in Computer Systems Engineering, which she switched to Electrical Engineering, and her industry co-op placements which gave her glimpses into the real world of high-tech. Her path was then molded by her graduate studies in Mathematics (Specialization in Probability and Statistics) and her government internship which showed her a different, yet equally exciting, world to that of high-tech and industry. Each experience not only helped guide her in what she liked, but also in what she did not like doing and why. Kathryn has designed software and electronic circuitry in products used by top telecommunications providers worldwide, performed statistical analyses on the Register of Electors and the Census, worked in the protection of electronic information and communication, graduated from a three-year specialized mathematics intern training program in the US, and currently manages a section tasked with providing advanced analysis in support of securing our nation's borders. She looks forward to sharing her experiences and insights on a government career path now.

De la cryptographie aux données volumineuses : un choix de carrière au gouvernement

Kathryn a entamé sa carrière à plein temps au gouvernement au sein d'une direction générale du ministère de la Défense nationale, après quoi elle a été affectée à une mission internationale au sein d'un organisme des États-Unis. Elle s'est ensuite jointe à l'Agence des services frontaliers du Canada (ASFC). Elle a ainsi combiné son bagage en génie et en statistiques en vue de contribuer à la sécurité nationale et publique. En ce moment, Kathryn dirige la Section des mathématiques et de l'exploration de données de l'ASFC. Cette section est chargée de procéder à des analyses mathématiques et statistiques poussées afin de contribuer à la protection de la frontière du pays. Que ce soit dans le domaine de l'évaluation des risques ou celui de l'affectation et de l'optimisation des ressources, la Section gère un vaste éventail d'analyses de l'ASFC qui englobent diverses données.

Le cheminement de carrière de Kathryn a commencé à se dessiner à l'école secondaire. Kathryn a alors présenté des demandes d'admission à des universités et décidé d'étudier en génie plutôt qu'en biologie. C'est une décision qui, en fait, devait lui ouvrir la voie dans l'éventualité où elle déciderait de ne pas faire carrière en médecine. Kathryn a façonné davantage sa carrière lors de ses études de premier cycle en génie informatique, après quoi elle a décidé de se tourner vers le génie électrique. Elle a ensuite fait,

dans le cadre d'un programme coop, un stage en industrie, qui lui a donné un aperçu du monde réel de la haute technologie. La voie de Kathryn a ensuite pris son sens lors de ses études de cycle supérieur en mathématiques (spécialisation en probabilité et en statistiques) et de son stage au gouvernement. Ce dernier lui a permis de découvrir un univers différent de celui de la haute technologie et de l'industrie, mais tout aussi stimulant. Chaque volet de l'expérience acquise par Kathryn l'a aidée à savoir ce qu'elle aimait, mais aussi ce qu'elle n'aimait pas, en plus de lui permettre de comprendre pourquoi elle n'y trouvait pas d'intérêt. Kathryn a conçu des logiciels et des circuits électroniques pour des produits utilisés par les principaux fournisseurs mondiaux de services de télécommunications. Elle a également réalisé des analyses statistiques du Registre national des électeurs et du Recensement, et a travaillé à la protection de l'information et des communications électroniques. En outre, elle a réussi un programme de formation et de stages de trois ans, spécialisé en mathématiques aux États-Unis, et elle dirige actuellement une section chargée d'assurer une analyse poussée aux fins de la protection de la frontière du pays. Kathryn se réjouit à l'idée de faire connaître son expérience et ses observations quant à son choix de carrière au gouvernement.

Skills Session / Session sur les compétences professionnelles

2:00-3:30pm

Participants may choose to attend either a CV/resume building workshop or a student panel discussing the challenges of graduate school.

Les participants pourront choisir d'aller soit à un atelier sur la préparation de curriculum vitae ou à une table ronde à propos des défis des cycles supérieurs.

Workshop / Atelier: 4 Pillars of Effective CVs and Resumes / 4 piliers des CV efficaces pour l'industrie et au milieu universitaire

Presenter: Michael Bodnar, University of Alberta Career Centre

Location / lieu : CAB 239

A well written, up-to-date and accurate resume/CV is an invaluable tool at the core of a successful job search. It communicates your qualifications by summarizing your experience, education and skills, with the goal of eliciting an interview invitation. This workshop is offered by the University of Alberta Career and Placement Services (CaPS). By attending it, you will learn how to describe and relate your experiences and skills to the jobs for which you are applying, and you will learn about the content, wording, format and presentation of effective CVs/resumes.

Un CV bien écrit, mise à jour et précis est un outil indispensable au cœur d'une recherche d'emploi réussie. Le CV décrit vos qualifications en résumant votre expérience, votre éducation et vos compétences, avec le but d'obtenir une invitation pour une entrevue. Cet atelier est offert par le Centre de carrière et de services de placement de l'Université de l'Alberta (CaPS). En participant à cet atelier, vous allez apprendre à décrire et relier vos expériences et compétences pour les positions auxquelles vous postulez, et vous allez apprendre quels sont le contenu, le texte, la mise en page et la présentation d'un CV efficace.

Panel / Table ronde: ***How to Survive Grad School / Comment survivre aux études des cycles supérieurs***

MC:

Zhihui (Amy) Liu, PhD student / étudiante de 3^e cycle (Biostat), McGill University

Panelists / Panélistes:

Jonathan Lee, PhD student / étudiant de 3^e cycle (Stat), University of Western Ontario;

Tim Guimond, PhD student / étudiant de 3^e cycle (Biostat), University of Toronto;

Zheng Sun, PhD student / étudiant de 3^e cycle (Stat), Simon Fraser University;

Oksana Chkrebti, PhD student / étudiante de 3^e cycle (Stat), Simon Fraser University.

Location / lieu: CAB 273

This Q&A panel consists of five PhD students in Statistics and Biostatistics from four Canadian universities. The panelists will share their own experiences of the challenges and successes during their student careers. Topics include applying for graduate schools, choosing a supervisor/thesis topic/committee, dissertation writing, scientific publishing, and time management. This session is intended for undergraduate students who are considering graduate studies and junior graduate students.

Ce panel de questions et réponses comprend cinq étudiants de troisième cycle en statistique et en biostatistique de quatre universités Canadiennes. Les panélistes partageront leurs propres expériences de défis et succès durant leurs études aux cycles supérieurs. Les sujets traités comprennent les demandes d'admission pour les écoles supérieures, le choix d'un superviseur, le sujet de thèse et le comité de défense, l'écriture de la thèse, la publication d'articles scientifiques, et la gestion du temps. Cette session est destinée aux étudiants de premier cycle qui considèrent entreprendre des études supérieures et aux étudiants juniors aux cycles supérieurs.

Contributed Poster Session

Séance de présentations par affiche

3:30-4:15pm

Location / lieu: CAB 269

* = eligible for the undergraduate poster presentation award / admissible pour le prix de présentation pour les étudiants de premier cycle

Local-EM and Mismeasured Data* / *EM local et données incorrectement mesurées**

Shiva Ashta ^c, Patrick Brown ^{a, b, c}, Paul Nguyen ^{a, b}, Jamie Stafford ^{b, c}

^a Prevention and Cancer Control, Cancer Care Ontario

^b Dalla Lana School of Public Health, University of Toronto

^c Department of Statistics, University of Toronto

In many spatial-temporal studies, the observations of interest cannot be directly measured. This could, in part, be due to the observation's characteristics, operating costs, or confidentiality issues. For instance, an individual may test negative after contracting an infectious disease with the actual disease being detected much later. The measured time of infection is then measured with error due to the lag in a true positive diagnosis. To address this measurement error, research was conducted to implement a local-EM algorithm on mismeasured data, and an improvement over the existing standard set by the deconvoluting kernel density estimator was found.

Dans plusieurs études spatio-temporelles, des mesures importantes ne peuvent être collectées directement en raison, entre autres, de leurs caractéristiques, de coûts opérationnels ou de protection de la confidentialité. Par exemple, malgré un résultat négatif à un test de détection d'une maladie infectieuse, on peut parfois la diagnostiquer beaucoup plus tard. Il y a donc une certaine dose d'erreur dans le temps d'infection mesuré due au délai dans l'obtention d'un diagnostic positif. Afin de pouvoir prendre en compte ce type d'erreur de mesure, une étude a été menée afin de mettre en oeuvre un algorithme EM local sur des données avec erreur de mesure. Cet algorithme est supérieur au standard existant de l'estimateur *deconvoluting kernel density*.

Investigating knot transition probabilities after strand passage on self-avoiding polygons on the cubic lattice* / *Une étude portant sur les probabilités de transition de noeuds suite au passage de brins sur des polygones auto-évitants sur le treillis cubique**

Marla Cheston (presenter)^a, Kevin McGregor (presenter)^a, Christine Soteris^a, Michael Szafron^b

^a Department of Mathematics and Statistics, University of Saskatchewan

^b School of Public Health, University of Saskatchewan

We computationally model the effects of a type II topoisomerase strand passage on DNA. In doing so, we generate self-avoiding polygons on the simple cubic lattice through the use of composite Markov chain computer simulations. We investigate specific strand passage structures, the theta-structure and the symmetric structure, to compare their limiting knot transition probabilities. We also show evidence that the probability of going from knot type K to K composed with K' is independent of the initial knot type K .

Nous modélisons par informatique les effets du passage d'un brin topoisomérase de type II sur l'ADN. Ce faisant, nous générons des polygones *auto-évitants* sur le treillis cubique simple à l'aide de simulations de chaînes de Markov composites. Nous examinons des structures spécifiques de passages de brins, la structure theta et la structure symétrique, afin de comparer leur distribution limite de transition de noeuds. Nous illustrons également que la probabilité d'une transition d'un noeud de type K à un noeud de type K composé avec K' est indépendante du type initial de noeud K .

Imputing from ordinal data in unevenly spaced category bounds / *Imputation à partir de données ordinales dans des catégories bornées par des frontières non-équidistantes*

Jack Davis

Department of Statistics & Actuarial Science, Simon Fraser University

Numeric data is sometimes recorded as ordinal data. This poster outlines a method estimating numeric values from ordinal data via a distribution from which i.i.d. random variables would feasibly produce the observed category counts. The method assumes a distribution, finds parameters with maximum likelihood with respect to the empirical distribution, then takes conditional means within each category.

Existing methods either assume that category counts are independent or require boundaries between categories to be equally spaced; the proposed method does neither.

This poster includes an application in survival analysis and a study of sensitivity to distribution misspecification.

Les données numériques sont parfois enregistrées comme des données ordinales. Cette affiche résume une méthode d'estimation de valeurs numériques à partir de données ordinales par l'intermédiaire de la génération de valeurs à partir de variables aléatoires i.i.d. suivant une distribution qui produirait le décompte observé dans les différentes catégories. La méthode est basée sur une loi présumée, dont les paramètres sont estimés par la méthode du maximum de vraisemblance à partir de la fonction de répartition empirique, et implique le calcul d'une moyenne conditionnelle dans chaque catégorie. Contrairement à la méthode proposée, les méthodes existantes sont fondées sur l'hypothèse d'indépendance entre le décompte dans chaque catégorie ou exigent que les frontières entre les

catégories soient équidistantes. L'affiche présente une application en analyse de survie et une étude de la sensibilité à la mauvaise spécification du modèle.

Mathematical Modelling of the 14th century (Black Death) pandemic using R programming language* / *Modélisation mathématique de la peste noire au 14e siècle à l'aide du langage de programmation R**

Hayedeh Jahandideh-Dariuni, Ashok Krishnamurthy

Department of Mathematics, Physics and Engineering, Mount Royal University

This project explores the use of partial differential equations to model the 14th century (Black Death) pandemic. The models that we used for this study are spatial variants of the common susceptible-infectious-removed (S-I-R) compartmental model of epidemiology. The removed state includes those who have either died, have been quarantined, or have recovered from the disease and become immune. We compare the simulated state of the epidemic in 1350 to the historical reconstruction of the unknown medieval population pattern with principal historical features of the Black Death. We present preliminary results from our implementation using R programming language.

Notre projet aborde l'utilisation d'équations différentielles partielles pour la modélisation de la peste noire, au 14e siècle. Les modèles utilisés sont des variantes spatiales du modèle épidémiologique compartimentalisé Susceptible-Infectieux-Rétabli (SIR). L'état «rétabli» inclut ceux qui sont morts, qui sont en quarantaine, ou qui se sont rétablis et sont maintenant immunisés. Nous comparons l'état simulé de l'épidémie en 1350 à une reconstruction de la structure inconnue de la population médiévale, comportant les caractéristiques principales de la peste noire. Nous présentons des résultats préliminaires tirés de notre mise en oeuvre du modèle basée sur le langage de programmation R.

Latent Variable Scoring Methods for Response Shift in Longitudinal Patient-Reported Outcomes / *Des méthodes de cotation de variables latentes appliquées à des changements de réponse dans des réponses longitudinales rapportées par les patients*

Courtney Kendall ^a, Lisa Lix ^b, Juxin Liu ^a

^a Department of Mathematics and Statistics, University of Saskatchewan

^b Department of Community Health Sciences, Faculty of Medicine, University of Manitoba

Response shift (RS), a change in a person's self-evaluation of a target construct such as quality of life, is a potential form of bias in the analysis of longitudinal patient-reported outcomes (PROs). Latent variable models have been proposed to test for RS. Factor scores, which are used to estimate latent variables, could be used to produce unbiased estimates of change in longitudinal PROs. This study compares three methods to estimate factor scores and test for change at two measurement occasions under varying conditions of sample size and correlation amongst observed and latent variables. Type I error rates are reported.

Un changement de réponse (abrév. *RS*, pour *response shift*), une variation dans l'auto-évaluation d'une mesure conceptuelle telle la qualité de vie, est une forme de biais potentiel dans l'analyse de réponses longitudinales rapportées par les patients (abrév. *PRO*, pour *patient-reported outcomes*). Des modèles de variables latentes ont été proposés afin de tester le *RS*. Des cotes factorielles, employées dans l'estimation de variables latentes, pourraient être utilisées pour obtenir des estimés non-biaisés du changement dans les *PRO* longitudinaux. Notre étude compare trois méthodes visant à estimer les cotes factorielles et à tester l'existence d'un changement à deux moments de collecte des mesures, sous une variété d'hypothèses de taille d'échantillon et de corrélation entre les variables observées et latentes. Le taux d'erreur de type I est rapporté.

Local dependence in Latent Class Analysis: A Computational Example / *La dépendance locale dans l'analyse de classes latentes: un exemple numérique*

Bogdan Lataianu ^a, Cindy Xin Feng ^a, Lisa Lix ^b

^a School of Public Health, University of Saskatchewan

^b Department of Community Health Sciences, Faculty of Medicine, University of Manitoba

Latent class analysis, a method to identify unobservable categories in a population from a set of discrete variables, rests on the assumption that the observed variables are independent conditional on group membership. However, this assumption is likely to be violated in many applications, which can result in bias in group membership assignment. This research focuses on implementation of a random-effects latent class model, which is insensitive to violations of the assumption of local independence, using SAS software. The program is illustrated in a numeric example about the identification of cases of inflammatory bowel disease in mismeasured diagnostic variables.

L'analyse de classes latentes, une méthode permettant l'identification de catégories inobservables dans une population tirée d'un ensemble de variables discrètes, repose sur l'hypothèse selon laquelle les variables observées sont indépendantes, sachant l'appartenance aux groupes. Cependant, il est probable que cette hypothèse soit invalide dans plusieurs applications, engendrant potentiellement un biais dans l'assignation aux groupes. Notre étude se concentre sur la mise en oeuvre en SAS d'un modèle de classes latentes avec effets aléatoires, insensible à l'invalidité de l'hypothèse d'indépendance locale. Son application est illustrée par un exemple numérique concernant l'identification de cas de maladies inflammatoires chroniques intestinales par l'intermédiaire de variables de diagnostic mesurées incorrectement.

Performance of Feature Selection Methods for Metabolomic Data* / *La performance de méthodes de sélection de caractéristiques pour les données métabolomiques**

Danny Lu ^a, Alex de Leon ^a, Oliver Bathe ^{a,b}, Hans Vogel ^a, Dr. Karen Kopciuk ^{a,b}

^a Department of Mathematics and Statistics, The University of Calgary

^b Alberta Health Services – Cancer Care

This poster will provide a brief overview of my summer research project on feature selection methods for metabolomic data, which are high dimensional data with the $n > p$ problem. An introduction to the field of metabolomics as well several popular feature selection methods will be discussed. Feature selection methods can be used to enhance the analysis of these data by improving estimation accuracy, model interpretability and computational costs. Performance metrics will be based on simulated data, with real cancer data sets used to illustrate the methods. Furthermore, the poster will describe the potential benefits of the research.

Cette affiche présente un bref survol de mon projet de recherche estival sur la sélection de caractéristiques pour les données métabolomiques, des données pluri-dimensionnelles avec le problème $n > p$. J'offrirai une introduction au domaine de la métabolomique et à plusieurs méthodes de sélection de caractéristiques. Les méthodes de sélection de caractéristiques peuvent améliorer l'analyse de ces données en augmentant la précision de l'estimation, en facilitant l'interprétation du modèle et en réduisant la charge computationnelle. Des métriques de performance seront basées sur des données simulées et nous utiliserons des jeux de données réels sur le cancer pour illustrer les méthodes. De plus, l'affiche décrira les bénéfices potentiels de ce projet de recherche.

Post Stratification in Aerial Elk Surveys / *La post-stratification dans les relevés aériens des populations de wapitis*

Megan McCorquodale, Boxin Tang, Carl Schwarz

Department of Statistics & Actuarial Science, Simon Fraser University

Aerial surveys are used to estimate elk populations in British Columbia. The survey area is divided into survey units and stratified by expected density. These units are then surveyed and it was determined that the initial stratification is inaccurate. Data from six surveys in the Kootenay region from 2004 to 2013 was used to develop a method of post stratification estimation using the classification matrix of initial and final strata. This was used to develop estimates with improved precision and an optimal allocation of survey units based on the initial stratification.

Des relevés aériens sont utilisés pour estimer les populations de wapitis en Colombie-Britannique. La zone du relevé est divisée en unités de relevé et stratifiée en fonction de la densité espérée. Ces unités sont par la suite survolées et on a déterminé que la stratification était inexacte. Des données provenant de six relevés de la région de Kootenay de 2004 à 2013 ont été utilisées dans le développement d'une méthode d'estimation par post-stratification faisant usage de la matrice de classification de strates initiales et finales. Cette méthode a été utilisée pour obtenir des estimés comportant une plus grande précision et une allocation optimale des unités de recensement basée sur la stratification initiale.

Contributed Oral Presentations Session 1

Présentations orales, séance 1

4:15-5:50pm

There will be a five minute break at 5pm so participants may switch rooms.

Il aura une pause de cinq minutes à 17h00 pour que les participants puissent changer de chambre.

Location / lieu : CAB 239

Chair / président : Nathaniel Payne (Simon Fraser University)

4:15-4:30pm

Confidence Interval Construction for Short Time Series with Applications in Administrative Health Databases / *Construction d'intervalles de confiance pour des séries chronologiques courtes et applications aux bases de données administratives en santé*

Amber Corkum, Ying Zhang, Paul Cabilio

Department of Mathematics & Statistics, Acadia University

In health policy research, we are interested in determining if there is a significant difference in the behaviour of prescribing physicians pre- and post-policy change. In this talk we use properties of the Wilcoxon Signed Rank Statistic (WSRS) to construct confidence intervals for this change. This involves computing the variance of the WSRS for time series data. We can derive this exactly when the error distribution is normal and dependence structure is known. For the more realistic case, when the error distribution or dependence structure is unknown, we demonstrate how to estimate the variance by block and sieve bootstrap methods.

En recherche sur les politiques de santé, on cherche à déterminer s'il y a un changement significatif dans le comportement des médecins prescripteurs avant et après un changement de politique. Dans cette présentation, nous utilisons les propriétés de la statistique de test *Wilcoxon Signed Rank* (WSRS) pour construire des intervalles de confiance pour ce changement. Cela implique le calcul de la variance de la WSRS pour des données de séries chronologiques. Nous sommes en mesure de dériver cela exactement quand la distribution des erreurs est normale et quand la structure de dépendance est connue. Pour le cas plus réaliste, quand la distribution des erreurs est inconnue, nous démontrons comment estimer la variance par des méthodes *bootstrap block and sieve*.

4:30-4:45pm

Smooth test of goodness-of-fit of normality for the errors of an ARMA process with unknown mean / Test lisse d'ajustement pour les erreurs d'un modèle ARMA avec moyenne inconnue

Pierre Duchesne, Pierre Lafaye de Micheaux, Joseph F. Tagne (presenter)

Département de mathématiques et de statistique, Université de Montréal

Neyman (1937) derived a goodness-of-fit test for standard uniform distribution. This test has been generalized to any distribution by Rayner and Best (1989). Ducharme and Lafaye de Micheaux combined this approach to the Ledwina (1994) data driven principle to find out a goodness-of-fit test for the errors of an ARMA process with known mean. This work generalizes the results of Ducharme and Lafaye de Micheaux (2004) for the unknown mean case. Some simulations study shown that our test held his level in the parameters space, and generally more powerful than existing test. An application to real data is provided.

Neyman (1937) a proposé un test d'adéquation à la loi uniforme qui a été généralisé à une densité quelconque par Rayner et Best (1989). Ducharme et Lafaye de Micheaux (2004) ont combinés cette approche à la procédure de sélection de modèle de Ledwina (1994), pour construire un test de normalité pour les erreurs d'un modèle ARMA avec moyenne connue. Nous généralisons les résultats de Ducharme et Lafaye de Micheaux lorsque la moyenne inconnue. Les simulations montrent que notre test conserve son niveau et possède en général une meilleure puissance que les tests existants. Une application aux données réelles est fournie.

4:45-5:00pm

Modelization of multivariate cointegrated time series / Modélisation des séries chronologiques multivariées et cointégrées

Joël Sango, Pierre Duchesne

Département de mathématiques et de statistique, Université de Montréal

This article considers the estimation and the validation of multivariate periodic cointegrated times series models. To fit the model, two approaches are discussed: full rank estimation and reduced rank estimation. To validate the estimated model, we study the sample matricial autocovariances of the innovations, the residual autocovariances and their asymptotic distributions. With these results, portmanteau type tests are proposed and their finite-sample properties are examined through a simulation study.

Le présent travail porte sur l'estimation et la validation des modèles périodiques multivariés qui peuvent afficher des phénomènes de cointégration de rang quelconque. Deux approches de modélisation sont discutées: l'estimation de plein rang et l'estimation de rang réduit. Pour la validation du modèle estimé, nous étudions les matrices d'autocovariance reposant sur le processus innovation non-observable, les matrices d'autocovariance résiduelles ainsi que leurs distributions asymptotiques. Basés sur ces résultats, des tests de type portemanteau sont construits. Nous procédons aussi à des simulations pour étudier le comportement de ces tests statistiques en échantillons finis.

5:05-5:20pm

A Spatial Cross-Sectional Credibility Model with Dependence Among Risks / Un modèle de crédibilité spatial transversal avec dépendance entre les risques

Jimmy Poon, Yi Lu

Department of Statistics and Actuarial Science, Simon Fraser University

A Bühlmann-Straub type credibility model with dependence structure among risk parameters and conditional spatial cross-sectional dependence is studied. Predictors of future losses for the model under both types of dependence are derived by minimizing the quadratic loss function and non-parametric estimators of structural parameters are considered in the spatial statistics context.

Predictions and estimations made for the proposed model are examined and compared to other models in an application with crop insurance data and a simulation study.

Un modèle de crédibilité de type Bühlmann-Straub avec une structure de dépendance entre les paramètres de risque et avec de la dépendance conditionnelle spatiale transversale est abordé. Nous dérivons des prédicteurs de pertes futures pour le modèle sous les deux types de dépendance par la minimisation de la fonction de perte quadratique et nous abordons des estimateurs non-paramétriques de paramètres structurels dans le contexte des statistiques spatiales. Nous examinons les prédictions et les estimations faites pour le modèle proposé et nous les comparons à celles d'autres modèles dans le cadre d'une application à des données d'assurance des récoltes et d'une étude de simulation.

5:20-5:35pm

Valuation of Contingent Capital Bonds in First-Passage Structural Models / Évaluation d'obligations conditionnelles en capital dans des modèles structurels *First-Passage*

Jingya Li ^a, Adam Metzler ^b, Mark Reesor ^a

^a Department of Applied Mathematics, University of Western Ontario

^b Department of Mathematics, Wilfrid Laurier University

Contingent capital bonds (CCB) are bonds that convert to common shares when a certain predetermined trigger is breached. We use a first-passage structural model to price CCBs based on a capital structure including deposit, equity, and senior and subordinated debt. Under infinite maturity, we derive a closed-form formula for the CCB's fair price and discuss how various factors affect issuing institution's contingent capital cost. Additionally, simulations confirm that broad conclusions drawn in the perpetual case also hold in the finite-maturity case. All the numerical experiments are based on real data and parameters that are calibrated to Canadian banks.

Les obligations conditionnelles en capital (*Contingent Capital Bonds, CCB*) sont des obligations qui se convertissent en actions ordinaires lors du dépassement d'un déclencheur prédéterminé. Nous utilisons un modèle structurel *First-Passage* pour évaluer les CCB, basé sur une structure de capital incluant des dépôts, de l'équité et une dette senior et subordonnée. Sous une maturité infinie, nous dérivons une formule à forme fermée pour le prix juste des CCB et nous expliquons comment de nombreux facteurs affectent le coût conditionnel en capital d'une institution. De plus, des simulations confirment que les conclusions générales obtenues dans le cas perpétuel tiennent également dans le cas d'une échéance déterminée. Toutes les expériences numériques sont basées sur des données réelles et des paramètres calibrés pour les banques canadiennes.

5:35-5:50pm

Declaration Guidelines in Test Cricket / *Consignes de déclaration dans un match-test de cricket*

Harsha Perera^a, Tim. B. Swartz^a, Paramjit Gill^b

^a Department of Statistics and Actuarial Science, Simon Fraser University

^b Mathematics, Statistics, and Physics Unit, Irving K. Barber School of Arts and Sciences, University of British Columbia

This project considers the decision problem of when to declare during the third innings of a test cricket match. There are various factors that affect the decision of the declaring team including the target score, the number of overs remaining, the relative desire to win versus draw, and the scoring characteristics of the particular match. Decision rules are obtained and these are assessed against historical matches. It is found that teams have traditionally been cautious in declaring, and that optimal decision making would lead to more frequent and earlier declarations.

Ce projet aborde le problème de décision concernant le moment de la déclaration durant la troisième manche d'un match-test de cricket. Il y a plusieurs facteurs qui affectent la décision de l'équipe déclarante, comme le pointage cible, le nombre restant d'*overs*, le désir relatif de gagner par rapport à celui d'accorder un *ex aequo* et les caractéristiques de pointage du match en question. Nous obtenons des règles de décision qui sont par la suite évaluées à l'aide de données historiques. Nous remarquons que les équipes ont été traditionnellement prudentes dans la déclaration et qu'une prise de décision optimale mènerait à des déclarations plus fréquentes et plus hâtives.

Contributed Oral Presentations Session 2

Présentations orales, séance 2

4:15-5:50pm

There will be a five minute break at 5pm so participants may switch rooms.

Il aura une pause de cinq minutes à 17h00 pour que les participants puissent changer de chambre.

Location / lieu : CAB 273

Chair / président : Ying Yan (University of Waterloo)

4:15-4:30pm

Functional model-based clustering for longitudinal data / Analyse de données longitudinales selon une approche fonctionnelle de la classification non supervisée

Folly Adjogou, Alejandro Murua

Département de mathématiques et de statistique, Université de Montréal.

We develop a flexible model for the analysis and clustering of complete or sparse longitudinal data. The model combines functional principal component analysis (based on splines) and model-based clustering. The main data groups arise from clusters in the space of spline coefficients. The clusters are modeled by a mixture of Student's t-distributions with unknown degrees of freedom. We develop an approximation of the marginal log-likelihood (MLL) for model selection which compares favourably with other popular criteria (AIC and BIC). We also consider an extension of our model to curves in multiple dimensions. We will present applications to gene expression and clinical data.

Nous développons un modèle flexible pour l'analyse et la classification de données longitudinales (balancées ou non). Le modèle combine l'analyse fonctionnelle en composantes principales et le regroupement en classes qui repose sur l'espace des coefficients dans la base des splines et un modèle de mélange de distributions de student de degrés de liberté inconnus. Nous développons une approximation de la log-vraisemblance marginale (MLL) pour la sélection de modèles qui se compare favorablement aux critères usuels (AIC et BIC). Nous considérons également une extension du modèle aux courbes multidimensionnelles. Nos applications principales concernent les données d'expression génétique et les essais cliniques.

4 :30-4:45pm

A Moderated Trace Test for the Growth Curve Model and its Application for Gene-filtering in Time-Course Microarray Data / Un test de la trace modéré pour le modèle de courbe de croissance et son application au filtrage de gènes avec données temporelles de micromatrices d'ADN

Sayantee Jana ^a, Narayanaswamy Balakrishnan ^a, Dietrich von Rosen ^b, Jemila S Hamid ^c

^a Department of Mathematics and Statistics, McMaster University

^b Department of Biometry and Informatics, Swedish university of agricultural sciences

^c Department of Clinical Epidemiology and Biostatistics, McMaster University

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

Les tests traditionnels pour l'analyse de données longitudinales ne sont pas applicables pour les données de grandes dimensions (où $n < p$). Nous proposons un test modéré utilisant le modèle de courbe de croissance qui est utile pour analyser des données longitudinales de grandes dimensions. De nombreuses simulations démontrent que notre test modéré est sans biais et est monotone par rapport à la taille d'échantillon et à la valeur du paramètre, quel que soit la covariance. Nous avons illustré l'application de notre méthode pour le filtrage de gènes avec des données temporelles de micromatrices d'ADN. Nous utilisons des données sur le cancer du poumon, où les mesures ont été prises sur des tissus de poumons humains normaux et traités chimiquement. Nous avons identifié 1053 (4.7%) gènes significatifs parmi les 22 227 gènes.

4:45-5:00pm

Comparing the direct and approximate approaches to implement marginal structural Cox models: An Application to Multiple Sclerosis / Comparaison d'approches directe et approximative pour mettre en œuvre des modèles de Cox structurels marginaux : Une application à la sclérose en plaques

Mohammad Ehsanul Karim ^a, Paul Gustafson ^a, John Petkau ^a, Afsaneh Shirani ^b, Yinshan Zhao ^b, Elaine Kingwell ^b, Charity Evans ^c, Mia van der Kop ^b, Joel Oger ^b, Helen Tremlett ^b

^a Department of Statistics, University of British Columbia

^b Division of Neurology, Department of Medicine, University of British Columbia

^c College of Pharmacy and Nutrition, University of Saskatchewan

Inverse probability weighted (IPW) pooled logistic regression is generally used to approximate marginal structural Cox model (MSCM) estimates. Recently, through simulation, researchers showed that direct use of the IPW Cox model produced MSCM estimates with better properties and suggested that the use of normalized weights yields more precise estimates of the treatment effect. Using a real longitudinal dataset (British Columbia, 1995-2008), we applied the direct MSCM approach and the normalized weights to investigate the association between beta-interferon treatment and disability progression in multiple sclerosis, and compared the estimates with those from the approximate approach.

La régression logistique pondérée selon les probabilités inverses (PPI) est généralement utilisée pour approximer les estimateurs du modèle de Cox structurel marginal (MCSM). Récemment, par simulation, des chercheurs ont montré que l'utilisation directe du modèle de Cox PPI produisait des estimateurs MCSM avec de meilleures propriétés et ont suggéré que l'utilisation de poids normalisés produisait des estimateurs plus précis de l'effet du traitement. À l'aide de réelles données longitudinales (Colombie Britannique, 1995-2008), nous avons appliqué la méthode MCSM et les poids normalisés pour étudier le lien entre le traitement interféron bêta et la progression de l'invalidité dans la sclérose en plaques, et comparé les estimateurs avec ceux de la méthode approximative.

5 :05-5:20pm

Statistical Inference for Discriminant Function Coefficients in Doubly Multivariate Data / Inférence statistique pour les coefficients de la fonction discriminante en matière de données doublement multivariées

Bolanle Dansu^a, Lisa M. Lix^a, Tolulope Sajobi^b

^a Department of Community Health Sciences, Faculty of Medicine, University of Manitoba

^b Faculty of Medicine, University of Calgary

Discriminant function coefficients (DFCs) from linear discriminant analysis are used to predict group membership and evaluate the relative importance of correlated variables for group separation. This study develops new statistical procedures to test for change in the relative importance of variables in doubly multivariate data, which are collected repeatedly over time. These DFC procedures use an asymptotic z-test, theoretical t-test, and empirical t-test based on a bootstrap critical value. A Monte Carlo study compares the Type I error and statistical power rates of the new procedures for two time points under different configurations of the group mean and covariance structures.

Les coefficients de la fonction discriminante (CFD) obtenus par analyse discriminante linéaire sont utilisés pour prédire les groupes et évaluer l'importance relative de variables corrélées pour la séparation des groupes. Cette étude développe de nouvelles méthodes statistiques pour tester les changements dans l'importance relative des variables associées à des données doublement multivariées, observées à plusieurs reprises. Ces méthodes CFD utilisent un test Z asymptotique, un test t théorique, et un test t empirique basé sur une valeur critique bootstrap. Une étude de Monte Carlo compare l'erreur de type I et les taux de puissance statistique des nouvelles méthodes pour deux temps pour des configurations différentes quant à la moyenne de groupe et structure de covariance.

5:20-5:35pm

Balanced k-nearest neighbor imputation / *Imputation équilibrée par les k plus proches voisins*

Yves Tillé, Caren Hasler (presenter)

Institut de Statistique, Université de Neuchâtel, Switzerland

In order to overcome the problem of item nonresponse, random imputations are often used because they tend to preserve the distribution of the imputed variable. Among the methods of random imputation, the random hot-deck has the interesting property that the imputed values are observed values. We present a new random method of hot-deck imputation which enables us to select the imputed values such that some balancing equations are satisfied and such that the donors are selected in a neighborhood of the recipients.

Des méthodes d'imputation aléatoire sont souvent utilisées pour traiter la non-réponse car de telles méthodes ont tendance à préserver la distribution des variables que l'on impute. Parmi les méthodes d'imputation aléatoire, on trouve les méthodes dites hot-deck ; celles-ci jouissent de l'intéressante propriété d'imputer des valeurs observées. Nous proposons une nouvelle méthode d'imputation hot-deck aléatoire permettant de sélectionner les valeurs imputées de telle manière que des équations d'équilibrage soient satisfaites et que les donneurs potentiels pour un receveur appartiennent à un voisinage de celui-ci.

Contributed Oral Presentations Session 3

Présentations orales, séance 3

4:15-5:50pm

There will be a five minute break at 5pm so participants may switch rooms.

Il aura une pause de cinq minutes à 17h00 pour que les participants puissent changer de chambre.

Location / lieu : CAB 281

Chair / présidente : Mahdis Azadbakhsh (York University)

* = eligible for the undergraduate oral presentation award / admissible pour le prix de présentation pour les étudiants de premier cycle

4:15-4:30pm

Two-level Star-based Factorial Designs with Randomization Restrictions: Checking Isomorphism and Equivalence * / Plan factoriel d'expérience « star-based » à deux niveaux avec restrictions aléatoires : Recherche d'isomorphisme et équivalence *

Neil Spencer, Pritam Ranjan

Department of Mathematics & Statistics, Acadia University

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

La recherche d'isomorphisme des plans factoriels d'expérience avec restrictions aléatoires est un sérieux défi en plan d'expérience. Dans cette présentation, je regarde la sous classe d'isomorphisme à deux étoiles de $PG(p-1, 2)$, équivalente à un plan d'expérience à 2^p facteurs. Une étoile est une collection de sous ensembles distincts de $PG(p-1, 2)$ avec une intersection commune constituant une couverture de $PG(p-1, 2)$. Je propose un algorithme de recherche des isomorphismes. L'innovation dans mon approche réside dans l'utilisation de la géométrie sous-jacente pour réduire la recherche de sous espace, ainsi qu'une nouvelle approche de recherche d'équivalence entre les étoiles.

4:30-4:45pm

Discrete Log-Concave Distribution in Aggregate Loss Modeling * / *Lois discrètes log-concaves dans la modélisation de perte totale* *

Gabriela González Martínez, Hanna Jankowski

Department of Mathematics and Statistics, York University

Insurance companies are interested in the behaviour of the collective set of insured risks, known as the aggregate loss distribution. Most current methods, to our best knowledge, require a parametric assumption. However, inference using these methods is questionable due to potential model misspecification. The non-parametric model we propose considers the probability mass function (PMF) of the number of claims as a mixture of a discrete log-concave distribution with a probability mass at zero. It therefore provides robustness to the model that, even under model misspecification, will give a close estimation of the true PMF.

Les compagnies d'assurance sont intéressées par le comportement de l'ensemble collectif de risques assurés, connus sous le nom de la loi de perte totale. La plupart des méthodes actuelles, à notre connaissance, nécessitent une hypothèse paramétrique. Cependant, l'inférence à l'aide de ces méthodes est douteuse en raison de la possibilité d'une mauvaise spécification du modèle. Le modèle non-paramétrique que nous proposons considère la fonction de masse de probabilité (FMP) du nombre de revendications comme un mélange de lois log-concaves discrètes avec une masse de probabilité à zéro. Cela donne un modèle robuste qui, même sous une mauvaise spécification du modèle, donnera une estimation proche de la vraie FMP.

4:45-5:00pm

Keeping Pace: Comparing Active, Moderately Active, and Inactive Children and Youth * / *Keeping Pace: Comparaison de l'Actif, Modérément Actif, et Inactif Enfants et Jeunes* *

Christopher M. van Bommel^a, Angela M. Kolen-Thompson^b, Laurie A. Wadsworth^c, and Xu (Sunny) Wang^a

^a Department of Mathematics, Statistics & Computer Science, St. Francis Xavier University

^b Department of Human Kinetics, St. Francis Xavier University

^c Department of Human Nutrition, St. Francis Xavier University

Keeping Pace is a study of physical activity of grade 3, 7, and 11 students and dietary intake of grade 7 and 11 students in Nova Scotia conducted in 2009. One question is to determine the differences among grade 7 students who achieved different activity levels: active (5 or more days of at least 60 minutes of moderate or more intense physical activity per week), moderately active (3-4 days) and inactive (0-2 days). I will discuss using Mixed-Effect Models to analyze the differences among the factors that affect physical activity among the participants of different activity levels and some results.

Keeping Pace est une étude réalisée en 2009 sur l'activité physique des étudiants de 3^e, 7^e, et 11^e année, et l'apport alimentaire des étudiants de 7^e et 11^e année en Nouvelle-Écosse. Une question est de déterminer les différences parmi les étudiants de 7^e année qui accomplissaient des niveaux d'activités différents: actif (5 ou plus jours avec au moins 60 minutes d'activité modérée ou plus intense par semaine), modérément actif (3-4 jours), et inactif (0-2 jours). Je vais discuter l'utilisation des Modèles à

effets mixtes d'analyser les différences parmi les facteurs qui affectent l'activité physique parmi les participants des niveaux d'activité physique différents et des résultats.

5:05-5:20pm

The effect of local geometry and solvent conditions on knotting in a self-avoiding polygon model / L'effet de la géométrie locale et des conditions de solvant sur les nœuds dans un modèle de polygone autoévitant

Matthew Schmirler^a, Chris Soteris^a, Michael Szafron^b

^a Department of Mathematics and Statistics, University of Saskatchewan

^b School of Public Health, University of Saskatchewan

The strand passage action of type II DNA topoisomerase - an enzyme that helps unknot DNA in the cell - is modeled here using self-avoiding polygons (SAPs) in the simple cubic lattice. Each SAP contains a fixed structure (which represents two segments of the SAP pinched close together) that can be used to implement a strand passage. A Composite Markov Chain is used to generate samples of fixed-structure SAPs according to varying solvent conditions. These samples are used to study how knotting after a strand passage varies depending on solvent conditions and the local geometry of the fixed structure.

L'action de passage de séquences d'ADN topoisomérase de type II, une enzyme qui aide à dénouer l'ADN dans la cellule, est ici modélisée en utilisant des polygones autoévitants (SAPs) dans le treillis cubique simple. Chaque SAP contient une structure fixe (représentant deux segments du SAP pincés près l'un de l'autre), qui peut être utilisée pour implémenter un passage de séquence. Une Chaîne de Markov Composée est utilisée pour générer des échantillons de SAPs à structure fixe en fonction de différentes conditions de solvant. Ces échantillons sont utilisés pour étudier comment les nœuds produits après un passage de séquence varient selon les conditions de solvant et la géométrie locale de la structure fixe.

5:20-5:35pm

A Multivariate Model for Gene Networks / Un modèle multidimensionnel pour les réseaux de gène

Abdollah Safari^a, Ali Sharifi^b, Hamid Pezeshk^c, Mehdi Sadeghi^d, Changiz Eslahchi^e

^a Department of Statistics and Actuarial Science, Simon Fraser University

^b Institute of Biophysics and Biochemistry, University of Tehran

^c Bioinformatics Research Group, School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran

^d National Institute of Genetic Engineering and Biotechnology, Tehran

^e Faculty of Mathematics, Shahid Beheshti University, Tehran

There are several methods for inference in gene networks, but there are few cases in which historical information has been considered. Our approach implements a Bayesian model that utilizes available information. Our model, which assumes a proper prior distribution, is an improvement over standard methods as it takes the dependency of parameters into account. We compare the results obtained from (1) method of moments estimation and (2) combining prior information and samples to estimate hyper

parameters. A simulation based on Gibbs samplers was used to compare the results. We also discuss the strengths and weaknesses of each method.

Il existe plusieurs méthodes d'extrapolation pour les réseaux de gène, mais très peu de cas tiennent compte de l'état des connaissances. Nous mettons en œuvre une approche intégrant l'information disponible. Notre modèle, basé sur l'hypothèse d'une distribution à priori, va au-delà des méthodes standards en tenant compte de la dépendance des paramètres. Nous comparons les résultats obtenus (1) par la méthode des moments avec celui (2) combinant l'information à priori et l'échantillon pour estimer les paramètres. Une simulation basée sur l'algorithme de Gibbs permet la comparaison des résultats. Nous discutons également des avantages et inconvénients de chacune des méthodes.

Keynote Address / Discours d'honneur

5:50-6:50pm

Location / lieu : CAB 243

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

Jeffrey Rosenthal, PhD

Lessons from a Twisted Career Path

In this talk, I will reflect upon the unique twists and turns of my own academic career path, which took me from enthusiastic pure-math undergrad, to worried Harvard PhD student, to struggling junior probability professor, to ultimately receiving the COPSS Presidents' Award for outstanding contributions to statistics. I will try to use my story to provide lessons and insights for budding statistical scholars and researchers.

Leçons tirées d'un cheminement de carrière tordu

Dans cette présentation, j'exhiberai ma singulière démarche universitaire, qui m'emmena d'étudiant en mathématiques pures, en passant par anxieux doctorant à Harvard et professeur junior en probabilité surchargé, jusqu'à finalement récipiendaire du prix des présidents du COPSS pour contributions exceptionnelles en statistiques. Je vais essayer d'utiliser mon histoire pour fournir des leçons et une certaine sagesse pour les statisticiens en développement.

Thank you ! Merci !

SSC executives and staff, including Christian Léger, Tim Swartz, Russell Steele, Alan Kelm, and Sarah Watson.

Staff at the University of Alberta, including Leona Guthrie

David Grubbs (CRC Press) and Bud Homsy (PIMS)

Our keynote and invited speakers: Jeffrey Rosenthal, Ori Stitelman, Matthias Schonlau, and Kathryn Mills

Student and faculty volunteers

Map of the 2nd floor of CAB (with room numbers)



Notes

Notes

Notes