



**SSC/WNAR/IMS 2001
Simon Fraser University
June 10-14, 2001**

**The Joint Meetings of the
Statistical Society of Canada**

**Western North American Region of the
International Biometric Society**

Institute of Mathematical Statistics

Contents/Table des matières

Welcome	2
Sponsors.....	2
General Information	
Registration.....	2
Food Service.....	3
The Opening Mixer/Poster Session	3
The Student Mixer	3
The Banquet.....	3
Grouse Mountain Social Event.....	4
Shuttle Service.....	4
Job Fair.....	4
Publishers Book Display (AQ 4000)	4
Rooms	4
Parking	4
Internet Access	5
Committee Meetings	6
Outline of Events.....	7
Workshops	9
Scientific Programme/Programme scientifique	10
Poster Abstracts.....	33
Résumés de Séance par affichage	41
Abstracts	51
Résumés.....	94
Index of Participants.....	144

Welcome:

The faculty members of the Department of Statistics and Actuarial Science at Simon Fraser University welcome you to Burnaby, British Columbia. We hope that you find the meeting professionally rewarding and enjoyable.

Sponsors:

SSC/WNAR/IMS 2001 extends special thanks to the sponsors of the meeting. In particular, we thank Simon Fraser University (SFU) for use of resources and financial support. We also thank the Pacific Institute of the Mathematical Sciences, the Centre de Recherche Mathématique, the Fields Institute, The International Indian Statistical Association, the Faculty of Science at SFU and the former Department of Mathematics and Statistics at SFU.



Statistics Canada
Statistique Canada



Simon Fraser
University

Activities of the Biostatistics Section have been supported in part by Quintiles Canada, Bayer Inc, Bristol-Myers-Squibb and Purdue Pharma.

Local Arrangements: Tim Swartz (Chair)

Programme Committee: Mary Lesperance (Chair)

Laura Salter (WNAR), Francois Bellavance (SSC Biostatistics), Robert Burton (IMS)
Patrick Farrell (SSC Survey Methods), Bovas Abraham (SSC Business and Industry),
Julie Zhou (Contributed Sessions Chair)

Programme Assistants: Kelly Choo, Irina Gavrilova, Ariana Clapton (University of Victoria)

French Translation:

Chair: Jean-François Angers (Dép. de mathématiques et de statistique Université de Montréal)

Members: Stéphane Courchesne
Geneviève Lefebvre
Sylvie Makhzoum.

(Not all of the programme and abstracts were translated into French because of technical difficulties. The Programme Chair assumes all responsibility.)

Registration:

The Registration desk will be open in the Diamond University Club from 18:00 to 21:00 on Sunday June 10 at the same location as the opening mixer/poster session.

The Registration desk will then be moved outside Images Theatre from 08:00 to 10:30 during the time of the opening Gold Medal Address.

For the remainder of the conference, the Registration desk will be located in AQ4125 nearby many of the smaller session rooms.

Food Services:

Various food services at Simon Fraser University will be available Monday through Thursday of the conference as follows:

Ravens Cafeteria in the West Mall Complex	08:00 to 16:00
MacKenzie Cafeteria in the north AQ hall	07:30 to 18:30 (closest to committee meetings)
White Spot (burgers) near James Douglas room	11:00 to 14:30
Convenience Store near James Douglas room	08:00 to 19:00
Renaissance Coffee I (good coffee and snacks)	07:00 to 19:00 in northeast corner of the AQ
Renaissance Coffee II (good coffee and snacks)	07:00 to 16:30 in the east section of Applied Science
Diamond University Club (better food but may)	11:30 to 14:00 require a reservation
Mr Submarine in the Maggie Benston Bldg	08:00 to 20:00
Highland Pub in the Maggie Benston Bldg	11:00 to 22:00 beverages and short-order food

The Opening Mixer/Poster Session:

The Opening Mixer/Poster Session will be held from 18:00 to 22:00 on Sunday June 10 at the Diamond University Club at SFU. The Registration desk will be open in the foyer. Nice hors d'oeuvres and a cash bar will be available. Don't miss it. A shuttle service between SFU and the four conference hotels (Executive Inn Burnaby, Inn at Westminster Quay, Best Western Coquitlam and Executive Plaza Coquitlam) will run on a continuous basis between the hours of 18:00 and 22:00.

The Student Mixer:

For students, those recently graduated and for those young at heart, The Student Mixer will be held at the Diamond University Club on Monday June 11 from 19:30 to 22:00. Pre-recorded music, food platters and a cash bar will be available.

The Banquet:

All registered delegates are invited to the Conference Banquet, Tuesday June 12 (19:00) at the Hyatt (Regency Ballroom), 655 Burrard Street in downtown Vancouver. A banquet ticket for a four-course dinner has been provided in the registration package. A shuttle service will be provided which leaves SFU (on the road north of the library, not far from the Diamond University Club) at 17:45. Buses will leave the four conference hotels (Executive Inn Burnaby, Inn at Westminster Quay, Best Western Coquitlam and Executive Plaza Coquitlam) at 18:00. Buses will depart from the Hyatt at 21:45 and drop people off at SFU and the four conference hotels.

For those driving to the Hyatt, take Hastings Street at the western base of SFU travelling west for approximately 15km to Commercial Street. Turn left on Commercial and within 1km turn right on Venables. Venables will lead you over the viaduct to Dunsmuir Street in the downtown whence you will turn left on Burrard Street.

Grouse Mountain Social Event:

On Wednesday June 13, a shuttle bus will leave SFU at 17:05 (on the road north of the library, not far from the Diamond University Club) for Grouse Mountain. Tickets need to be purchased in advance. Those attending, can choose between taking the effortless cable car to the top of the mountain or doing the "Grouse Grind"; a steep 1000 metre ascent that takes between 1 and 1.5 hours. A salmon barbeque will be served from 19:15 to 20:15 in the Timber Room. While at the top, enjoy the free "Theatre in the Sky" presentation (every half hour, last show at 21:00), hike around, check out the totem poles and visit the Gift Shop. Be wary of bears! The buses will depart Grouse Mountain for SFU and the conference hotels at 20:50.

Shuttle Service:

For those staying at the Executive Inn Burnaby, a daily shuttle bus has been arranged departing the hotel at 08:00 Monday through Thursday and leaving SFU (on the road north of the library, not far from the Diamond University Club) at 17:05 Monday through Wednesday. On Thursday, the shuttle will leave SFU at 12:05.

Job Fair:

Interviews for the Job Fair will take place in the ten small interview rooms on the 1300 level of the Maggie Benston Building. Individuals participating in the Job Fair should check at the Registration Desk for times of interviews and further updates.

Publishers Book Display:

Make sure that you visit the AQ4000 lounge where representatives from Springer, McGraw-Hill, Nelson and Wiley will be displaying some of their best books and journals.

Rooms:

The meeting rooms at SFU are spread out across the campus. Consult the maps. The first map provides more detail; the second map indicates locations relevant to the meeting. Note that coffee breaks are daily at 10:00 and 15:00 in the James Douglas Room.

Parking:

Those who drive to SFU should park in the Visitors Parkade. Upon exiting, if you identify yourself as belonging to SSC/WNAR/IMS 2001, you will be charged the discounted daily rate of \$3.

Internet Access:

The following computing facilities are available for those attending the conference.

	MWTR	F	SAT	SUN
Microcomputer PC facility AQ 3146	0830-2200	0830-1800	CLOSED	1200-1800
CIC PC Lab AQ 3145.1	0830-2200	0830-1800	1100-2200	1100-2200
CIC Mac Lab AQ 3145.1	0830-2200	0830-1800	1100-2200	1100-2200
Wordstation, Library second floor	0800-2345	0800-1800	1100-2200	1100-2200
Westmall Lab WMX 2506	0830-2200	0830-1800	1100-2200	1100-2200

You will have to follow the instructions on the computer you use. Generally these involve logging in with an ID and password which are printed on the machine. Once logged in that way you can use telnet (to make a telnet connection to your own computer) and Netscape (to browse the internet or use netmail).

When you try to access off campus services a small menu will pop up asking you to supply your SFU computer ID. In this case, use the computer account "statconf" and the password "YESAB2K1" - do not type the quotation marks.

Committee Meetings

Saturday, June 9		
6:00	Executive Committee	Executive Plaza Coquitlam Hotel
Sunday, June 10		
9:00-11:00	Finance	AQ4100
10:30-11:30	Accreditation	AQ4150
11:00-12:00	Publications	AQ4140
12:00-5:00	Board	AQ4100 (catered lunch)
Monday, June 11		
7:15-8:15	Biostatistics Executive	AQ4100
7:15-8:15	Survey Executive	AQ4140
7:15-8:15	Program	AQ4150
12:15-1:15	CJS Editorial Board	AQ4100
12:15-1:15	Public Relations	AQ4140
12:15-1:15	Liaison Editorial Board	AQ4150
12:15-1:15	WNAR Young Investigators' Lunch	DUC (catered lunch)
5:00-6:00	Biostatistics General Meeting	Images
5:00-6:00	Survey Methods General Meeting	WMX3520
5:00-6:00	BISS General Meeting	AQ4150
5:30-7:00	WNAR Executive Meeting	DUC down (catered dinner)
6:00-7:00	BISS Executive Meeting	AQ4150
Tuesday, June 12		
7:15-8:15	Accreditation	AQ4100
7:15-8:15	Research	AQ4140
12:15-1:15	Bilingualism	AQ4120
12:15-1:15	Statistical Education	AQ4140
12:15-1:15	Committee for Women	AQ4150
12:15-1:15	WNAR Advisory Board	AQ4100 (catered lunch)
12:15-1:15	NSERC/Research Committee	AQ4130 (catered lunch)
Wednesday, June 13		
7:15-8:15	Biostatistics Executive	AQ4100
7:15-8:15	Professional Development	AQ4140
12:15-1:15	Statistics Department Heads	DUC (catered lunch)
12:15-1:15	WNAR Executive	DUC (catered lunch)
5:00-6:00	SSC General Meeting	Images
Thursday, June 14		
12:30-3:00	SSC Board	AQ4100 (catered lunch)
3:00-5:00	SSC Executive	AQ4100

Outline of Events

Sunday, June 10 /Dimanche 10 juin	9:00-4:00	Biostatistics Workshop - DUC upstairs Survey Methods Workshop - DUC downstairs BISS Workshop (Data Mining) - WMX 3520
	6:00-10:00	Contributed Poster Session, DUC and Opening Mixer
Monday, June 11th /Lundi 11 juin	8:30-10:00	Welcome and SSC Gold Medal Address, Images
	10:30-12:00	WNAR Presidential Invited Address, Images
		Contrasting US 2000 and Canada 2001 Censuses, Halpern
		Unusual Stochastic Models, AQ 4150
	1:30-3:00	Session in Honour of V.P. Godambe's 75 th Anniversary, Halpern
		Spatial Statistics, Images
		Robust Inference, AQ 4150
		Contributed Session: Survival Analysis, AQ 4140
		Case Studies Session I: Climate Change/ Changements climatiques WMX 3520
		WNAR Student Paper Presentations, AQ 4130
	3:30-5:00	IMS Special Invited Session I, Halpern
		Business and Industry Section Special Invited Session, Images
		Gene Ancestry, Identity and Allelic Associations, AQ 4130
		Survey Methods Contributed Session: Imputation and Variance Estimation, AQ 4150
		Contributed Session: Optimal Designs, AQ 4140
	7:30-10:00	Student Mixer, DUC
Tuesday, June 12 /Mardi 12 juin	8:30-10:00	[8:30 – 9:15] Pierre Robillard Award Presentation, Images [9:15 – 10:00] Canadian Journal of Statistics Award Presentation, Images
		Measurement Error, Halpern
		Survey Methods Contributed Session: Survey Applications, AQ 4150
		WNAR Student Presentations, AQ 4130
	10:30-12:00	On Becoming a Successful Statistician: A Panel Discussion, Images
		Small Area Estimation, AQ 4150
		Likelihood, Halpern
		Statistics in Industry and Technology, AQ 4130
		Contributed Session: Timeseries, AQ 4140
	1:30-3:00	Session in Honour of J.N.K. Rao's Retirement, Images
		Bioinformatics Tutorial I, Halpern
		Discrete Probability and Combinatorics, AQ 4150
	3:30-5:00	IMS Special Invited Session II, Images
		Bioinformatics Tutorial II, Halpern
		Canadian Initiative on Social Statistics, AQ 4150
	7:00-9:45	Banquet, Hyatt, Regency Ballroom, 655 Burrard St.

Wednesday, June 13 /Mercredi 13 juin	8:30-10:00	SSC Presidential Invited Address, Images
	10:30-12:00	Statistics in Fisheries I, Images
		Nonparametric Econometrics, AQ 4130
		Contributed Session: Longitudinal and Correlated Data Analysis, AQ 4140
		Contributed Session: Inference, AQ 4120
		Case Studies Session II: Old Growth Forest/En forêt ancienne, Halpern
		WNAR Student Presentations, AQ 4150
	1:30-3:00	Statistics in Forestry, Halpern
		Statistics in Fisheries II, Images
		Innovative Applications of Statistics in Business, AQ 4150
		Recent Results on Normal and Related Approximations, AQ 4140
		Block Designs: Directions and Inspirations, AQ 4130
		Survey Methods Contributed Session: Estimation 1, AQ 4120
	3:30-5:00	Statistics in Fisheries III, Images
		Brownian Sheets and Stochastic Partial Differential Equations, AQ 4140
		Survey Methods Contributed Session: Estimation 2, AQ 4150
		NSERC Grant Selection Committee - Open Meeting, Halpern
		Contributed Session: Regression, AQ 4130
		Contributed Session: Assorted, AQ 4120
	5:00-6:00	SSC Annual General Meeting, Images
	5:05-	Grouse Mountain Social Event, Grouse Mountain
Thursday, June 14 /Jeudi 14 juin	8:30-10:00	Canadian Journal of Statistics Read Paper and Discussion, Images
		Environmental Toxicology, Halpern
		Contributed Session: Assorted, AQ 4150
	10:30-12:00	Nonlinear Timeseries, Halpern
		Diagnostic Testing, Images
		Contributed Session: Assorted, AQ 4150
	9:00-4:00	Workshop - Beyond MCMC, WMX 3520

Workshops

Sunday, June 10/Dimanche 10 juin

9:00-4:00, DUC upstairs

Biostatistics Section and WNAR Workshop

Inferences from Genetic Data on Pedigrees

Workshop Leader: Elizabeth A. THOMPSON (University of Washington)

9:00-4:00, DUC downstairs

Survey Methods Section Workshop

The Analysis of Sample Survey Data

Workshop Leader: John L. ELTINGE (U.S. Bureau of Labor Statistics and Texas A&M)

9:00-4:00 WMX 3520

Business and Industry Workshop

Data Mining

Workshop Leader: Hugh CHIPMAN (University of Waterloo)

Thursday, June 14/Jeudi 14 juin

9:00-4:00, WMX 3520

SSC Workshop

Beyond MCMC: Monte Carlo Methods in Bayesian Computation

Workshop Leaders: Joseph IBRAHIM (Harvard School of Public Health), Ming-Hui CHEN (Worcester Polytechnic Institute)

Scientific Programme/Programme scientifique
Poster Session/ Séance par affichage

Sunday, June 10 /Dimanche 10 juin

6:00 - 10:00 p.m.

Contributed Poster Session (Diamond University Club)

Luc ADJENGUE (École Polytechnique de Montréal)

An Estimation Procedure for the Parameters of a Partially Observable Markov Decision Process/ Un Procédé D'Estimation Pour Les Paramètres D'Un Processus De Décision De Markov Partiellement Observable

Ejaz AHMED (University of Regina)

Inference Procedures for Process Capability Indices in Arbitrary Populations/ Procédures D'Inférence Pour Des Processus D'Indice De Capacité Dans Des Populations Arbitraires

Paul D. BELL and Timothy E. O'BRIEN (Loyola University)

Predicting attendance at Major League Baseball Games/ Prediction de l'assistance aux jeux de major league baseball

David BELLHOUSE (University of Western Ontario)

The Reverend Thomas Bayes FRS: a Biographical Sketch/ Le Révérend Thomas Bayes FRS: Un Essai Biographique

Willard BRAUN (University of Western Ontario)

Kernel Spline Regression/ La Régression Spline Avec Noyau

David BRILLINGER (University of California, Berkeley)

Process Analysis Of Extreme Flow Rates Along The Mississippi River/ Le Processus D'Analyse Des Taux De Débits Extrêmes Le Long Du Fleuve De Mississippi

Noel CADIGAN (Fisheries and Oceans)

Generalized Local Influence, With Application To Sequential Population Analysis/ L'Influence Locale Généralisée Avec Application À L'Analyse De La Population Séquentielle

Hugh CHIPMAN (University of Waterloo)

Additive Models For Survey Data Via Penalized Least Squares/ Modèles Additifs Pour Des Données De Sondage Par L'Intermédiaire Des Moindres Carrés Pénalisés

Laura COWEN (Simon Fraser University)

Adjusting Survival Estimates Using Radio-Tag Failure Curves/ L'Ajustement Des Estimateurs De Survie En Utilisant Des Courbes De Panne De Marqueur Radio

Monica DRAGOMIR (University of Toronto)

Wavelet Regression for Conditional Medians/ La Régression En Ondelettes Pour Des Médianes Conditionnelles

James DUNN (University of Arkansas)

Variations On Multinomial Logit Models: Alternative Links And Stepwise Selection/Variations Sur Les Modèles Logit Multinomiaux : Liens Alternatifs Et La Sélection Pas À Pas

Paramjit GILL (Okanagan University College)

Spatio-Temporal Modelling of Teenage Birth Data/ La Modélisation Spatio-Temporelle Pour Les Données De La Grossesse Des Adolescentes

Hyang Mi KIM (Carleton University)

Preliminary Test Prediction Of Population Total Under Simple Regression Model With Measurement Errors

Milorad KOVACEVIC (Statistics Canada)

Inference about domain parameters when the domain size depends on nuisance parameters/ Inférence sur des paramètres de domaine quand la taille du domaine dépend de paramètres de nuisance.

Salim LARDJANE (Crest-Ensay)

Marginal Density Estimation For Approximable Stochastic Processes/ Une Procédure De Comparaison Multiple Unilatéral Pour Des Modèles Normaux Restreints

Jack LEE (Department of Statistical & Actuarial Science)

Wavelets Application (Wave Shrinkage) On Time Series Models/ L'Application Des Ondelettes À Rétrécisseur Sur Les Modèles De Série Chronologique

Juan Pablo LEWINGER (University of Toronto)

Confidence Intervals for Penalized Likelihood Estimates of Logistic Regression Parameters/ Les Intervalles De Confiance Pour Des Estimateurs De Vraisemblance Pénalisés Des Paramètres D'Une Régression Logistique

Chang LIE (Seoul National University)

Joint Reliability Importance of Partial Redundant System with Independent Components/ L'Importance De La Fiabilité Conjointe Des Systèmes Redondants Partiels Avec Des Composantes Indépendantes

Erling LUNDEVALLER (Svenska Statistiker Samfundet)

Test Of Random Effects In Heteroskedastic Linear Models/ Test Des Effets Aléatoires Pour Des Modèles Linéaire Hétéroskédastique

Etsuo MIYAOKA (Science University of Tokyo)

Statistical Models for Binary Crossover Data/ Modèles Statistiques Pour Des Données Binaires Croisées

Theophile NIYONSENGA (Sherbrooke University Hospital Center)

Primary Care Physician's Approach In Treatment Of Dyslipidemias Among Women: Application Of Hierarchical Models/ L'Approche Du Médecin De Premier Soins Pour Le Traitement De Dyslipidemias Chez Les Femmes: Application Des Modèles Hiérarchiques

Sudhir PAUL (University of Windsor)

Generalized Linear Model, Zero-Inflation and Over-Dispersion

Ruxandra PINTO (University of Toronto)

Improving Markov Chain Monte Carlo Estimators by Coupling to an Approximating Chain/
L'Amélioration Des Estimateurs De Monte Carlo Markoviens En Les Regroupant Avec Une Chaîne
Approximative

Annie QU (Oregon State University)

Testing For Nonignorability Of Dropouts For Longitudinal Studies

Timothy RAMSAY (Statistics Canada)

Small Domain Estimation and Non-parametric Regression in the Canadian Census

Mark ROBINSON (University of British Columbia)

Bayesian Neural Networks for Regression -- Some Thoughts on Priors/ Réseaux Neuronaux Bayésiens
Pour La Régression - Quelques Pensées Sur Les Distributions A Priori

Ehsanes SALEH (Carleton University)

Estimation of the Parameters of Multiple Regression Model with Measurement Errors/

Meriem SAÏD (Université Laval)

Testing For Dependent Censoring In A Parametric Survival Model/ Tester Des Données Censurées
Dépendantes Dans Un Modèle Paramétrique De Survie

Bruce SMITH (Dalhousie University)

The Tides, They Are A Changing/ Les Marées Changent

Rolf TURNER (University of New Brunswick)

Misaligned Alignments/ Alignements Mal Alignés

María Inés URRUTIA

Sociocultural Mapping of Groups of Immigrants Settled in the Province of Buenos Aires (Other than
Spanish and Italians)/ CARTOGRAPHIE SOCIO-CULTURELLE DES GROUPES D'IMMIGRANTS
ÉTABLIS DANS LA PROVINCE DE BUENOS AIRES (À L'EXCEPTION DES ESPAGNOLS ET
DES ITALIENS)

Zilin WANG (University of Western Ontario)

On Estimation of A Partial Linear Semiparametric Regression Model in Complex Survey/ L'Estimation
D'Un Modèle De Régression Partielle Linéaire Et Semi-Paramétrique Dans Un Sondage Complexe

Larry WELDON (Simon Fraser University)

Directions in the Evolution of Statistics/

Mahmoud ZAREPOUR (University of Ottawa)

A Recursive Method For Functionals Of Poisson Processes

Xiaoling ZHANG (University of Western Ontario)

Graphic Display Of Orientation Data/ L'Affichage Graphique Des Données D'Orientation

Monday, June 11/Lundi 11 juin

Scientific Programme/Programme scientifique

Monday, June 11th (morning)/Lundi 11 juin (matin)

8:30-10:00

Welcome and SSC Gold Medal Address, Images

Organizer and Chair: Louis-Paul RIVEST (Université Laval)

[8:30 - 10:00] Jim ZIDEK (University of British Columbia)
The Weighted Likelihood/ La vraisemblance pondérée

10:30-Noon

WNAR Presidential Invited Address, Images

Organizer and Chair: Ed BEDRICK (University of New Mexico)

[10:30 - 12:00] Noel CRESSIE (The Ohio State University)
A Bird's Eye View of Earth and its Atmosphere/ Une vue à vol d'oiseau de la
Terre et de son atmosphère

Contrasting US 2000 and Canada 2001 Censuses, Halpern

Organizer and Chair: Patrick FARRELL (Carleton University)

[10:30 - 11:00] Mike BANKIER (Statistics Canada)
Calibration Estimation in the Canadian Census/ Estimations de calibrage dans le
recensement canadien

[11:00 – 11:30] Peter DICK (Statistics Canada)
Small Domain Estimation of Missed Persons in the 2001 Census/ Estimation de
petit domaine des personnes non dénombrées au Recensement 2001

[11:30 – 12:00] Robert FAY (U.S. Census Bureau)
Census 2000 in the U.S.: A Neighborly Overview/ Recensement 2000 aux É.U.:
un survol comparatif

Unusual Stochastic Models, AQ 4150

Organizer and Chair: Priscilla GREENWOOD (Arizona State University)

[10:30 – 11:00] Tom TAYLOR (Arizona State University)
A Stochastic Adding Machine And Complex Dynamics

[11:00 – 11:30] Jeffrey ROSENTHAL (University of Toronto)
Convergence Of Gibbs Samplers For Uniform Distributions/ Convergence des
échantillonneurs de Gibbs pour distributions uniformes

[11:30 – 12:00] Georghe STOICA (St. Johns University of New Brunswick)
Market Completeness: a Return to Order/ Complétude du marché : retour à l'ordre

Monday, June 11 (afternoon)/Lundi 11 juin (après-midi)

1:30-3:00

Session in Honour of V.P. Godambe's 75th Anniversary, Halpern

Organizer and Chair: Mary THOMPSON (University of Waterloo)

Sponsored by: The International Indian Statistical Association

- [1:30 – 2:00] Subhash LELE (University of Alberta)
Smooth Simulated Composite Likelihood (SSCL) based inference for Hierarchical Models/ Inférence basée sur la vraisemblance composée lisse et simulée (VCLS) pour modèles hiérarchiques
- [2:00 – 2:30] M.B. RAJARSHI, (University of Pune)
Abundance Estimation Based On Daily Pellet Heap Counts: An Estimating Function Approach
- [2:30 – 3:00] Harold MANTEL and Milorad KOVACEVIC (Statistics Canada)
Estimating Functions and Inference for Survey Data/ Fonctions d'estimation et inférence pour données de sondages

Spatial Statistics, Images

Organizer and Chair: Fangliang HE (Pacific Forestry Centre)

- [1:30 – 2:00] Julian BESAG (University of Washington)
The Good, The Bad And The Ugly: 30 Years Of Markov Random Fields/ Le bon, la brute et le truand: 30 ans de champs aléatoires de Markov
- [2:00 – 2:30] Charmaine DEAN (Simon Fraser University) and Y.C. MACNAB
Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates
- [2:30 – 3:00] Mark DALE (University of Alberta)
Accounting For Spatial Autocorrelation In Statistical Tests/ L'inclusion de l'autocorrélation spatiale dans les tests statistiques

Robust Inference, AQ 4150

Organizer and Chair: Julie ZHOU (University of Victoria)

- [1:30 – 2:00] Min TSAO (University of Victoria)
On The Robustness Of Empirical Likelihood Ratio Confidence Intervals For Location/ Sur la robustesse d'intervalles de confiance du rapport de vraisemblance empirique pour le paramètre de position
- [2:00 – 2:30] Doug WIENS (University of Alberta)
Robust Sequential Designs for Misspecified Models/ Plans séquentiels robustes pour modèles mal-spécifiés
- [2:30 – 3:00] David TYLER (Rutgers University)
On The Role Of Redescending M-Estimates In Computer Vision (And Other Areas)/ Sur le rôle d'estimateurs M redescendant en vision numérique (et autres domaines)

Contributed Session: Survival Analysis, AQ 4140

Chair: Eric AGEYKUM (University of Victoria)

- [1:30 – 1:45] Marylène TROUPÉ and J. VAILLANT (Université des Antilles-Guyane)
Selection Of Survival Predictors Using A Nonparametric Method/Sélection de prédicteurs de survie à l'aide d'une méthode non paramétrique

- [1:45 – 2:00] Jennifer ORLICH and Susan MURPHY (University of Michigan)
Checking the Fit of a Discrete-time Multi-level Survival Model/ Vérification de l'ajustement d'un modèle de survie à plusieurs Niveaux à temps discrets
- [2:00 – 2:15] Jong Sung KIM (Portland State University)
Efficient Estimation for the Proportional Hazards Model with Left Truncated and “Case1” Interval-Censored Data/ Estimation Efficace Pour le Modèle de Risque Proportionnel Tronqué à Gauche et avec Données Censurées par Intervalle de Type 1
- [2:15 – 2:30] Mark UDEVITZ (Alaska Biological Science Centre)
Estimating Wildlife Survival Rates With Time Series Of Age-Structure Data/ Estimation des taux de survie fauniques à l'aide de séries chronologiques à partir de données par structure d'âge
- [2:30 – 2:45] Xuewen LU (Agriculture and Agri-Food Canada)
A Partially Linear Proportional Hazards Model/ Un modèle de risque proportionnel partiellement linéaire
- [2:45 – 3:00] Jaroslaw HAREZLAK (Indiana University School of Medicine)
Using Auxiliary Information In Interval-Censored Survival And Event History Data Analysis/ Utilisation d'information auxiliaires pour des intervalles de survie censurés et pour l'analyse des données des événements historiques

1:30 – 3:15

Case Studies Session I: Climate Change/ Changements climatiques WMX 3520

Chair: Peter MACDONALD (McMaster University)

- [1:30 - 1:45] **Introduction:** Francis ZWIERS (Canadian Centre for Climate Modelling and Analysis)
- [1:45 - 2:00] Sophia LEE, Baifang XING (York University)
- [2:00 - 2:15] Eishita ALI, Audrey FU, Jafar KHAN, Lisa KURAMOTO, Eugenia YU (University of British Columbia)
- [2:15 - 2:30] Patrick SHERIDAN, Baktiar HASAN, Fuping HUANG, Carlos CINTORA, Frank YIN, Jhona ROSE, Jason WENCZLER (University of Guelph)
- [2:30 - 2:45] Jinnan LIU, Jian XIONG, Peter DAMIANAKOS (York University)
- [2:45 - 3:00] Theodoro KOULIS, Yun-Hee CHOI (University of Waterloo)
- [3:00 - 3:15] Discussant: Francis ZWIERS (Canadian Centre for Climate Modelling and Analysis)

WNAR Student Paper Presentations, AQ 4130

Chair: Laura SALTER (University of New Mexico)

- [1:30 – 2:00] Shenghai ZHANG (University of Waterloo)
Vector-valued Markov Models for Analyzing Disease History Data with Multi-Variate Outcomes and Incomplete Observations/ Modèle markovien vectoriel

pour l'analyse des données historique d'une maladie avec des résultats multidimensionnels et des observations incomplètes

[2:00 – 2:30] Yuanyuan WANG, Hugh CHIPMAN, and William WELCH (University of Waterloo)
Comparison of Statistical Methods for High Throughput Screening Data/
Comparaison des méthodes statistiques pour des données traitées par filtration à haut rendement

[2:30 – 3:00] Jason SUTHERLAND, C.J. SCHWARZ, J.G. FALL, B. DORNER, K.P. LERTZMAN (Simon Fraser University)
Multi-List/Capture-Recapture Methods With Incomplete And Partial Stratification Applied To Estimating Forest Fire Incidence/ Méthodes de listes multiples et de capture-recapture avec stratification partielle ou incomplète appliquées à l'estimation de l'incidence des feux de forêt

3:30-5:00

IMS Special Invited Session I, Halpern

Organizers: IMS Executive

Chair: Robert BURTON

[3:30 – 5:00] Steffen LAURITZEN and Thomas S. RICHARDSON (Aalborg University, Denmark)
Chain Graph Models and Their Causal Interpretations/

Business and Industry Section Special Invited Session, Images

Organizer and Chair: Bovas ABRAHAM (University of Waterloo)

[3:30 – 5:00] Bill MEEKER and Luis A. ESCOBAR (Iowa State University)
Accelerated Testing: A Method for Obtaining Reliability Information Quickly

Gene Ancestry, Identity and Allelic Associations, AQ 4130

Organizer and Chair: Elizabeth THOMPSON (University of Washington)

[3:30 – 4:00] Sandrine DUDOIT (MSRI, Berkeley)
A Score Test For The Linkage Analysis Of Qualitative And Quantitative Traits Based On Identity By Descent Data/ Un test de score pour l'analyse de liaison de traits qualitatifs et quantitatifs basé sur des données d'identité par descendance

[4:00 – 4:30] Laura LAZZERONI (Stanford University)
Distinguishing Between Causative And Linked Genetic Loci/ Distinction entre loci génétiques causaux et liés

[4:30 – 5:00] Matthew STEPHENS (University of Washington)
A New Statistical Method For Haplotype Reconstruction/ Une nouvelle méthode statistique pour la reconstruction de l'haplotype

Survey Methods Contributed Session: Imputation and Variance Estimation, AQ 4150

Chair: Mike HIDIROGLOU

[3:30 – 3:50] David HAZIZA (Statistics Canada)
Inference for Regression Parameters under Imputation for Missing Data/

Inférence sur les paramètres de régression avec imputation des données manquantes

- [3:50 – 4:10] Wesley YUNG, Mike HIDIROGLOU and J.N.K. RAO (Statistics Canada)
Variance Estimation for Estimating Equations in the Presence of Missing Values/
Estimation de la variance pour les équations d'estimation en présence de valeurs manquantes
- [4:10 – 4:30] Wisner JOCELYN and Garriguet DIDIER (Statistics Canada)
Variance Estimation for the Annual Survey of Manufactures/ Estimation de la variance pour l'enquête annuelle sur les manufactures
- [4:30 – 4:50] Charlie ARCARO and Wesley YUNG (Statistics Canada)
Variance Estimation in the Presence of Imputation/ Estimation de la variance lors d'imputation

Contributed Session: Optimal Designs, AQ 4140

Chair: Mary THOMPSON (University of Waterloo)

- [3:30 – 3:45] Alex SMITH, Katherine S. PANAGEAS, Mithat GONEN, Paul B. CHAPMAN (Memorial Sloan-Kettering Cancer Centre)
A Two-Stage Phase II Design with a Multinomial Endpoint/ Un plan de phase II à deux niveaux avec point d'arrêt multinomial
- [3:45 – 4:00] Saumendranath MANDAL, K.C. CARRIERE (University of Alberta)
Constructing Optimal Designs With Constraints/ Construction de plans optimaux avec contraintes
- [4:00 – 4:15] Wei WEI and R.J.COOK (University of Waterloo)
Trial Design For Negative Binomial Responses Based On ANCOVA/ Plan d'expérience pour réponses suivant une binomiale négative basé sur l'ANCOVA
- [4:15 – 4:30] David VERBEL, Jaya M. SATAGOPAN, E.S. VENKATRAMAN, Colin B. BEGG (Memorial Sloan-Kettering Cancer Centre)
Two-Stage Designs For Genome Scanning/ Un plan à deux niveaux pour le scan du génome
- [4:30 – 4:45] Vartan CHOULAKIAN (Université de Moncton), M. A. Stephens (Simon Fraser University)
On Modelling of Exceedances of Riverflows over Thresholds by Generalized Pareto distribution/ Sur la modélisation des excédents de l'écoulement des rivières au-dessus des seuils par la distribution de Pareto généralisée (GPD)
- [4:45 – 5:00] Christian BOUDREAU (University of Waterloo)
Event History Analysis and Ignorable Sampling/ Analyse d'événements historiques et échantillonnage ignorable

Tuesday, June 12 (morning)/Mardi 12 juin (matin)

8:30-10:00

[8:30 – 9:15] Pierre Robillard Award Presentation, Images

Chair: Mike EVANS (University of Toronto)

[9:15 – 10:00] Canadian Journal of Statistics Award Presentation, Images

Chair: David BELLHOUSE (University of Western Ontario)

Measurement Error, Halpern

Organizer and Chair: John SPINELLI (B.C. Cancer Agency)

[8:30 – 9:00] Donna SPIEGELMAN and R. LOGAN (Harvard School of Public Health)

Extended Partial Likelihood Methods For Failure Time Models, With
Measurement Error In A Time-Varying Covariate

[9:00 - 9:30] Dan SCHAFFER (Oregon State University)

Measurement error regression and nonparametric maximum likelihood/
Régression d'erreurs de mesures et maximum de vraisemblance non paramétrique

[9:30 – 10:00] Paul GUSTAFSON (University of British Columbia)

Case-Control Analysis with Partial Knowledge of Exposure Misclassification
Probabilities/ Analyse de cas-témoins avec connaissances partielles des
probabilités de mauvais classement de l'exposition

Survey Methods Contributed Session: Survey Applications, AQ 4150

Chair: Don ROYCE

[8:30 – 8:50] Hélène BÉRARD (Statistics Canada)

The Redesign of the Monthly Wholesale and Retail Trade Survey of Statistics
Canada/ La nouvelle conception de l'enquête mensuelle du commerce en gros et
au détail des Statistique Canada

[8:50 – 9:10] Patricia WHITRIDGE (Statistics Canada)

Challenges in Measuring Electronic Commerce/ Défis dans la quantification du
commerce électronique

[9:10 – 9:30] Yves BÉLAND (Statistics Canada)

Sample Design of the Canadian Mental Health Survey/ Plan échantillonnal de
l'enquête sur la santé mentale canadienne

[9:30 – 9:50] Brad W. THOMAS, Doreen DUCHESNE (STC), Marek WYSOCKI (STC),

Robert BOURBEAU (University of Montréal) (Statistics Canada)
What's New for the 1995-1997 Canada Life Tables/ Quoi de Neuf Pour les
Tableaux de Mortalité 1995-1997 Pour le Canada

WNAR Student Presentations, AQ 4130

Chair: Laura SALTER (University of New Mexico)

[8:30 – 9:00] Misoo ELLISON (University of Colorado Health Sciences Centre/National Jewish
Medical and Research Centre)

Responder Cell Frequency Estimation and Binomial Three-level Nonlinear Mixed Effects Model in Limiting Dilution Assays/

[9:00 – 9:30] Amalia MEIER and Barbra RICHARDSON (University of Washington)
Uncertainty In Outcomes In Discrete Time Proportional Hazards Modelling/
Incertitude des résultats pour la modélisation des risques proportionnels à temps discret

[9:30 – 10:00] Melissa DOBBIE and A.H. WELSH (Australian National University)
Models For Zero-Inflated Count Data Using The Neyman Type A Distribution

10:30-12:00

On Becoming a Successful Statistician: A Panel Discussion, Images

Organizer: WNAR Advisory Board

Chair: Betty NELSON (Genentech, Inc.)

John CROWLEY (Southwest Oncology Group, Statistical Centre)

Abe Sue RUBIN (Immunex Corporation)

Elizabeth A. THOMPSON (University of Washington)

Small Area Estimation, AQ 4150

Organizer and Chair: John ELTINGE (US Bureau of Labor Statistics and Texas A&M)

[10:30 – 11:00] David SMITH, G.S. DATTA and J.N.K. RAO (University of Georgia)
Second Order Asymptotics in Small Area Estimation

[11:00 – 11:30] Narasimha PRASAD and Sharon LOHR (University of Alberta)
Small Area Estimation using Multiple Surveys

[11:30 – 12:00] John ELTINGE (U.S. Bureau of Labor Statistics and Texas A&M)
Development of Small Domain Estimators for the U.S. Current Employment Statistics Program

Likelihood, Halpern

Organizer and Chair: Cynthia STRUTHERS (University of Waterloo)

[10:30 – 11:00] Susan MURPHY (University of Michigan)
Likelihood Inference, High Dimensional Information and Missing at Random/
Inférence basée sur la vraisemblance, information à haute dimension et manquante de façon aléatoire

[11:00 – 11:30] Nan LAIRD (Harvard School of Public Health)
Likelihood Models for Assessing Familial Aggregation/ Modèles de vraisemblance pour l'évaluation de l'agrégation familiale

[11:30 – 12:00] Nancy REID (University of Toronto)
Adjustments To Profile Likelihood/Ajustements au profil de vraisemblance

Statistics in Industry and Technology, AQ 4130

Organizer and Chair: Randy SITTER (Simon Fraser University)

[10:30 – 11:00] Jock MACKAY and Stefan STEINER (University of Waterloo)
Some Simple Research Problems from Industrial Consulting

[11:00 – 11:30] Wei LI, Shaowei CHENG, Jack HU and Justin SHRIVER (University of Washington)
Statistical Investigation of Resistance Spot Welding Quality Using a Two-Stage, Sliding Level Experiment

[11:30 – 12:00] Derek BINGHAM and William LI (University of Michigan)
Non-Regular Robust Parameter Design

Contributed Session: Timeseries, AQ 4140

Chair: Bruce SMITH (Dalhousie University)

[10:30 – 10:45] Jeff TERPSTRA (North Dakota State University)
Generalized Signed Rank Autoregression/ Autorégression généralisée des rangs signés

[10:45 – 11:00] Ying ZHANG and A.I. MCLEOD
The Exact Maximum Likelihood Estimate For Stationary Normal AR (1) Processes and Its Applications in Unit Root Test/ L'estimateur du maximum de vraisemblance exact pour les processus stationnaires normaux AR(1) et applications au test pour la racine unité

[11:00 – 11:15] Ante ROZGA (University of Split)
Applying Seasonal Adjustment Methods to Difficult Time Series: Croatian Case/ Application de méthodes d'ajustement saisonnier à des séries chronologiques difficiles : le cas croate

[11:15 – 11:30] Don MCLEISH (University of Waterloo)
Financial Ups and Downs: Some properties of the Extremes of Stochastic Processes/ Les fluctuations financières : Quelques propriétés des processus stochastiques extrêmes

[11:30 – 11:45] Pierre DUCHESNE and Yongmiao HONG (Ecole des Hautes Etudes Commerciales)
Detecting Duration Clustering and Adequacy of Autoregressive Conditional Duration Models Using Wavelets/ Détection des grappes de durée et adéquation des modèles conditionnels auto-régressifs de durée en utilisant les ondelettes

Tuesday, June 12 (afternoon)/Mardi 12 juin (après-midi)

1:30-3:00

Session in Honour of J.N.K. Rao's Retirement, Images

Organizer and Chair: Patrick FARRELL (Carleton University)

[1:30 – 2:00] David BELLHOUSE (University of Western Ontario)
Jon Rao: an Appreciation of His Research Contributions/ Jon Rao: survol de ses contributions en recherche

[2:00 – 2:30] Randy SITTER and Jeff WU (Simon Fraser University)
Resampling Methods in Complex Surveys: An Overview/ Méthodes de
rééchantillonnage pour sondages complexes : un survol

[2:30 – 3:00] Mike HIDIROGLOU (Statistics Canada)
Contributions To Survey Methodology At Statistics Canada By J.N.K. Rao/
Contributions faites par J.N.K. Rao à la méthodologie de sondage à Statistiques
Canada

Bioinformatics Tutorial I, Halpern

Organizers: Anna BARON, Imran SHAH, Larry HUNTER (University of Colorado)
Chair: Anna BARON (University of Colorado)

[1:30 – 3:00] Larry HUNTER (University of Colorado)
Inference In High-Throughput Molecular Biology/ Inférence en biologie
moléculaire à haut débit

Discrete Probability and Combinatorics, AQ 4150

Organizer and Chair: J. Michael STEELE (University of Pennsylvania)

[1:30 – 2:15] Ken ALEXANDER (University of Southern California)
Exponential Decay And Mixing Properties For Lattice Systems in Finite
Volumes/ Décroissance exponentielle et propriétés de mélange

[2:15 – 3:00] Gordon SLADE (University of British Columbia)
Critical Two-Point Functions For Self-Avoiding Walk, Percolation, Lattice Trees
And Lattice Animals/ Fonctions critiques à deux points pour marche s'auto-
évitant, percolation, arbres en treillis et animaux en treillis

3:30-5:00

IMS Special Invited Session II, Images

Organizers: IMS Executive
Chair: Elizabeth THOMPSON

[3:30 – 5:00] Simon TAVARE (University of Southern California)
Interpreting Genomic Polymorphism Data: What History Has To Tell/

Bioinformatics Tutorial II, Halpern

Organizers: Anna BARON, Imran SHAH, Larry HUNTER (University of Colorado)
Chair: Anna BARON (University of Colorado)

[3:30 – 5:00] Imran SHAH (University of Colorado)
Predicting Metabolic Networks/ Prévission de réseaux métaboliques

Canadian Initiative on Social Statistics, AQ 4150

Organizer and Chair: Georgia ROBERTS (Statistics Canada)
Perspectives on the Canadian Initiative on Social Statistics

[3:30 – 4:00] Doug NORRIS (Statistics Canada)
Overview of CISS and Statistics Canada's Involvement

Tuesday, June 12/Mardi 12 juin

Scientific Programme/Programme scientifique

[4:00 – 4:30] Paul BERNARD (Université de Montréal)
Perspective Of The Sociologist Involved In CISS

[4:30 – 5:00] Mary THOMPSON (University of Waterloo)
Perspective Of The Statistician Involved In Social Science Research

Wednesday, June 13 (morning)/Mercredi 13 juin (matin)

8:30-10:00

SSC Presidential Invited Address, Images

Organizer and Chair: Louis-Paul RIVEST (Université Laval)

[8:30 – 10:00] Keith WORLSEY (McGill University)

The Geometry Of Random Images In Astrophysics And Brain Mapping.

10:30-12:00

Statistics in Fisheries I, Images

Organizer and Chair: Rick ROUTLEDGE (Simon Fraser University)

[10:30 – 11:00] Chris WOOD (Fisheries and Oceans Canada)

Defining Conservation Units Based On Capacity For Local Adaptation: Testing Predictions Of A Heuristic Model./ Définition d'unités de conservation basées sur la capacité d'adaptation locale : vérification des prédictions d'un modèle heuristique

[11:00 – 11:30] Ashley STEEL, Blake E. FEIST, George R. PESS, and Robert E. BILBY (U. S. National Marine Fisheries Service)

A Hierarchical Statistical Model To Link Spatially Explicit Salmon Abundance Records To Habitat Characteristics In The Salmon, Snohomish, And Willamette River Basins./ Un modèle statistique hiérarchique pour lier des données d'abondance du saumon spatialement explicites aux caractéristiques de l'habitat des bassins des rivières Salmon, Snohomish et Willamette

[11:30 – 12:00] Eric ANDERSON (University of Washington)

Bayesian Methods For Inference In Populations Of Hybridizing Fish./

Nonparametric Econometrics, AQ 4130

Organizer: Nilanjana ROY (University of Victoria)

Chair: Min TSAO (University of Victoria)

[10:30 – 11:15] Joris PINKSE (University of British Columbia)

Nonparametric Regression Estimation Using Weak Separability

[11:15 – 12:00] Thanasis STENGOS (University of Guelph)

Economic Development and the Return to Human Capital/ Développement économique et les retours au capital humain

Contributed Session: Longitudinal and Correlated Data Analysis, AQ 4140

Chair: David HAMILTON (Dalhousie University)

[10:30 – 10:45] Ying MACNAB (University of British Columbia)

Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates

[10:45 – 11:00] Grace YI and Richard J. COOK (University of Waterloo)

Inverse Probability Weighted Second Order GEE for Clustered Data with Intermittently Missing Observations/ GEE de second ordre pour des données en grappes avec des observations manquantes intermittentes

- [11:00 – 11:15] Hongtu ZHU, Mingguo GU, Julie ZHOU, Fanliang HE (University of Victoria)
Maximizing Spatial Likelihood And Its Applications/ Maximisation de la vraisemblance spatiale et ses applications
- [11:15 – 11:30] Janet TOOZE, Gary K. GRUNWALD, Richard H. JONES (National Cancer Institute)
Analysis of Repeated Measures Data with Clumping at Zero/ Analyses de données à mesures répétées avec groupement à zéro
- [11:30 – 11:45] Zhenguo QIU, Peter X.-K. SONG and Ming TAN (York University)
Simplex Mixed Models For Longitudinal Proportional Data/ Modèles de simplexe mixte pour des proportions expérimentales longitudinales
- [11:45 – 12:00] Claudia CZADO (Munich University of Technology) and Peter X.-K. Song (York University)
State Space Mixed Models for Binomial Longitudinal Observations/ Modèles mixtes de l'espace d'état pour des observations longitudinales binomiales

Contributed Session: Inference, AQ 4120

Chair: Gerarda DARLINGTON (University of Guelph)

- [10:30 – 10:45] Edit GOMBAY (University of Alberta)
Sequential Testing Of Composite Hypotheses/ Test séquentiel pour des hypothèses composées
- [10:45 – 11:00] Hari MUKERJEE and H. El BARMÍ (Wichita State University)
Consistent Estimation Of Distributions With Type II Bias/ Estimation convergente des distributions avec un biais de type II
- [11:00 – 11:15] R. KARUNAMUNI and S. ZHANG (University of Alberta)
On Semiparametric Estimates of Population Density in Line Transect Sampling/ Sur l'estimation semi-paramétrique de la densité d'une population dans un échantillonnage en lignes transversales
- [11:15 – 11:30] Alexandre LEBLANC and Jean-François Angers (Université de Montréal)
Prior Density Estimation via Haar Deconvolution/ Estimation de la densité a priori via la déconvolution de Haar
- [11:30 – 11:45] Humberto VAQUERA, Jose VILLASEÑOR and B.C. ARNOLD (ITESM-CP)
A Statistical Test for Comparing Poisson, Negative Binomial, and Binomial Populations/ Un test statistique pour comparer des populations binomiale, Poisson et binomiale négative
- [11:45 – 12:00] Jiahua CHEN, Hanfeng CHEN, Jack KALBFLEISCH (University of Waterloo)
Testing The Number Of Components Of Finite Mixture Model of Finite Mixture Models With Modified Likelihood/ Test pour le nombre de composantes de modèles mixtes finis pour des modèles mixtes finis avec vraisemblance modifiée

Case Studies Session II: Old Growth Forest/En forêt ancienne, Halpern

Chair Peter MACDONALD (McMaster University)

[10:30 - 10:45] Introduction: Fangliang HE (Pacific Forestry Centre)

[10:45 - 11:00] Ying LUAN, Juqing Ellen SHI, Victor NAWA (University of Waterloo)

[11:00 - 11:15] Scott BRADFIELD, Amos CHONG, Juan DU, Shira DUNSIGER, Jourdan GOLD, Jeong-ae KIM, Susanne Macmahon, Hong QIAN, Eddy REMPEL (McMaster University)

[11:15 - 11:30] Geneviève LEFEBVRE, Annie LEVERT, Urs MAAG, (Université de Montréal)

[11:30 - 11:45] Andrew WILLIAMS (Carleton University)

[11:45 - 12:00] Discussant: Fangliang HE (Pacific Forestry Centre)

WNAR Student Presentations, AQ 4150

Chair: Laura SALTER (University of New Mexico)

[10:30 - 11:00] Mireya DIAZ and J. Sunil RAO (Case Western Reserve University)
Mammographic Computer-Aided Diagnosis Using Bootstrap Aggregation/ Diagnostic mammaire assisté par ordinateur à l'aide d'agrégation de rééchantillonnage

[11:00 – 11:30] Xiaoqiang LI and Duncan MURDOCH (University of Western Ontario)
Perfect Simulation of Stochastic Differential Equations/ Simulation parfaite d'équations différentielles stochastiques

[11:30 - 12:00] Lucia MIREA and Shelley B. BULL (Samuel Lunenfeld Research Institute, University of Toronto)
Detection of Genetic Heterogeneity Among Families with Affected Relatives/
Détection de l'hétérogénéité génétique au sein de familles avec une parenté infectée

Wednesday, June 13 (afternoon)/Mercredi 13 juin (après-midi)

1:30-3:00

Statistics in Forestry, Halpern

Organizer and Chair: Vera SIT (B.C. Ministry of Forests)

[1:30 - 2:00] Gay BRADSHAW and Jeffrey BORCHERS (USDA Forest Service)
Bayesian Statistics To Make Complex Decisions In Natural Resources/
Statistiques bayésiennes pour des décisions complexes relatives aux ressources naturelles

[2:00 – 2:30] William REED and Kevin S. MCKELVEY (University of Victoria)
The Size Distribution Of Wildfires/

[2:30 – 3:00] Wendy BERGERUD (B.C. Ministry of Forests)
Silviculture Survey System: Statistical Properties and Anticipated Yields/
Système de sondage en sylviculture : propriétés statistiques et production anticipée

Statistics in Fisheries II, Images

Organizer and Chair: S. SMITH (Fisheries and Oceans Canada)

- [1:30 – 2:00] Zane Z. ZHANG (Department Fisheries & Oceans, Nanaimo)
Ideal Survey Designs for Various Spatial Distributions of Benthic Invertebrate Populations/ Plans de sondages idéaux pour diverses distributions spatiales d'invertébrés de fonds marins
- [2:00 – 2:30] Terry QUINN (University of Alaska, Juneau)
Adaptive Cluster Sampling of Gulf of Alaska Rockfish/ Échantillonnage en grappe adaptatif du sébaste du Golfe de l'Alaska
- [2:30 – 3:00] Carl SCHWARZ (Simon Fraser University)
Assessing Fish Population With Capture-Recapture: A How-To-Guide/ La détermination de population de poissons utilisant la capture-recapture: un guide pratique

Innovative Applications of Statistics in Business, AQ 4150

Organizer: Marty PUTERMAN (University of British Columbia)

Chair: Mike EVANS (University of Toronto)

- [1:30 – 2:00] Ernest URBANOVICH, Ella E. YOUNG, Martin L. PUTERMAN, Sidney O. FATTEDAD (Workman's Compensation Board)
Early Detection of High Risk Claims at the Workers Compensation Board of British Columbia
- [2:00 – 2:30] Jason GOTO, Isabelle SMITH, Kelly CHUNG and Hong CHEN (Centre for Operations Excellence)
Analyzing Telecommunications Network Call Traffic Flow At Telus
- [2:30 – 3:00] Remi DESMEULES (University of British Columbia)
Dynamic Linear Models for Motion Pictures Box-Office Forecasting

Recent Results on Normal and Related Approximations, AQ 4140

Organizer and Chair: Rabi BHATTACHARYA (Indiana State University)

- [1:30 - 2:00] V. BENTKUS (Vilnius Institute of Mathematics)
An Inequality For Tail Probabilities Of Martingales With Bounded Differences
- [2:00 – 2:30] Susan HOLMES and Gesine REINERT (Stanford University)
Stein's Method For Computing Error Bounds
- [2:30 – 3:00] Soumendra N. LAHIRI (Iowa State University)
Asymptotic Expansions For Studentized Statistics Under Weak Dependence

Block Designs: Directions and Inspirations, AQ 4130

Organizer: Dibyen MAJUMDAR (University of Illinois at Chicago) and John MORGAN (Virginia Tech)

Chair: Dibyen MAJUMDAR (University of Illinois at Chicago)

- [1:30 - 1:35] **Introductory Remarks:** John MORGAN (Virginia Tech)
- [1:35 – 1:55] Brian RECK (Old Dominion University) and J. P. Morgan (Virginia Tech)
Block Design Optimality: Results for Combinatorially Problematic Settings/
Optimalité des designs par bloc : résultat pour le cadre combinatoire

- [1:55 – 2:15] M. JACROUX (Washington State University)
A- And MV-Efficient Designs For Comparing A Set Of Test Treatments To A Set Of Controls/
- [2:15 – 2:35] John STUFKEN (Iowa State University)
Polygonal Designs/ Plans polygonaux
- [2:35 – 3:00] W. NOTZ (Ohio State University)
Design Of Computer Experiments To Optimize The Mean Of A Response Subject To Constraints

Survey Methods Contributed Session: Estimation 1, AQ 4120

Chair: Milorad KOVACEVIC (Statistics Canada)

- [1:30 – 1:50] Rubin BLEUER(Statistics Canada)
A Test for Survival Distributions using Data from Complex Samples/ Un test pour des fonctions de survie utilisant des données d'échantillons complexes
- [1:50 – 2:10] Georgia R. ROBERTS (Statistics Canada)
Complexities in Analysis of Duration Data from Longitudinal Surveys - with Application to the Survey of Labor and Income Dynamics/ Complexités dans l'analyse de données de durée dans des enquêtes longitudinales - avec application à l'étude sur la dynamique du travail et du revenu
- [2:10 – 2:30] Asma ALAVI (Statistics Canada)
On the Treatment of Influential Observations in Household Surveys/ Sur le traitement des observations influente pour l'enquête sur les ménages
- [2:30 – 2:50] Steven COHEN (Agency for Healthcare Quality and Research)
The Accuracy of a Prediction Model to Oversample Individuals with High Levels of Medical Expenditures in the Medical Expenditure Panel Survey/ L'exactitude d'un modèle de prévision sur des individus suréchantillonnés ayant des niveaux élevés de dépenses médicales dans l'enquête sur les dépenses médicales

3:30-5:00

Statistics in Fisheries III, Images

Organizer and Chair: S. SMITH (Fisheries and Oceans Canada)

- [3:30 – 4:00] Sherri DRESSEL (University of Alaska, Fairbanks)
Using Poststratification to Incorporate Habitat Information into Abundance Estimates from Multispecies Surveys/ Utilisation de la post-stratification pour l'incorporation d'information sur l'habitat pour des estimations d'abondance lors de sondages comportant plusieurs espèces
- [4:00 – 4:30] Brenda NORCROSS (University of Alaska, Fairbanks)
Sampling Problems for Which Textbooks Did Not Prepare You/ Problèmes d'échantillonnages auxquels les livres ne nous ont pas préparés
- [4:30 – 5:00] Steven SYRJALA (NMFS, Seattle)
A Bootstrap Approach to Make Sample-Size Calculations for Trawl Studies/ Une approche bootstrap pour études au chalut servant à effectuer des calculs de taille d'échantillon

Brownian Sheets and Stochastic Partial Differential Equations, AQ 4140

Organizers: KHOSHNEVISAN (University of Utah) and John B. WALSH (University of British Columbia)

Chair: Robert BURTON (Oregon State University)

[3:30 – 4:15] Robert DALANG (Ecole Polytechnique Federale, Switzerland)
Some Recent Results Of Interest

[4:15 - 5:00] Ed PERKINS (University of British Columbia)
Symbiotic branching

Survey Methods Contributed Session: Estimation 2, AQ 4150

Chair: Karla NOBREGA

[3:30 – 4:00] Changbao WU and Randy R. SITTER (University of Waterloo)
Efficient Estimation of Quadratic Finite Population Functions in the Presence of Auxiliary Information/ Évaluation efficace des fonctions quadratiques de population finies en présence d'information auxiliaire

[4:00 – 4:30] Ioana SCHIOPU-KRATINA (Statistics Canada)
Existence and Consistency of GEE Estimators - Application to Design-based Inference/Existence et uniformité des estimateurs EEG. Application à l'inférence basée sur un plan d'expérience

[4:30 – 5:00] Ian CAHILL (Statistics Canada)
Benchmarking Parameter Estimates in Logit Models of Binary Choice/
Évaluations de paramètres par calibrage dans des modèles logit à choix binaire

NSERC Grant Selection Committee - Open Meeting, Halpern

Contributed Session: Regression, AQ 4130

Chair: Harold MANTEL (Statistics Canada)

[3:30 – 3:45] Jean-Francois ANGERS (Université de Montréal) and Mohan DELAMPADY, (ISI-Bangalore)
Bayesian Nonparametric Regression Using Wavelets/Régression non paramétrique bayésienne à l'aide d'ondelettes

[3:45 – 4:00] Peiming WANG (Nanyang Technological University)
A Markov Switching Negative Binomial Regression Model and Its Applications/
Un modèle de régression binomiale négative transitif de Markov et ses applications

[4:00 – 4:15] Cynthia STRUTHERS and D. L. MCLEISH (St. Jeromes University and University of Waterloo)
Estimation of Regression Parameters with Missing Covariates/ Estimation des paramètres de régression avec des covariables manquantes

[4:15 – 4:30] Grace CHIU, Richard ROUTLEDGE and Richard LOCKHART (Simon Fraser University)

Abrupt Change or Not? -- From the Broken Stick to the Bent Cable/ Changement brusque ou pas? -- « Du bâton rompu au câble courbé »

[4:30 – 4:45] Renjun MA (University of New Brunswick)
A Paired Poisson Modelling Approach to Binomial Mixed Models/ Une approche avec un modèle de Poisson apparié pour l'analyse de modèle mixtes binomials

[4:45 – 5:00] Bingshu CHEN and Richard COOK (University of Waterloo)
Event History Regression Models for Interval Censored Covariates/ Modèles de régression sur des événements historiques pour les covariables à intervalle censuré

Contributed Session: Assorted, AQ 4120

Chair: William REED (University of Victoria)

[3:30 – 3:45] Shahajahan KHAN (University of Southern Queensland)
The Role of the Shape Parameter for the Shrinkage Estimators of the Mean/ Rôle du paramètre de forme pour les estimateurs à rétrécisseur de la moyenne

[3:45 – 4:00] Hong GU and Hugh CHIPMAN (University of Waterloo)
Interpretable Dimension Reduction/ Interprétation de la réduction de la dimension

[4:00 – 4:15] Tina LEVESQUE (Université Laval)
Reduction Of The Bias For The Estimator Of A Population Size/ Réduction du biais de l'estimateur de la taille d'une population animale

[4:15 – 4:30] Liqun WANG and James C. FU (University of Manitoba)
A Noniterative Monte Carlo Method for Multivariate Sampling and Integration/
Une méthode de Monte Carlo non itérative pour l'échantillonnage multidimensionnel et l'intégration

[4:30 – 4:45] Zhenlin YANG (National University of Singapore)
Normal Transformation-Based Percentile Limits for Lifetime Distributions/
Limites percentiles basées sur des transformations normales pour des distributions de durée de vie

[4:45 – 5:00] Susan MIKULICH, Gary O. ZERBE and Thomas J. CROWLEY (University of Colorado)
On Linear and Nonlinear Mixed Model Approaches to Cosinor Analysis/
Approches linéaires et non linéaires de modèle mixte à l'analyse de « cosinor »

Thursday, June 14 (morning)/Jeudi 14 juin (matin)

8:30-10:00

Canadian Journal of Statistics Read Paper and Discussion, Images

Organizer and Chair: Christian GENEST (Laval University)

Peter HOOPER (University of Alberta)

Flexible Regression Modelling With Adaptive Logistic Basis Functions.

Environmental Toxicology, Halpern

Organizer and Chair: Sylvia ESTERBY (Okanogan University College)

[8:30 – 9:00] John HUBERT (University of Guelph)

Models and Methods for Time-Dependent Bioassays

[9:00 – 10:00] Jim SHERRY and Abdel EL-SHAARAWI (National Water Research Institute)

Statistical Issues in Canadian Pulp and Paper Environmental Effects

Contributed Session: Assorted, AQ 4150

Chair: Bruce SMITH (Dalhousie University)

[8:30 – 8:50] Ivan MIZERA (University of Alberta)

Robustness Of Depth-Based Procedures/ Robustesse des procédures basées sur la profondeur

[8:50 – 9:10] Bernard RACHE, M. Abrahamowicz and A.J. Sasco

Flexible Estimation Of Lag Distribution Of The Effect Of Instantaneous Exposures/ Estimation flexible d'une distribution à décalage de l'effet d'expositions instantanées

[9:10 – 9:30] Jane GENTLEMAN and Gary CATLIN (National Centre for Health Statistics)

Collaborative Activities Between Canadian And U.S. Governmental Health Statistics Organizations/ Activités de collaboration entre les organismes gouvernementaux de statistiques sur la santé canadien et américain

[9:30 – 9:50] Wenyu JIANG and John D. KALBFLEISCH (University of Waterloo)

An Examination of the Bootstrap in the Cox Relative Risk Model/ Une examen du Bootstrap pour le modèle de risque relatif de Cox

10:30-12:00

Nonlinear Timeseries, Halpern

Organizer and Chair: Gemai CHEN (University of Manitoba)

[10:30 – 10:50] Dag Bjarne TJOSTHEIM (University of Bergen, Norway)

Nonlinear And Nonstationary Dependence Concepts In Time Series/ Concepts de dépendance non linéaire et non stationnaire en séries chronologiques

[10:50 – 11:10] Reg J. KULPERGER and Hao YU (University of Western Ontario)

Nonlinear Time Series: Some Aspects Of Testing Model Adequacy And

Detection/ Séries chronologiques non linéaires: quelques aspects pour la détection de modèles et tester leur adéquation

[11:10 – 11:30] Wai Keung LI (Hong Kong University)

Varying Parameter Time Series Models/ Modèles de séries chronologiques à paramètres variables

[11:30 - -11:50] Min CHEN (Chinese Academy of Sciences)

Nonparametric Tests of Conditional Heteroscedasticity in Time Series/ Tests non paramétriques pour l'hétéroscédasticité conditionnelle dans les séries chronologiques

Diagnostic Testing, Images

Organizer and Chair: Tim HANSON (University of New Mexico)

[10:30 – 11:00] Wesley O. JOHNSON (University of California-Davis)

Models, Methods and Inferences for Prevalence and Test-Accuracy Based on Imperfect-Dependent Screening Tests/ Modèles, méthodes et inférences pour la prévalence et la précision de tests basés sur des tests de dépistage imparfaits-dépendants

[11:00 – 11:30] Ruth PFEIFFER (National Cancer Institute)

Efficiency Of DNA Pooling To Estimate Joint Allele Frequencies And Measure Linkage Disequilibrium/ Efficacité de la mise en commun d'ADN pour l'estimation de fréquences conjointes d'allèles et pour la mesure de déséquilibre de liaison

[11:30 – 12:00] Nandini DENDUKURI (St. Mary's Hospital)

Bayesian Sample Size Estimation For Studies Involving Two Diagnostic Tests/ Estimation bayésienne de taille d'échantillon pour études impliquant deux tests diagnostiques

Contributed Session: Assorted, AQ 4150

Chair: William REED (University of Victoria)

[10:30 – 10:45] Xiaoli HOU (ASA)

Confidence Interval for Dose Response Covers Subject to Certain Increase/ Intervalle de confiance pour la dose sujet à une certaine augmentation de la réponse à cette dose

[10:45 – 11:00] Otto MOESCHLIN

Traffic Control Based On An Ergodicity Concept/ Contrôle du trafic basé sur un concept d'ergodicité

[11:00 – 11:15] Xikui WANG (University of Manitoba)

Structural Properties Of A Bandit Process/ Propriétés de structure d'un processus de bandit

[11:15 – 11:30] Brad BIGGERSTAFF (US Centres for Disease Control and Prevention)

Estimation Of A Bernoulli Probability From Pooled Samples With Application To Virus Infection Rates In Mosquitoes/ Estimation d'une probabilité de Bernoulli à partir d'échantillons groupés avec des applications aux taux d'infection par un virus transmis par les moustiques

- [11:30 – 11:45]** Karen LEFFONDRÉ, J. LELLOUCH, C. COM-NOUGUÉ, and T. MOREAU
(McGill University Health Centre)
Optimality Of Nonparametric Tests For Association Between Survival Time And
Continuous Covariates/ Optimalité des Tests Non Paramétriques Pour
L'Association Entre le Temps de Survie et les Covariables Continues
- [11:45 – 12:00]** Eric AGYEKUM, C. LUDGATE et al (University of Victoria)
Neo-Adjuvant Hormone Therapy + Radiation For T3, Margin Positive Prostate
Cancer

Poster Session

Luc ADJENGUE (École Polytechnique de Montréal)

An Estimation Procedure for the Parameters of a Partially Observable Markov Decision Process

A class of partially observable Markov decision Processes (POMDP) is considered. Such processes are defined with two stochastic processes: the core process and the observation process.

The core process is a Markov process with a finite set of states that are not directly observable. A core process can, for instance, represent a production process (a machine or a series of machines) which at times, fails to produce items that are as good as expected. The state of the production process remains unknown until the production cycle is stopped and a thorough inspection is carried out. The observation process represents the observable output products and is supposed to follow a given probability distribution. The objective is to estimate the parameters of the transition probability matrix of the core process and the parameters of the probability distribution, given a sample of the observation process.

The estimation procedure is based on the EM (Expectation-Maximization) algorithm (Dempster et al. 1977). Various probability distributions are considered including those of the continuous type. A simulation study allows to analyze the estimates obtained for different sample sizes. Numerical examples of the procedure are also presented.

Ejaz S. AHMED (University of Regina)

Inference Procedures for Process Capability Indices in Arbitrary Populations

The simultaneous estimation of process capability indices are considered when random samples are drawn from arbitrary populations. Further, statistical procedures are compared for testing the homogeneity of process capability indices. One of the procedures is based on the estimated large sample variance. The other is based on the assuming a common known value of the index.

A log-transformation is also proposed for the testing purposes. Large sample properties of the proposed methods are obtained. A simulation study is carried out to appraise the performance of the suggested methods for moderate sample.

Paul D. BELL and Timothy E. O'BRIEN (Loyola University Chicago)

PREDICTING ATTENDANCE AT MAJOR LEAGUE BASEBALL GAMES

This poster provides predictive regression models for game attendance for four Major League Baseball teams. Independent variables include day of the week, climate conditions, date, opponent, game time, and current success of the respective teams. Noteworthy findings include the mixed significance of weather in predicting attendance, the high increase of attendance when certain teams visit, and a parabolic pattern for attendance versus date. Although the predictive models are based on the figures of the 2000 baseball season, they have been tested for the 2001 season with notable success. Ramifications of this study include accurate predictions of ticket and concession sales and of staffing needs of stadium personnel, local police, businesses, and restaurants.

David R BELLHOUSE (University of Western Ontario)

The Reverend Thomas Bayes FRS: a Biographical Sketch

The year 2001 probably marks the 300th anniversary of the birth of Thomas Bayes of Bayes Theorem fame. This biography is written in celebration of this anniversary. In writing scientific biography Karl Pearson stated, "it is impossible to understand a man's work unless you understand something of his character and unless you understand something of his environment. And his environment means the state of affairs social and political of his own age." In this spirit, the main events of Bayes's life and work are covered as well as the scientific and theological background in which he worked. The current biography contains well-known material as well as newer information not contained in previously published biographies of Bayes.

Willard J. BRAUN (University of Western Ontario)

Kernel Spline Regression

A method of combining spline regression and kernel smoothing is proposed. A simulation study shows that the resulting curve estimators can succeed in capturing the main features of the true curve more effectively than local polynomial regression when the curvature rapidly changes. Kernel splines also exhibit somewhat less sensitivity to choice of bandwidth than local polynomial regression. In many situations, knots can be chosen manually. This is an advantage over variable bandwidth and variable order approaches in kernel polynomial regression where a manual approach to bandwidth or order selection is not often possible. When automatic knot selection is required, the knot deletion method seems to work well. A

direct-plug-in bandwidth selector is available as well. Some illustrative examples will also be presented.

David R BRILLINGER (University of California, Berkeley)
Process analysis of extreme flow rates along the Mississippi River

This paper considers the case of stationary point processes and their use in a setup that is reasonably well understood physically. The situation is that of the passage of the extreme water flows through a series of locks along the Mississippi River. The focus is on an investigation of the validity of partial coherency analysis, an analog of partial correlation analysis. A maximum likelihood analysis is also presented.

Noel G CADIGAN (Fisheries and Oceans)
Generalized local influence, with application to Sequential Population Analysis

The purpose of this presentation is to describe a general methodology for measuring local influence that is particularly useful for complex models. An example of this type of model is Sequential Population Analysis (SPA). SPA is an important model used to provide abundance information for many major commercial fisheries worldwide. With SPA it is not always easy to understand how data and modelling assumptions influence results, or why different assumptions produce different results; however, fish stock assessment scientists require this information to assess the reliability of the model used to estimate stock size. The basic approach we use to measure influence is to assess how small perturbations of SPA inputs affect important SPA outputs. Our main contribution is to show how this information can be obtained in routine applications. An example is presented that illustrates the practical utility of the diagnostics.

Hugh CHIPMAN (University of Waterloo)
Additive models for survey data via penalized least squares

The use of penalized least squares in the sampling context is studied as a tool for the analysis of a general trend in a finite population. We focus on smooth regression with a normal error model. This model is complicated in the sampling context by the design of the survey, which could include stratification, cluster sampling, and other complex survey designs. Ties in covariates abound for large scale surveys resulting in the application of scatterplot smoothers to means. The estimation of smooths (for example smoothing splines) is seen to depend on the sampling design only via the sampling weights, meaning that standard software can be used for estimation. Inference for these curves is more challenging, due to correlations induced by the sampling design. We propose and illustrate tests which account for the sampling design. Illustrative examples are given using the Ontario health survey, including scatterplot smoothing, additive models, model diagnostics, and various significance tests.

Laura L COWEN (Simon Fraser University)
Adjusting survival estimates using radio-tag failure curves.

Estimating survival rates and testing hypotheses concerning survival process are essential to understanding the dynamics of animal populations. These are often studied using mark-recapture methods where animals are marked with individually identifiable tags, and released. Subsequent recaptures are recorded and estimates of survival and catchability can be obtained. Radio-tags have been used previously in mark-recapture studies, but only recently have been used in the study of salmon survival. However, unlike past work with radio telemetry in capture-recapture problems occur when fish cannot be detected due to tag failure. Tag life may have a profound effect on survival estimates if tag failure occurs in the middle of the study. In some cases, information is available on the lifetime of the radio-tags from a subset of animals kept to monitor tag lifetimes. Given the tag failure curve, adjustments can be made to the known detections to account for the proportion of the tags that could not be detected because a portion of the tags were no longer active. The distribution of time to tag failure can be fit non-parametrically using Kaplan -Meier estimates, or parametrically using, for example, a Weibull distribution. Capture history data can then be adjusted using the fitted distribution of time to failure to account for fish undetected due to tag failure. The methodology to be developed will be quite general allowing for different types of radio tags with different time-to-failure distributions and several cohorts of tagged fish following the general modelling approach of Lebreton et al, 1992.

Monica L. DRAGOMIR (University of Toronto)
Wavelet Regression for Conditional Medians

Classical wavelet methods are naturally suited to estimating conditional mean function of a response. However, in many problems the conditional median response may be a more useful function to estimate. The extensions of classical methods to estimating conditional medians (or conditional quantiles) are conceptually straightforward but are computationally more difficult. An algorithm is proposed to use wavelets for conditional median estimation, and some of its properties are studied.

James E DUNN (University of Arkansas)
Variations on multinomial logit models: Alternative links and stepwise selection

Users of standard statistical software, e.g., SAS procedure CATMOD, will recognize these deficiencies: (1) for multinomial, unordered categories of response, only use of baseline logit link functions admits to maximum likelihood estimation, and (2) no provision exists for automated model building, such as that typically implemented by SELECTION= options. This paper illustrates how both might be implemented, based in the first case on a class of generalized additive models previously published by the authors, and improvising on the basis of sweep operations in the second. Agresti-level data sets are used to illustrate the methodology. Even in these simple cases, we demonstrate that either model fit or power may improve by choosing alternatives to baseline logits. Considerably simplified models often result by zeroing parameters in addition to those suggested by the ANOVA table. For instance, many individual odds ratios might be set to one whereas otherwise they would be individually estimated and interpreted. Statisticians and other scientists faced with modeling multinomial response should find this paper of interest.

Paramjit GILL (Okanagan University College)
Spatio-Temporal Modelling of Teenage Birth Data

The teenage pregnancy in the USA is among the highest in the western nations. Four out of ten American teens - nearly a million every year - become pregnant at least once before they turn 20. Teen pregnancy has serious social, health and financial consequences. As a social issue, the focus has been on reducing current teen pregnancy rates. Indeed, the rate has been showing a downward trend since the early 1990's. Although the recent downward trend cuts across geographic, racial and ethnic lines, the drop has not been uniform in these subgroups. As a case study, we use teenage births counts data from the state of Texas. The main aim of our analysis is to measure the magnitude of variation and to assess the role of contributing factors at the geographical level. Key analytical goals are to provide county-specific estimates of the birth rates in the 1990's and to derive estimates of covariate effects, such as race, poverty level, availability of family planning services, degree of urbanity etc. Birth counts data for the two teen age groups (15-17 years old and 18-19 years old) from the 254 counties over the years 1989-1999 are used. Over-dispersion and spatial variation are incorporated by fitting a fully Bayesian formulation of the generalised linear mixed models (glmm) using the WinBUGS software.

Hyang Mi KIM (Carleton University)
Preliminary test prediction of population total under simple regression model with measurement errors

In this paper we investigate the predictions of the population total under regression superpopulation models when the explanatory variable is measured with errors. In studying of predictors of the population total which are function of the estimators of the slope of regression models with a finite sample, we consider both the ordinary least squares estimator (OLS) as well as the preliminary test estimator (PTE) of the slope and biased estimator (B) as well as its corresponding PTE of the slope under conditional model. As a result, we find that measurement errors tend to increase the variability of the predictors compared to the predictors without measurement errors. And the OLS predictor and the PTE predictor based on OLS are better than the B predictor and the PTE predictor based on B with measurement errors.

Salim LARDJANE (CREST-ENSAI)
Marginal density estimation for approximable stochastic processes

We first define approximable stochastic processes and then deal with nonparametric marginal density estimation for stationary approximable processes and for stationary processes with regular autocovariances. Conditions and rates of convergence in quadratic mean are obtained. We apply our results to non-mixing linear processes, non-mixing nonlinear autoregressive processes and Weak Bernoulli dynamical processes.

Chu-In Charles LEE (Memorial University of Newfoundland)

A One-Sided Multiple Comparison Procedure in Restricted Normal Models

There is a considerable interest in statistical inference for the one-way analysis of variance model with k quantitative levels. A new simple one-sided multiple comparison procedure is introduced to construct one-sided simultaneous confidence lower bounds for pairwise comparisons. It is a useful technique to categorize treatment levels. The new procedure is more efficient than Hayter's OSRT except when one compares the means of two adjacent treatment levels. Its efficiency in some occasions may exceed that of the least significant difference procedure.

Jack Jaxk Wai LEE (Department of Statistical & Actuarial Science)

Wavelets application (Wave shrinkage) on Time series models

Examining Wavelets smoothing methods applied on Time models such as AR, MA, ARMA, ARIMA, ARCH, etc. The selection of Threshold functions, rules and scaling rules applying on these models will be discussed. The Spectral density estimation by wavelets associated with above models will also be discussed.

Juan Pablo LEWINGER (Department of Statistics, University of Toronto)

Confidence Intervals for Penalized Likelihood Estimates of Logistic Regression Parameters

Logistic regression is one of the most widely used statistical tools in the analysis of binary data. However, with small sample sizes or sparse data, the commonly used maximum likelihood estimates can be very biased away from zero, and in some cases they can be infinite. Traditional confidence intervals based on a quadratic approximation to the log-likelihood break down in the latter situation, but open-ended confidence intervals can still be constructed using profile likelihoods. To address the estimation problem with sparse data, Firth (1993) introduced a penalized likelihood method that removes first order bias and always yields finite estimates. In this paper we consider two approaches for the construction of confidence intervals for penalized likelihood estimates of logistic regression parameters, one based on the Fisher information and the other based on the profile likelihood, and evaluate their properties via simulations. We also show how the penalized likelihood method can be applied to conditional logistic regression. We illustrate some practical advantages of the penalized likelihood approach with a sparse data set in a randomized prevention trial, and a study of genetic association of HLA markers.

Chang H LIE (Seoul National University)

Joint Reliability Importance of Partial Redundant System with Independent Components

Joint reliability importance of partial redundant systems comprised of independent components is analyzed. This paper consists of two parts: One is to deal with the joint reliability importance of a partial redundant system (k -out-of- n system) and the other is to investigate how to use the joint reliability importance in comparing various k -out-of- n systems. In specific, we show that the joint reliability importance of k -out-of- n system with identically-independently-distributed components has a simple closed form and the sign of the joint reliability importance of any two components in k -out-of- n system with non-identical components is non-positive (non-negative) if the reliabilities of all components are equal to or greater (less) than $(k-1)/(n-1)$. Some comparisons of the joint reliability importance are depicted and expressed for interpreting the meaning of the joint reliability importance with respect to the level of redundancy.

Erling H. LUNDEVALLER (svenska statistiker samfundet)

TEST OF RANDOM EFFECTS IN HETEROSKEDASTIC LINEAR MODELS

A new test of random subject effects in linear regression models for repeated measurements data is presented in this paper. The proposed test is based on a test statistic which is robust against heteroskedasticity. The asymptotic distribution of the test statistic is derived using a sequence of local alternatives. Results from a simulation study and an empirical study illustrates the finite sample performance of the test. Results presented shows that the test works well in comparison with earlier test proposed. It is also indicated that the test works well when adapted for test of random effects in generalized linear models.

Tara MADDALA (University of Texas School of Public Health)

A Parametric Approach for Analyzing Recurrent Event Survival Data in the Presence of Competing Risks

In clinical trials and observational studies, we frequently encounter situations in which individuals experience repeated failures. The analysis is complicated in these types of studies, due to dependence among repeated failures for each subject, censoring, and competing risks. A number of survival data techniques have been developed to address multivariate failure data, some of which have incorporated dependent censoring. This paper proposes a model in which subjects may experience recurrent events, which are repetitions of essentially the same event. However, the last event observed for a specific subject may be a nonfatal recurrent event, or a fatal recurrent event which terminates the event process. The method developed in this paper is based on a time-dependent Yule process with the proportional intensity assumption. Under this model, we assume that the ratio of the hazards is independent of time. We compare the proposed model to popular methods developed by Anderson and Gill, Lawless, Prentice, Williams and Peterson, and Wei, Lin and Weissfeld using both simulation studies and real data.

Amalia S MEIER (University of Washington)

Uncertainty in outcomes in discrete time proportional hazards modelling

Estimation of failure time has typically been done when failure is observed with certainty. However, in a context in which false positive and false negative test results occur, a time to event analysis is also of interest. A procedure is desired that addresses information loss due to imperfect sensitivity/specificity in the outcome. Richardson and Hughes (2000) provide an algorithm for estimating the product limit of discrete time data with uncertain outcomes. Following their work, a method is developed that permits the estimation of both the survival curve and covariate effects. Of particular interest are differences in survival between treatment groups, under the assumption of proportional hazards. It is shown that the greatest bias in both hazard and covariate effect estimation stems from failure to account for false positive tests. However, simulations show that correct specification of these false testing rates permit (nearly) unbiased estimation of both the baseline hazard and covariate effects. A graphical comparison to Cox models is made, evaluating the robustness of the estimates with regard to misspecification of the false testing rates. The method is most applicable to research in the developing world, where the accuracy of tests can be limited by cost and technological difficulty.

Etsuo MIYAOKA (Science University of Tokyo)

Statistical Models for Binary Crossover Data

We discuss a statistical modeling for binary crossover data with likelihood based approaches. Log-linear models and generalized liner transition models for 2 by 2 and 3 by 3 crossover data are examined in particular. We present some results from simulation study.

Theophile NIYONSENGA (Sherbrooke University Hospital Center)

PRIMARY CARE PHYSICIAN'S APPROACH IN TREATMENT OF DYSLIPIDEMIAS AMONG WOMEN: APPLICATION OF HIERARCHICAL MODELS.

The purpose of this paper is to evaluate whether general practitioners are following the current consensus recommendations with respect to the treatment of dyslipidemias among women in the prevention of cardiovascular disease (CVS). In order to evaluate the pattern of daily practice of primary care providers, the FAMUS (Family Medicine University of Sherbrooke) research team has set a primary care register. Data came from a computerized network of 233 family physicians, in the province of Quebec, who collected patient's relevant information on their CVS disease risk factors. Using Hierarchical Bayesian Linear Models (HBLM), we will answer to the following questions: (1) Which patient and physician characteristics do motivate the physician to request a complete lipid profile? (2) Which patient and physician characteristics do determine the physician decision to prescribe lipid-lowering therapies? (3) Can the treatment of dyslipidemias among these women be considered optimal as compared to the general recommendations? Preliminary analyses show that the treatment of dyslipidemias is sub-optimal compared to guidelines. They show as well that, on the patient level, family history of CVS events, the number of CVS risk factors, the type of risk factors (and/or the different combinations of them) and age are significant predictors of the decision to check lipid profiles and initiate therapy. On the physician level, physicians' attitudes towards lipid tests as clinical markers of CVS risk factors and treatment efficacy to lower them and their clinical experience are important factors affecting the physician decision to order a patient lipid profile.

Sudhir R PAUL (University of Windsor)

Generalized Linear Model, Zero-inflation and Over-dispersion

A discrete generalized linear model (Poisson or binomial) may fail to fit a set of data having a lot of zeros purely because of presence of zero-inflation in the data or because there is zero-inflation as well as over-dispersion in the data. In this paper we derive procedures based on score tests for analysing such data. We first consider a zero-inflated over-dispersed generalized linear model. The over-dispersed generalized linear model used here is of the form considered by Cox (1983) and Dean (1992). We then obtain score test (i) for zero-inflation in presence of over-dispersion, (ii) score test for over-dispersion in presence of zero-inflation, and (iii) omnibus score test simultaneously for zero-inflation and over-dispersion. For Poisson and binomial data these score tests are obtained as special cases. Further, score tests are obtained using zero-inflated negative binomial model for Poisson data and zero-inflated beta-binomial model for binomial data. These score tests are then compared with those obtained from the zero-inflated over-dispersed generalized linear model considered earlier. We show that for Poisson data (i) the score test statistic for testing over-dispersion in the presence of zero-inflation using the zero-inflated over-dispersed generalized linear model is identical to the score test statistic using the zero-inflated negative binomial model and (ii) the score test statistic for testing simultaneously zero-inflation and over-dispersion using the zero-inflated over-dispersed generalized linear model is identical to the score test statistic using the zero-inflated negative binomial model. Illustrative examples are given for choosing an appropriate model using the score tests developed here. Extensive simulations are performed for Poisson data to study level properties of the tests and conclusions are drawn. Simulations show that the omnibus test performs well.

Ruxandra L PINTO (University of Toronto)

Improving Markov Chain Monte Carlo Estimators by Coupling to an Approximating Chain

We show how large improvements in the accuracy of MCMC estimates for posterior expectations can sometimes be obtained by coupling a Markov chain that samples from the posterior distribution with a chain that samples from a Gaussian approximation to the posterior. Use of this method requires a coupling scheme that produces high correlation between the two chains. An efficient estimator can then be constructed that exploits this correlation, provided an accurate value for the expectation under the Gaussian approximation can be found, which for simple functions can be done analytically. Good coupling schemes are available for many Markov chain samplers, including Gibbs sampling with standard conditional distributions. For many moderate-dimensional problems, the improvement in accuracy using this method will be much greater than the overhead from simulating a second chain.

Annie QU (Oregon State University)

Testing for nonignorability of dropouts for longitudinal studies

Assessing whether or not missing data are non-ignorable or informative is often crucial to the choice of modelling strategy. Current existing diagnostic tools are either based on empirical explanatory data analysis plots or logistic models for the dropout process (Diggle & Kenward, 1994) which depend on untested distribution assumptions. The goodness-of-fit test of the quadratic inference function (Qu, Lindsay & Li, 2000) is applicable for testing informativity of missing data. The quadratic inference function approach is similar to the generalized score test except that the dimension of score in the quadratic inference function is longer than the dimension of parameters. The test statistics have an asymptotic chi-squared distribution under the null hypothesis. Simulations and schizophrenia trial data will be illustrated to evaluate its performance.

Timothy RAMSAY (Statistics Canada)

Small Domain Estimation and Non-parametric Regression in the Canadian Census

A wide variety of model-based small domain estimators have been proposed in the literature, and many of these have proven to be useful in practice. Most of these estimators model the quantity being estimated, or some transformation thereof, as a linear function of one or more auxiliary variables. In order to accurately estimate the Canadian population, it is necessary to estimate the rate of undercoverage of the Canadian census by sex and single year of age. Although the undercoverage rate is believed to depend on age, no standard transformation of this rate can reasonably be modelled as a linear function of age. This paper presents a small domain estimator based on a robust cubic smoothing spline. The smoothed rate estimates are calibrated to estimates of missed persons for aggregated age groups and a bootstrap variance estimator is derived for the resulting calibrated estimates.

Mark D ROBINSON (UBC)

Bayesian Neural Networks for Regression -- Some Thoughts on Priors

In recent years, Neural Networks (NN) have graced the literature a considerable amount in the statistical and computer science fields. NNs can be used as universal approximators, that is, a tool for regressing a dependent variable on a possibly complicated function of the explanatory variables. The NN parameters, unfortunately, are notoriously hard to interpret. Here, we take a Bayesian view. We propose and discuss prior distributions for some of the network parameters which encourage parsimony and reduce overfit, by promoting orthogonality, linearity, or additivity. Thus we consider more senses of parsimony than are discussed in the existing literature. We investigate the predictive performance of networks fit under these various priors.

EHSANES A.K.Md. SALEH (Carleton University)

Estimation of the Parameters of Multiple Regression Model with Measurement Errors

The object of the paper is to provide preliminary test and shrinkage estimators of the parameters of the multiple regression model with measurement errors and study their properties under quadratic loss and under the condition (i) when the reliability coefficient matrix is known, (ii) when the covariance matrix of the measurement errors is known. Some asymptotic analysis is provided under local alternatives.

MERIEM F SAÏD (Université Laval)

Testing for dependent censoring in a parametric survival model

For right censored survival data, testing the independence between the failure time T and the censoring time U is impossible without additional hypotheses about the distribution of (T, U) . Specifying the marginal distributions, it is possible to construct identifiable models for (T, U) , where one of the parameters "measure" dependence (cf. Emoto & Matthews, 1990 or Bhattacharyya, 1997). We want to test the independence, without parametric hypothesis about the marginal distribution of U , assuming only a distribution for T and an archimedean copula for the dependence between T and U . To do this, we construct a pseudo-likelihood and deduce a score statistics for the dependence parameter of the specified copula. We show how the dependence parameter can be estimated using the pseudo-likelihood. We also present a simulation study.

Bruce SMITH (Dalhousie University)

The tides, they are a changing

There is plenty of evidence that sea level has been gradually rising over the past century. The Bay of Fundy is in near resonance with the semi-diurnal tidal component, and physical theory suggests that as the mean level of water in the bay increases, the amplitude of this tidal component should also increase. We have examined the historical sea level records at several locations in and around the bay, and have found substantial evidence to support the theory.

Rolf TURNER (University of New Brunswick)

Misaligned Alignments

Molecular phylogeny attempts to reconstruct a tree showing the evolutionary relationships among several extant species from which molecular sequence data are available. The idea is that the degree of DNA homology between a pair of species at a fixed genetic locus should be a measure of the evolutionary distance between the organisms. There are several ways that an ancestral gene can be modified through evolutionary time. Point mutation, where one of the bases A, T, C, G is replaced by another, is the most common agent of change. In addition bases can be inserted into, or deleted from, the DNA sequence. Therefore, when starting out to construct an evolutionary tree, one typically has available DNA sequences of different lengths containing deleted or inserted portions, relative to the ancestral gene. Prior to any tree reconstruction the sequences have to be aligned. This process gives the sequences a common starting point and marks the putative positions of deleted bases. The most commonly used alignment methods work in a pairwise fashion, using some measure of distance between sequences, adding penalties for gaps, and incorporating a variety of other ad hoc procedures. The pairwise nature of the approach may not result in good overall alignments when several sequences are available, and there has been interest in developing likelihood or Bayesian methods for simultaneous alignment of multiple sequences. A hidden Markov model proposed by Krough et al is now in fairly common use. The novel statistical aspect of the algorithm is that it accommodates data sequences of different lengths, with missing values at unknown locations. We will describe our attempts to understand and implement the algorithm.

Maria Ines URRUTIA

Sociocultural Mapping of Groups of Immigrants Settled in the Province of Buenos Aires (Other than Spanish and Italians)

Argentina has a population characterized by its diversity of origins which confers it a truly multicultural profile. Buenos Aires, setting in which our research was developed, is the Province which has historically received European, Asian, African, and American migrations. It shows not only a particular ethnical physiognomy but also a cultural, economical, social and political one. Traditionally, research on migrations has been referred to the two largest majorities: Spanish and Italians and a few anthropological works on small and medium-sized immigrant groups have been carried out. Our purpose is to study Polish, Cape Verdean, Greek, Lithuanian, Ukrainian, Dutch, Danish, Japanese, etc communities, in order to characterize the modifications that are produced as a consequence of cultural changes, to make a first characterization of the sociocultural profile of these communities and to geographically locate the immigrant groups and their offsprings in the province of Buenos Aires. Methodological emphasis was based on a quantitative analysis of the information through documentary and field research. Fifty cities were selected to carry out the study. In a first stage, open interviews to qualified informants who had been in contact with other members of the community and who knew the group problems were conducted. The analyzed facts, were obtained through a selection of items from a emistructured and precodified survey, which was designed in order to find out a plurality of dimensions: personal data (age, marital status, profession etc) and sociocultural ones (mother tongue, second tongue, religion, music, food, dance, etc).

Zilin WANG (University of Western Ontario)

On Estimation of A Partial Linear Semiparametric Regression Model in Complex Survey

The aim of this work is to develop an estimation method for a semiparametric regression model for large-scale survey. In this semiparametric regression model, the explanatory variables are represented separately in two parts: nonparametric part and the parametric linear part. We estimate both the function of the nonparametric part of the model and the parameters included in the parametric part of the model. The estimation method for this model combines the local polynomial regression in complex survey by Bellhouse and Stafford (1999) and the classical least squares estimate. Moments of the related estimates have been derived. An empirical illustration is carried out using the 1990 Ontario Health survey.

Larry WELDON (Simon Fraser University)

Directions in the Evolution of Statistics

Despite forty years of revolution in the tools available for statistical analysis, the current academic tradition in statistics is remarkably similar to the pre-computer tradition. This tradition is rooted in parametric modeling, least squares, and linear models. This poster display argues for a shift in emphasis away from this tradition, from parametric modeling and data reduction to graphical display, from omnibus optimal techniques to those that are more context-specific, and from goals of objectivity to goals of revelation. Some strategies that tend to be viewed as central to our discipline are portrayed as constraints to modernization of the discipline.

Mahmoud ZAREPOUR (University of Ottawa)

A recursive method for functionals of Poisson processes

Functionals of Poisson processes arise in many statistical problems. In problems involving heavy tailed distributions they appear naturally in the study of limiting processes, while in Bayesian nonparametric statistics they are used as constructive representation of nonparametric priors. In this paper we describe a simple recursive method requiring only the use of conditional probability that is useful for characterizing Poisson process functionals. Some applications of this technique to vonvex hulls, extremes, stable measures, infinitely divisible random variables and Bayesian nonparametric priors are discussed.

Xiaoling ZHANG (University of Western Ontario)

Graphic display of orientation data

Data describing the orientation of a rigid body in three dimensional space is known as orientation data. Orientation data is commonly represented in four ways: 3x3 rotation matrices, unit quaternions, 3x3 skew-symmetric matrices, and Euler angles. We have been involved in research fitting regression models to such data. In this presentation, we explore graphic displays developed to help in model fitting. We need displays that both distinguish different orientations, and point us in the appropriate directions for improving our models.

Séance par affichage

Luc ADJENGUE (École Polytechnique de Montréal)

Un Procédé D'Estimation Pour Les Paramètres D'Un Processus De Décision De Markov Partiellement Observable

Une classe des procédés de décision de Markov partiellement observables (POMDP) est considérée. De tels processus sont définis avec deux procédés stochastiques: le procédé avec noyau et le procédé d'observations. Le processus de noyau est un processus de Markov avec un ensemble fini d'états qui ne sont pas directement observable. Un processus de noyau peut, par exemple, représenter un procédé de production (une machine ou des séries de machines) qui parfois, ne réussit pas à produire des bons items. L'état du procédé de production demeure inconnu jusqu'à ce que le cycle de production soit arrêté et une inspection complète soit effectuée. Le procédé d'observation représente les produits de sortie observables et est sensé suivre une distribution de probabilité donnée. L'objectif est d'estimer les paramètres de transitions de la matrice de probabilité du processus de noyau et les paramètres de la distribution de probabilité, sachant qu'on a un échantillon du procédé d'observation. Le procédé d'estimation est basé sur l'EM (« Espérance-maximisation ») (Dempster et autres 1977). Des divers distributions de probabilité sont considérées y compris ceux du type continu. Une étude de simulation permet d'analyser les estimateurs obtenues pour différentes grandeurs de l'échantillon. Des exemples numériques du procédé sont également présentés.

Ejaz S. AHMED (University of Regina)

Procédures D'Inférence Pour Des Processus D'Indice De Capacité Dans Des Populations Arbitraires

L'estimation simultanée des processus d'indice de capacité est considérée lorsque des échantillons aléatoires sont tirés des populations arbitraires. De plus, des procédures statistiques sont comparées pour tester l'homogénéité des processus d'indice de capacité. Une des procédures est basée sur l'estimation de la variance quand on a un grand échantillon. L'autre est basée sur le fait qu'on connaît une valeur de l'index. On propose également une transformation logarithmique dans le but de faire des tests. Des propriétés pour un grand échantillon sont obtenues par les méthodes proposées. Une étude de simulation est effectuée pour évaluer la performance des méthodes suggérées pour un échantillon modéré.

Paul D. BELL et Timothy E. O'BRIEN (Loyola University Chicago)

PREDICTION DE L'ASSISTANCE AUX JEUX DE MAJOR LEAGUE BASEBALL

Cet affiche fournit les modèles de régression predictive pour l'assistance de jeu pour quatre équipes de Baseball de Ligue Majeures. Variables indépendants inclut jour de la semaine, date, l'adversaire, l'heure de jeu, et succès des équipes respectifs. Résultats remarquables inclure l'a mélangé la signification de temps dans prédit l'assistance, l'haute augmentation d'assistance quand la certaine visite d'équipes, et un modèle de parabolic pour l'assistance contre la date. Bien que les modèles de predictive soient basés sur les figures de la saison de l'année 2000, ils ont été essayés pour la saison 2001 avec le succès notable. Ramifications de ceci étudie inclut la prediction précis de vende des billets, et des besoins de personnel de stade, local, affaires, et restaurants.

David R BELLHOUSE (University of Western Ontario)

Le Révérend Thomas Bayes FRS: Un Essai Biographique

L'année 2001 marque le 300^e anniversaire de la naissance de Thomas Bayes, l'auteur du théorème de Bayes. Cette biographie est écrite pour la célébration de cet anniversaire. Dans la biographie scientifique, Karl Pearson a dit : « il est impossible de comprendre le travail d'un homme à moins que vous connaissez un peu sa personnalité et à moins que vous connaissiez un peu son environnement. Et son environnement signifie, l'état social et la politique de son propre âge. » Dans cet esprit, les événements principaux de la vie de Bayes et son travail sont étudiés, de même que l'environnement scientifique et théologique dans lequel il a travaillé. La biographie actuelle contient du matériel bien connu ainsi que des nouvelles informations qui ne se trouvent pas dans les biographies précédentes sur Bayes.

Willard J. BRAUN (University of Western Ontario)

La Régression Spline Avec Noyau

On propose une méthode qui combine la régression spline et le lissage par noyau. Une étude de simulation prouve que les estimateurs de courbe résultants peuvent réussir à capturer les caractéristiques principales de la vraie courbe plus efficacement que la régression polynomiale locale lorsque la courbure change rapidement. Les noyaux splines montrent également moins de sensibilité pour le choix de la largeur de la bande que la régression polynomiale locale. Dans beaucoup de situations, les nœuds peuvent être choisis manuellement. C'est un avantage par rapport aux approches avec une largeur de fenêtre variable et dans la régression polynomiale par noyau d'ordre variable, où le choix de la largeur ou de l'ordre doit se faire manuellement. Quand une sélection automatique de nœud est exigée, la méthode de suppression de nœud semble bien fonctionner. Une directe approche dans la sélection de la largeur de bande est aussi disponible. Quelques exemples d'illustrations seront également présentés.

David R BRILLINGER (University of California, Berkeley)

Le Processus D'Analyse Des Taux De Débits Extrêmes Le Long Du Fleuve De Mississippi

Cet article considère le cas des processus de points stationnaires et de leur utilisation dans une installation qui est raisonnablement bien comprise physiquement. La situation est celle du passage des débits d'eau extrême qui traverse une série de barrages le long du fleuve de Mississippi. Le but est de faire une enquête sur la validité de l'analyse de cohérence partielle, un analogue à l'analyse de corrélation partielle. Une analyse du maximum de vraisemblance est également présentée.

Noel G CADIGAN (Fisheries and Oceans)

L'Influence Locale Généralisée Avec Application À L'Analyse De La Population Séquentielle

Le but de cette présentation est de décrire une méthodologie générale pour mesurer l'influence locale qui est particulièrement utile pour des modèles complexes. Un exemple de ce type de modèle est l'analyse de population séquentielle (SPA). Le SPA est un modèle important employé pour fournir beaucoup d'informations pour plusieurs grandes usines de pêches commerciales dans le monde entier. Avec le SPA, il n'est pas toujours facile de comprendre comment les données et les hypothèses de modélisation influencent les résultats ou pourquoi des différentes hypothèses produisent des résultats différents. Cependant, les scientifiques qui évaluent la quantité de poisson exigent cette information pour évaluer la fiabilité du modèle employée pour estimer la taille de la population. L'approche de base que nous utilisons pour mesurer l'influence, est d'évaluer comment les petites perturbations des entrées du SPA impliquent des effets importants sur les sorties du SPA. Notre contribution principale, est de montrer comment cette information peut être obtenue dans des applications courantes. On présente un exemple qui illustre l'utilité pratique du diagnostic.

Hugh CHIPMAN (University of Waterloo)

Modèles Additifs Pour Des Données De Sondage Par L'Intermédiaire Des Moindres Carrés Pénalisés

L'utilisation des moindres carrés pénalisés dans le contexte d'échantillonnage est étudiée comme un outil pour l'analyse d'une tendance générale pour une population finie. Nous nous concentrons sur la régression lisse avec un modèle d'erreur normal. Ce modèle est compliqué dans le contexte d'échantillonnage à cause du design du sondage, qui pourrait inclure la stratification, l'échantillonnage par grappe, et d'autres designs de sondages complexes. Les égalités pour les covariables arrivent fréquemment dans des enquêtes à grande échelle ayant comme résultat l'application des lissages basés sur le nuage de points des moyennes. L'estimation du lissage (par exemple les splines de lissage) dépend du design de l'échantillon seulement par l'intermédiaire de la pondération de l'échantillon, c.-à-d. que le logiciel standard peut être utilisé pour l'estimation. Faire de l'inférence pour ces courbes est un plus grand défi, en raison des corrélations induites par le design de l'échantillon. Nous proposons et illustrons les tests qui expliquent le design de l'échantillon. Des exemples d'illustration sont donnés en utilisant l'enquête de santé de l'Ontario, y compris le lissage du diagramme de dispersion, le diagnostic de modèle, le modèle additif, et divers tests significatifs.

Laura L COWEN (Simon Fraser University)

L'Ajustement Des Estimateurs De Survie En Utilisant Des Courbes De Panne De Marqueur Radio

Estimer des taux de survie et tester des hypothèses sur le processus de survie sont essentiels pour comprendre la dynamique des populations animales. Celles-ci sont souvent étudiées en utilisant des méthodes de capture-recapture, c'est-à-dire les animaux sont identifiés par des étiquettes différentes et identifiables, ensuite les animaux sont libérés. Les données de recaptures sont enregistrées et on obtient des estimations sur la survie et sur le taux de capture. Les marqueurs radio ont été utilisées précédemment dans des études de capture-recapture, mais ils ont été utilisés seulement récemment pour l'étude de la survie du saumon. Cependant, contrairement, aux travaux fait précédemment sur la méthode de capture-recapture radio-téléométrique, des problèmes se produisent quand les poissons ne peuvent pas être détectés à cause de la panne des étiquettes. La durée de vie des étiquettes peut avoir un effet important sur l'estimation de survie, si la panne de l'étiquette se produit au milieu de l'étude. Dans certains cas, l'information est disponible sur la vie des marqueurs radio d'un sous-ensemble d'animaux qui sont gardés pour étudier la durée de vie des étiquettes. Étant donné la courbe d'échec pour les étiquettes, des ajustements peuvent être faits aux détections connues pour tenir compte de la proportion des étiquettes qui ne pourraient pas être détectées parce qu'une partie de ces étiquettes ne sont plus en activité. La distribution du temps de panne des étiquette peut être adaptée d'une manière non paramétrique en utilisant les estimateurs de Kaplan-Meier, ou paramétriquement, en utilisant par exemple, une distribution de Weibull. Les données historiques de capture peuvent être ajustées en utilisant la distribution ajustée du temps d'échec pour tenir compte de la non-détection du poisson causé par l'échec de l'étiquette. La méthodologie à développer sera générale, permettant pour des différents types de marqueurs radio avec différentes distributions de temps d'échec et plusieurs cohortes de poissons étiquetés. Cette méthodologie sera développée suivant l'approche de Lebreton et al, 1992.

Monica L. DRAGOMIR (Department of Statistics, University of Toronto)

La Régression En Ondelettes Pour Des Médianes Conditionnelles

Des méthodes classiques pour les ondelettes sont bonnes pour estimer la fonction d'espérance conditionnelle d'une réponse. Cependant, dans plusieurs problèmes, la réponse avec médiane conditionnelle peut être une fonction plus utile à estimer. Les autres méthodes classiques pour estimer les médianes conditionnelles (ou des quartiles conditionnels) sont plus direct mais les calculs sont plus difficiles. Un algorithme est proposé pour utiliser les ondelettes pour l'estimation de la médiane conditionnelle, et certaines de ses propriétés sont étudiées.

James E DUNN (University of Arkansas)

Variations Sur Les Modèles Logit Multinomiaux : Liens Alternatifs Et La Sélection Pas À Pas

Users of standard statistical software, e.g., SAS procedure CATMOD, will recognize these deficiencies: (1) for multinomial, unordered categories of response, only use of baseline logit link functions admits to maximum likelihood estimation, and (2) no provision exists for automated model building, such as that typically implemented by SELECTION= options. This paper illustrates how both might be implemented, based in the first case on a class of generalized additive models previously published by the authors, and improvising on the basis of sweep operations in the second. Agresti-level data sets are used to illustrate the methodology. Even in these simple cases, we demonstrate that either model fit or power may improve by choosing alternatives to baseline logits. Considerably simplified models often result by zeroing parameters in addition to those suggested by the ANOVA table. For instance, many individual odds ratios might be set to one whereas otherwise they would be individually estimated and interpreted. Statisticians and other scientists faced with modeling multinomial response should find this paper of interest.

Paramjit GILL (Okanagan University College, Kelowna, BC)

La Modélisation Spatio-Temporelle Pour Les Données De La Grossesse Des Adolescentes

La grossesse des adolescentes aux États-Unis est une des plus haute parmi les nations occidentales. Quatre adolescentes américaines sur dix, soit presque un million chaque année, deviennent enceintes au moins une fois avant qu'elles aient 20 ans. La grossesse à l'adolescence a des conséquences sérieuses, au niveau social, financier et au niveau de la santé. Du point de vue social, le but est de réduire le taux actuel de la grossesse à l'adolescence. En effet, depuis le début des années 90, le taux diminue. Bien que l'évolution récente soit à la baisse par région géographique, groupes raciaux et ethnique, celle-ci n'est pas uniforme. Comme étude de cas, nous utilisons le nombres d'adolescentes qui ont donné naissances dans l'état du Texas. Le but principal de notre analyse est de mesurer l'étendue de la variation et d'évaluer le rôle des facteurs qui contribuent au niveau géographique. Les buts analytiques principaux sont de fournir des estimateurs par comté des taux de naissance dans les années 90 et de dériver les estimateurs des effets des covariables, tels que la race, le niveau de pauvreté, la disponibilité des programmes de planification familiaux et le degré d'urbanisme etc. Le compte, de données de naissances pour les deux catégories d'âge de l'adolescence (15-17 ans et 18-19 ans) des 254 comtés au cours des années 1989-1999, est utilisé. La dispersion et la variation spatiale sont incorporées en adaptant une formulation entièrement bayésienne pour les modèles mixtes linéaires généralisés (glmm), en utilisant le logiciel WinBUGS.

Hyang Mi KIM (Carleton University)

Preliminary test prediction of population total under simple regression model with measurement errors

In this paper we investigate the predictions of the population total under regression superpopulation models when the explanatory variable is measured with errors. In studying of predictors of the population total which are function of the estimators of the slope of regression models with a finite sample, we consider both the ordinary least squares estimator (OLS) as well as the preliminary test estimator (PTE) of the slope and biased estimator (B) as well as its corresponding PTE of the slope under conditional model. As a result, we find that measurement errors tend to increase the variability of the predictors compared to the predictors without measurement errors. And the OLS predictor and the PTE predictor based on OLS are better than the B predictor and the PTE predictor based on B with measurement errors.

Milorad KOVACEVIC e Lenka MACH (Statistique Canada)

Inférence sur des paramètres de domaine quand la taille du domaine dépend de paramètres de nuisance.

Un domaine d'étude est une sous-population pour laquelle l'inférence statistique est demandée. Parfois la taille du domaine dépend des paramètres inconnus qui ne sont pas intrinsèquement d'intérêt pour l'inférence mais qui doivent être estimés à partir du même échantillon. Les exemples typiques incluent le domaine des ménages à bas revenu, le domaine des enfants de poids insuffisant, et une population après avoir éliminé les valeurs aberrantes à l'aide de la « règle des trois sigmas ». La définition de ces domaines dépend des paramètres de nuisance de la population : le revenu moyen, le poids moyen pour un enfant dans une catégorie d'âge donnée, la moyenne et l'écart type de la population, respectivement. Les paramètres d'intérêt du domaine, tels que les moyennes, les totaux, et les taux, sont alors des fonctions non lisses dépendantes des paramètres de nuisance. Ils sont habituellement estimés à l'aide des estimateurs à substitution après avoir fait la substitution des paramètres de nuisance par leurs estimateurs. Sous certaines conditions de régularité pour des échantillonnages complexes, nous prouvons la convergence de ces estimateurs et dérivons leurs erreurs types en utilisant l'approche des équations d'estimation. Nous supportons nos résultats théoriques par des illustrations basées sur l'enquête de Statistique Canada sur la dynamique du revenu et du travail.

Salim LARDJANE (CREST-ENSAI)

L'Estimation De Densité Marginale Pour L'Approximation De Processus Stochastique

Nous définissons d'abord, l'approximation des processus stochastiques et nous traitons ensuite de l'estimation de la densité marginale non paramétrique pour des processus stationnaires approximable et pour des processus stationnaires avec des auto-covariances régulières. Des conditions et des taux de convergence pour des moyennes quadratiques sont obtenus. Nous appliquons nos résultats à des processus linéaires non mixte, des processus auto-régressifs non linéaires et non mixte et des processus dynamiques faibles de Bernoulli.

Chu-In Charles LEE (Department of Mathematics and Statistics, Memorial University of Newfoundland)
Une Procédure De Comparaison Multiple Unilatéral Pour Des Modèles Normals Restreints

Il y a un intérêt considérable dans l'inférence statistique pour l'analyse unilatérale du modèle de variance avec k niveaux quantitatifs. Une nouvelle procédure simple de comparaisons multiples unilatérales est présentée pour construire des limites de confiance inférieures unilatérales pour des comparaisons appariées. C'est une technique utile pour classer les niveaux de traitement. Le nouveau procédé est plus efficace que celui de Hayter OSRT à moins qu'on compare les moyennes de deux niveaux de traitement adjacents. Dans quelques occasions, son efficacité peut dépasser les procédures avec les différences les moins significatives (« LSD procedure »).

Jack Jaxk Wai LEE (Department of Statistical & Actuarial Science)
L'Application Des Ondelettes À Rétrécisseur Sur Les Modèles De Série Chronologique

Examiner les méthodes de lissage des ondelettes, appliqués sur des modèles de séries chronologiques tels que Ar, MA, ARMA, ARIMA, ARCH, etc. La sélection des fonctions de seuil, les règles et les règles d'échelle s'appliquant sur ces modèles seront discutées. L'estimation de la densité spectrale par les ondelettes associée aux modèles ci-dessus sera également discutée.

Juan Pablo LEWINGER (Department of Statistics, University of Toronto)
Les Intervalles De Confiance Pour Des Estimateurs De Vraisemblance Pénalisés Des Paramètres D'Une Régression Logistique

La régression logistique est un des outils statistiques les plus utilisées dans l'analyse des données binaires. Cependant, avec des petites tailles d'échantillon ou des données difficiles à observer, les estimateurs du maximum de vraisemblance généralement utilisés, peuvent être très biaisés loin de zéro, et dans certains cas, ils peuvent être infinis. Les intervalles de confiance traditionnels, basés sur une approximation quadratique du log-vraisemblance ne fonctionnent pas dans la dernière situation, mais des intervalles de confiance ouverts peuvent être construits en utilisant le profil des vraisemblances. Pour régler le problème d'estimation avec des données rares, Firth (1993) a présenté une méthode de vraisemblance pénalisée qui enlève le biais de premier niveau et donne toujours des estimateurs finis. Dans cet article, nous considérons deux approches pour la construction des intervalles de confiance pour des estimateurs de vraisemblance pénalisés des paramètres de régression logistique ; l'une est basée sur l'information de Fisher et l'autre est basée sur le profil de vraisemblance, et on évalue sa propriété à l'aide de simulations. Nous montrons également comment la méthode de vraisemblance pénalisée peut être appliquée à la régression logistique conditionnelle. Nous illustrons quelques avantages pratiques de l'approche de vraisemblance pénalisée avec des données rares pour un essai de prévention randomisé, et une étude d'association génétique à l'aide de marqueur "HLA".

Chang H LIE
Seoul National University, Dept. of Industrial Engineering
L'Importance De La Fiabilité Conjointe Des Systèmes Redondants Partiels Avec Des Composantes Indépendantes

L'importance de la fiabilité conjointe des systèmes redondants partiels, qui contiennent des composantes indépendantes, est analysée. Cet article se compose de deux parties : la première partie traite de l'importance de la fiabilité conjointe d'un système redondant partiel (k parmi n systèmes) et l'autre est d'étudier comment utiliser l'importance de la fiabilité conjointe en comparant divers k parmi n systèmes. Spécifiquement, nous prouvons que l'importance de la fiabilité de k parmi n systèmes avec des composantes indépendantes et identiquement distribuées (i.i.d) a une forme analytique simple ; et que le signe de l'importance de la fiabilité conjointe pour deux composantes quelconques dans k parmi n systèmes avec des composantes non identiques, est non positif (non négatif), si les fiabilités de toutes les composantes sont égales à ,ou plus grandes (petites), que $(k-1)/(n-1)$. Quelques comparaisons sur l'importance de la fiabilité conjointe sont exprimées pour l'interpréter le sens de l'importance de la fiabilité conjointe en ce qui concerne le niveau de la redondance.

Erling H. LUNDEVALLER (Svenska Statistiker Samfundet)
Test Des Effets Aléatoires Pour Des Modèles Linéaire Hétéroscédastique

Un nouveau test pour des sujets à effets aléatoires dans un modèles de régression linéaire avec des données à mesures répétées, est présenté dans cet article. Le test proposé est basé sur une statistique de test qui est robuste contre l'hétéroscédasticité. La distribution asymptotique de la statistique de test est dérivée en utilisant une séquence d'alternative locale. Des résultats d'une étude de simulation et d'une étude empirique illustrent la performance d'un test avec un échantillon fini. Les résultats trouvées montrent que le test fonctionne mieux que les tests proposés auparavant. Il est montré aussi que le test travail bien quand il est adapté pour tester les effets aléatoires dans des modèles linéaires généralisés.

Tara MADDALA (University of Texas School of Public Health)
Une Approche Paramétrique Pour Analyser Des Données D'Événement De Survie Récurrents En Présence De Risques Concurrents

Dans des essais cliniques et des études empiriques, nous rencontrons fréquemment des situations dans lesquelles les individus ont eu des expériences répétées d'échec. L'analyse est compliquée dans ces types d'études due à la dépendance entre la répétition d'échec pour chaque sujet, la censure, et les risques concurrents. Un certain nombre de techniques pour des données de survie ont été développées pour manier les données d'échec multidimensionnelles, dont certaines ont incorporé une censure dépendante. Cet article propose un modèle dans lequel les sujets peuvent éprouver des événements récurrents, qui sont essentiellement des répétitions du même événement. Cependant, le dernier événement observé pour un sujet spécifique peut être un événement récurrent non fatal, ou un événement récurrent fatal qui termine le processus d'événement. La méthode développée dans cet article est basée sur un processus de Yule dépendant du temps avec l'hypothèse d'intensité proportionnelle. Sous ce modèle, nous supposons que le rapport des risques est indépendant du temps. Nous comparons le modèle proposé aux méthodes populaires développées par Anderson et Gill, Lawless, Prentice, Williams et Peterson, et Wei, Lin et Weissfeld en utilisant des études de simulation et des vraies données.

Etsuo MIYAOKA (Science University of Tokyo)
Modèles Statistiques Pour Des Données Binaires Croisées

Nous discutons un modèle statistique pour des données binaires croisées, avec des approches basées sur la méthode de vraisemblance. Des modèles log-linéaires et des modèles généralisés de transition pour des données croisées 2 par 2 et 3 par 3 sont examinés plus particulièrement. Nous présentons quelques résultats tirés des études de simulation.

Theophile NIYONSENGA (Sherbrooke University Hospital Center)
L'Approche Du Médecin De Premier Soins Pour Le Traitement De Dyslipidémias Chez Les Femmes: Application Des Modèles Hiérarchiques

Le but de cet article est d'évaluer si les médecins généralistes suivent les recommandations actuelles de consensus en ce qui concerne le traitement des dyslipidémias chez les femmes, pour la prévention de la maladie cardio-vasculaire (CVS). Afin d'évaluer la pratique quotidienne des médecins de premier soins, l'équipe de recherche de FAMUS (Family Medicine University of Sherbrooke) a placé un registre de premier soins. Les données sont obtenues par un réseau d'ordinateurs de 233 médecins de famille au Québec qui ont rassemblé l'information appropriée des patients sur les facteurs de risque pour leurs maladie CVS. En utilisant les modèles linéaires bayésiens hiérarchiques (HBLM), nous répondrons aux questions suivantes: (1) quelles caractéristiques du patient et du médecin motivent le médecin à demander un profil complet de lipide? (2) quelles caractéristiques du patient et du médecin déterminent la décision du médecin à prescrire des thérapies pour la réduction de lipides? (3) est-ce que le traitement des dyslipidémias chez ces femmes peut être considéré optimal par rapport aux recommandations générales? Des analyses préliminaires montrent que le traitement des dyslipidémias est suboptimal comparativement aux directives. Elles montrent aussi qu'au niveau du patient, les antécédents familiaux des événements de CVS, le nombre de facteurs de risque de CVS, le type de facteurs de risque (et/ou leurs différentes combinaisons) et l'âge sont les prédicteurs significatifs pour la décision de vérifier les profils de lipide et initiée la thérapie. Au niveau du médecin, les attitudes des médecins envers le test de lipide utilisé comme repères cliniques pour les facteurs de risque de CVS et l'efficacité du traitement pour les abaisser et leurs expériences cliniques sont des facteurs importants affectant la décision du médecin d'inscrire au patient un profil de lipide.

Sudhir R PAUL (Department of Mathematics and Statistics, University of Windsor)
Generalized Linear Model, Zero-inflation and Over-dispersion

A discrete generalized linear model (Poisson or binomial) may fail to fit a set of data having a lot of zeros purely because of presence of zero-inflation in the data or because there is zero-inflation as well as over-dispersion in the data. In this paper we derive procedures based on score tests for analysing such data. We first consider a zero-inflated over-dispersed generalized linear model. The over-dispersed generalized linear model used here is of the form considered by Cox (1983) and Dean (1992). We then obtain score test (i) for zero-inflation in presence of over-dispersion, (ii) score test for over-dispersion in presence of zero-inflation, and (iii) omnibus score test simultaneously for zero-inflation and over-dispersion. For Poisson and binomial data these score tests are obtained as special cases. Further, score tests are obtained using zero-inflated negative binomial model for Poisson data and zero-inflated beta-binomial model for binomial data. These score tests are then compared with those obtained from the zero-inflated over-dispersed generalized linear model considered earlier. We show that for Poisson data (i) the score test statistic for testing over-dispersion in the presence of zero-inflation using the zero-inflated over-dispersed generalized linear model is identical to the score test statistic using the zero-inflated negative binomial model and (ii) the score test statistic for testing simultaneously zero-inflation and over-dispersion using the zero-inflated over-dispersed generalized linear model is identical to the score test statistic using the zero-inflated negative binomial model. Further, we show that for binomial data (i) the score test statistic for testing over-dispersion in the presence of zero-inflation using the zero-inflated over-dispersed generalized linear model is identical to the score test statistic using the zero-inflated beta-binomial model and (ii) the score test statistic for testing simultaneously zero-inflation and over-dispersion using the zero-inflated over-dispersed generalized linear model is identical to the score test statistic using the zero-inflated beta-binomial model. Illustrative examples are given for choosing an appropriate model using the score tests developed here. Extensive simulations are performed for Poisson data to study level properties of the tests and conclusions are drawn. Simulations show that the omnibus test performs well.

Ruxandra L. PINTO (Department of Statistics, University of Toronto)

L'Amélioration Des Estimateurs De Monte Carlo Markoviens En Les Regroupant Avec Une Chaîne Approximative

Nous montrons comment de grandes améliorations, dans l'exactitude des estimateurs MCMC pour des espérances a posteriori, peuvent parfois être obtenues en utilisant une chaîne de Markov qui prend un échantillon provenant d'une distribution a posteriori avec une chaîne qui prend un échantillon provenant d'une approximation gaussienne a posteriori. L'utilisation de cette méthode exige un regroupement qui produit une forte corrélation entre les deux chaînes. On peut alors construire un estimateur efficace qui exploite cette corrélation, si une valeur précise pour l'espérance sous l'approximation gaussienne peut être trouvée ou, qui pour des fonctions simples, peut être fait analytiquement. Les bons regroupements sont disponibles pour beaucoup d'échantillons de chaîne de Markov, y compris l'échantillonnage de Gibbs avec des distributions conditionnelles standard. Pour beaucoup de problèmes de dimensions modérés, l'amélioration dans l'exactitude utilisant cette méthode sera beaucoup plus grande que le temps nécessaire pour simuler une deuxième chaîne.

Annie QU (Oregon State University)

Testing for nonignorability of dropouts for longitudinal studies

Assessing whether or not missing data are non-ignorable or informative is often crucial to the choice of modelling strategy. Current existing diagnostic tools are either based on empirical explanatory data analysis plots or logistic models for the dropout process (Diggle & Kenward, 1994) which depend on untested distribution assumptions. The goodness-of-fit test of the quadratic inference function (Qu, Lindsay & Li, 2000) is applicable for testing informativity of missing data. The quadratic inference function approach is similar to the generalized score test except that the dimension of score in the quadratic inference function is longer than the dimension of parameters. The test statistics have an asymptotic chi-squared distribution under the null hypothesis. Simulations and schizophrenia trial data will be illustrated to evaluate its performance.

Timothy RAMSAY (Statistics Canada)

Small Domain Estimation and Non-parametric Regression in the Canadian Census

A wide variety of model-based small domain estimators have been proposed in the literature, and many of these have proven to be useful in practice. Most of these estimators model the quantity being estimated, or some transformation thereof, as a linear function of one or more auxiliary variables. In order to accurately estimate the Canadian population, it is necessary to estimate the rate of undercoverage of the Canadian census by sex and single year of age. Although the undercoverage rate is believed to depend on age, no standard transformation of this rate can reasonably be modelled as a linear function of age. This paper presents a small domain estimator based on a robust cubic smoothing spline. The smoothed rate estimates are calibrated to estimates of missed persons for aggregated age groups and a bootstrap variance estimator is derived for the resulting calibrated estimates.

Mark D. ROBINSON (M.Sc. Student, UBC)

Réseaux Neuronaux Bayésiens Pour La Régression - Quelques Pensées Sur Les Distributions A Priori

Ces dernières années, les réseaux neuronaux (RN) sont très nombreux dans la littérature surtout dans les domaines statistiques et les domaines informatiques. RN peut être utilisé comme des approximations universels, c'est-à-dire, un outil pour régresser une variable dépendante sur une fonction complexe avec des variables explicatives. Malheureusement, il est difficile d'interpréter les paramètres des RN. Ici, nous adoptons une position bayésienne. Nous proposons et discutons des distributions a priori pour une partie des paramètres de réseau qui encouragent la parcimonie et réduisent le sur-ajustement, en utilisant l'orthogonalité, la linéarité, ou l'additivité. Ainsi, nous parlons du sens de la parcimonie beaucoup plus que dans la littérature existante. Nous étudions la performance prédictive des réseaux qui sont ajustés sous ces divers distributions a priori.

EHSANES A.K.Md. SALEH (Carleton University, Ottawa)

Estimation Of The Parameters Of Multiple Regression Model With Measurement Errors

The object of the paper is to provide preliminary test and shrinkage estimators of the parameters of the multiple regression model with measurement errors and study their properties under quadratic loss and under the condition (i) when the reliability coefficient matrix is known, (ii) when the covariance matrix of the measurement errors is known. Some asymptotic analysis is provided under local alternatives.

MERIE M SAÏD (Université Laval)

Tester Des Données Censurées Dépendantes Dans Un Modèle Paramétrique De Survie

Pour des données de survie censurées à droite, tester l'indépendance entre le moment de panne T et le moment de censure U , est impossible sans des hypothèses supplémentaires sur la distribution de (T, U) . En spécifiant les distributions marginales, il est possible de construire des modèles identifiables pour (T, U) , où un des paramètres « mesurent » la dépendance (cf. Emoto & Matthews, 1990 ou Bhattacharyya, 1997). Nous voulons tester l'indépendance, sans hypothèse paramétrique au sujet de la distribution marginale de U , assumant seulement une distribution pour T et une archimédienne pour la dépendance entre T et U . Pour faire cela, nous construisons une pseudo-vraisemblance et nous déduisons des statistiques de score pour le paramètre de dépendance du copule spécifiée. Nous montrons comment le paramètre de la dépendance peut être estimé en utilisant la pseudo-vraisemblance. Nous présentons également une étude de simulation.

Bruce SMITH (Dalhousie University)

Les Marées Changent

Il y a une grande évidence que le niveau de la mer a graduellement augmenté au cours du siècle passé. La Baie de Fundy est en proche résonance avec la composante de la marée semi-diurnale, et la théorie physique suggère qu'à mesure que le niveau moyen de l'eau dans la baie augmente, l'amplitude de cette composante de marée devrait également augmenter. Nous avons examiné les enregistrements historiques du niveau de la mer dans plusieurs endroits, c'est-à-dire dans et autour de la baie, et nous avons trouvé des preuves importantes pour supporter la théorie.

Rolf TURNER (Department of Maths and Stats, University of New Brunswick)

Alignements Mal Alignés

La phylogénie moléculaire essaie de reconstruire un arbre montrant l'évolution des rapports parmi plusieurs espèces éteintes pour lesquelles une séquence de données moléculaires sont fournies. L'idée est que le degré d'homologie d'ADN entre une paire d'espèces à un lieu génétique fixe devrait être une mesure de la distance évolutive entre deux organismes. Il y a plusieurs manières qu'un gène héréditaire peut être modifié durant l'évolution. Le point de mutation, où une des bases A, T, C, G est remplacée par une autre, est l'agent le plus commun du changement. En outre, des bases peuvent être insérées dans la séquence d'ADN ou encore effacées. Par conséquent, pour construire un arbre évolutif, on a typiquement une séquence disponible d'ADN de différentes longueurs contenant des parties effacées ou insérées, relativement au gène héréditaire. Avant n'importe quelle reconstruction d'arbre, les séquences doivent être alignées. Ce processus donne aux séquences un point de départ commun et marque les positions putatives des bases effacées. Les méthodes les plus généralement utilisées d'alignement fonctionnent pour des paires, en utilisant une certaine mesure de la distance entre les séquences, en ajoutant des pénalités pour les espaces, et en incorporant une variété de procédures « ad hoc ». À cause des paires, cette approche peut causer un mauvais alignement global, quand plusieurs séquences sont disponibles, de plus il y a eu un intérêt pour développer des méthodes de vraisemblance ou des méthodes bayésiennes pour l'alignement simultané des séquences multiples. Un modèle de Markov sous-jacent, qui avait été proposé par Krough et al, est maintenant assez souvent utilisé. Ce qui est bon dans cet algorithme est qu'il est conçu pour des séquences de données avec différentes longueurs, et avec des valeurs manquantes à des emplacements inconnus. Nous décrivons nos essais pour comprendre et de mettre en application l'algorithme.

María Inés URRUTIA, Marta Maffia · María Apezteguía (Université Nationale de La Plata)

CARTOGRAPHIE SOCIO-CULTURELLE DES GROUPES D'IMMIGRANTS ÉTABLIS DANS LA PROVINCE DE BUENOS AIRES (À L'EXCEPTION DES ESPAGNOLS ET DES ITALIENS)

L'Argentine présente une population caractérisée par la diversité d'origines lui donnant un profil vraiment multiculturel. La province de Buenos Aires où nous avons développé notre recherche, montre une particulière physionomie ethnique, culturelle, économique, politique et sociale, puisque c'est la région du pays qui a historiquement reçu des vagues d'immigrants européennes, asiatiques, américaines et africaines. Les recherches sur les migrations ont traditionnellement visé les deux groupes plus nombreux: les Espagnols et les Italiens; par contre, il existe très peu de travaux spécifiquement anthropologiques concernant les groupes d'immigrants petits ou moyens. Nous nous proposons d'étudier des communautés de Polonais, de Cap-verdiens, de Grecs, de Lituanais, d'Ucrainiens, de Hollandais, de Danois, de Japonais..., dans le but, d'une part, de caractériser les modifications qui se produisent comme conséquence du changement culturel, d'autre part, de réaliser une première caractérisation du profil socio-culturel de ces communautés, et enfin de localiser géographiquement les groupes de ces immigrants et de leurs descendants dans la province de Buenos Aires. L'approche méthodologique est partie d'une analyse quantitative de l'information, à travers la recherche documentaire et d'une recherche sur terrain. Nous avons retenu 50 villes pour y faire l'étude. Dans une première étape nous avons mené des entretiens ouverts à des rapporteurs qualifiés en contact avec d'autres membres de la communauté et en connaissance de la problématique du groupe. Les données analysées ont été retenues à partir de la sélection des items d'une enquête semi-structurée et précodée, conçue pour rechercher une pluralité de dimensions, à savoir: renseignements personnels (âge, état civil, occupation, etc.) et socio-culturels (langue maternelle, langue seconde, religion, musique, alimentation, danses...)

Zilin WANG (University of Western Ontario, Department of Statistical & Actuarial Sciences)

L'Estimation D'Un Modèle De Régression Partielle Linéaire Et Semi-Paramétrique Dans Un Sondage Complexe

Le but de ce travail est de développer une méthode d'estimation pour un modèle de régression semi-paramétrique pour un très grand sondage. Dans ce modèle de régression semi-paramétrique, les variables explicatives sont représentées séparément en deux parties : la partie non paramétrique et la partie linéaire paramétrique. Nous estimons la fonction de la partie non paramétrique du modèle et aussi les paramètres inclus dans la partie paramétrique du modèle. La méthode d'estimation pour ce modèle regroupe la régression polynomiale locale pour des sondages complexes donnée par Bellhouse et Stafford (1999) et l'estimation des moindres carrés classiques. Les moments des estimateurs ont été dérivés. Une illustration empirique est effectuée en utilisant l'enquête sur la santé faite en Ontario en 1990.

Larry WELDON (Stats Department, Simon Fraser University)

Directions in the Evolution of Statistics

Despite forty years of revolution in the tools available for statistical analysis, the current academic tradition in statistics is remarkably similar to the pre-computer tradition. This tradition is rooted in parametric modeling, least squares, and linear models. This poster display argues for a shift in emphasis away from this tradition, from parametric modeling and data reduction to graphical display, from omnibus optimal techniques to those that are more context-specific, and from goals of objectivity to goals of revelation. Some strategies that tend to be viewed as central to our discipline are portrayed as constraints to modernization of the discipline.

Mahmoud ZAREPOUR (University of Ottawa)

A recursive method for functionals of Poisson processes

Functionals of Poisson processes arise in many statistical problems. In problems involving heavy tailed distributions they appear naturally in the study of limiting processes, while in Bayesian nonparametric statistics they are used as constructive representation of nonparametric priors. In this paper we describe a simple recursive method requiring only the use of conditional probability that is useful for characterizing Poisson process functionals. Some applications of this technique to vonnev hulls, extremes, stable measures, infinitely divisible random variables and Bayesian nonparametric priors are discussed.

Xiaoling ZHANG (University of Western Ontario)

L'Affichage Graphique Des Données D'Orientation

Les données décrivant l'orientation d'un corps rigide dans l'espace tridimensionnel sont connues comme des données d'orientation. Les données d'orientation sont généralement représentées de quatre manières : matrices de la rotation 3×3 , quaternions de norme unité, matrices 3×3 antisymétriques, et les angles d'Euler. Nous avons été impliqués dans la recherche pour l'ajustement des modèles de régression à de telles données. Dans cette présentation, nous explorons les affichages graphiques développés pour aider dans l'ajustement de modèle. Nous avons besoin des affichages qui distinguent différentes orientations, et qui nous dirige dans les directions appropriées pour améliorer nos modèles.

Abstracts

Welcome and SSC Gold Medal Address

Monday, June 11th, 8:30 – 10:00, Images

The Weighted Likelihood

Jim ZIDEK (University of British Columbia)

Arguably the likelihood has been one of the greatest contributions to Statistics. Its associated estimators, tests and their large sample theories have helped to make it one of the most powerful tools in the practitioner's toolbox. However as the scale and scope of statistical inquiry expanded with technology and increasing demands for information, a variety of new forms had to be created. The weighted likelihood is one such form that derives from Stein's celebrated work and the resulting failure of the most famous likelihood estimator of all, the sample average. In what may have been the single most important contribution coming out of Wald's theory, Stein showed that information from independent (but related) samples could be swapped to obtain estimators for the separate population means that were overall, more precise. The idea of trading bias for precision was born. In the context of modern statistical inquiry that idea assumes immense importance as in disease mapping or small area estimation, for example. In this talk I will describe the how Stein's estimator may be found as an adaptively weighted likelihood estimator (WLE). Also, I will survey recent work with a number of co-investigators, Feifang Hu, Weimin Sun, William Rosenberger, Constance van Eeden and Malay Ghosh showing what can be done with the WLE in the context of simultaneous estimation and small area estimation among other things.

WNAR Presidential Invited Address

Monday, June 11th, 10:30 – 12:00, Images

A Bird's Eye View of Earth and Its Atmosphere

Noel CRESSIE (Ohio State University)

Polar orbiting satellites remotely sense the earth and its atmosphere, producing data sets that give daily global coverage. NASA's Earth Observing System (EOS) will consist of 10-15 multi-instrument satellites within the next 10 years. By 2010, there will be on the order of 100 EOS instruments looking at the earth, each producing about 80Gb of data per day. For any given instrument on any given day, the data are many and spatially irregular. Our goal here is rather modest. For a particular data product, we wish to predict values that are spatially regular at different resolutions; such values are often used as input to general circulation models (GCMs) and the like. Not only do we wish to predict optimally, but because data acquisition is relentless, our algorithm must also process the data very rapidly. This talk presents a statistical-prediction methodology that preserves "mass balance" across spatial resolutions and computes spatial predictions and prediction (co)variances extremely fast. Similar ideas can be used when the data are multivariate and spatial-temporal. Data from the Total Ozone Mapping Spectrometer (TOMS) instrument, on the Nimbus-7 satellite, are used for illustration. The research presented in this talk is joint with Hsin-Cheng Huang, Academia Sinica, Taiwan.

Contrasting US 2000 and Canada 2001 Censuses

Monday, June 11th, 10:30 – 11:00, Halpern

Calibration Estimation in the Canadian Census

Mike BANKIER (Statistics Canada)

In the 1991 Canadian Census of Population, calibration or regression estimation was used to calculate a single set of household level weights to be used for all Census estimates based on a 1 in 5 national sample of approximately two million households. Because many auxiliary variables were available, only a subset of them could be used. Otherwise, some of the weights would have been less than one or even negative. A forward selection procedure was used to discard auxiliary variables which caused weights to be less than one or which caused a large condition number for the calibration weight matrix being inverted. Also, two calibration adjustments were done to achieve close agreement between auxiliary population counts and estimates for small areas. In 1991, the latter objective was achieved but differences between population counts and estimates at the provincial level were sometimes larger than those achieved in the 1986 Census under raking ratio estimation. This may have been the result of the accumulation of bias present in the sample or introduced through estimation. The impact of the reduction in the number of auxiliary variables available for use in the 1996 Census will be described. Finally, plans for the 2001 Census will be outlined.

Monday, June 11th, 11:00 – 11:30, Halpern
Small Domain Estimation of Missed Persons in the 2001 Census
 Peter DICK (Statistics Canada)

As part of the production of the population estimates in 2001, estimates of net missed persons by single year of age for each sex for all provinces and territories are required. While the coverage studies can provide reliable estimates of missed persons for each province and territory, the sample size is not large enough to provide reliable estimates at the detail needed. A mixture of procedures are used to produce the estimates. Direct survey estimates are used to create the provincial and territorial marginal totals. A spline smoothing method is used to produce national age sex totals for the other margin. An Empirical Bayes regression model creates the estimates for broad age groups, while a synthetic method is used for the detailed estimates. This paper discusses the methods used in the 1996 program and the improvements planned for 2001.

Monday, June 11th, 11:30 – 12:00, Halpern
Census 2000 in the U.S.: A Neighborly Overview
 Robert FAY (U.S. Census Bureau)

By June 2001, the U.S. Census Bureau has released the most important initial data products from Census 2000--counts of the population down to the census block level--although work remains on other data products, including those based on sample data. The paper summarizes the design and current status of Census 2000. A primary intent of the overview is to invite comparisons with Canada's 2001 Census. Accordingly, part of the paper reviews the design of estimation methods for sample data from Census 2000. Both Statistics Canada and the Census Bureau have researched census undercoverage of population for decades. Statistics Canada has incorporated estimates of net census undercoverage in postcensal population estimates, but most official estimates in the U.S. have not. Plans for Census 2000, however, once included adjustment for all data products. But as required by the U.S. Supreme Court, the Census Bureau released unadjusted state counts in December, 2000, for the apportionment of the House of Representatives. During the winter of 2001, the Census Bureau decided whether to statistically adjust subsequent official data products from Census 2000. The paper reports the outcome of this decision and reviews the evidence available at the time.

Unusual Stochastic Models

Monday, June 11th, 10:30 – 11:00, AQ 4150
A Stochastic Adding Machine And Complex Dynamics
 Tom TAYLOR (Arizona State University)

This paper considers properties of a Markov chain on the natural numbers which models a binary adding machine in which there a nonzero probability of failure each time a register attempts to increment the succeeding register and resets. This chain has a family of natural quotient Markov chains, and extends naturally to a chain on the 2-adic integers. The transition operators of these chains have a self-similar structure, and have a spectrum, which is, variously, the Julia set or filled Julia set of a quadratic map of the complex plane.

Monday, June 11th, 11:00 – 11:30, AQ 4150
Convergence Of Gibbs Samplers For Uniform Distributions
 Jeffrey ROSENTHAL (University of Toronto)

We consider Gibbs samplers for uniform distributions on regions. These samplers generalize the "slice sampler" Markov chains, which sample uniformly on the region underneath the graph of a density function. We examine the mathematical properties of these algorithms. In particular, we prove quantitative bounds on the convergence times of slice samplers, as well as very general qualitative convergence rates. We also show that certain convergence properties depend crucially on the smoothness of the boundary of the region itself. Finally, we describe the "polar slice sampler" modified algorithm. This is joint work with G.O. Roberts (Lancaster University).

Monday, June 11th, 11:30 –12:00, AQ 4150

Market Completeness: a Return to Order

Georghe STOICA (St. Johns University of New Brunswick)

If the cash flow space, in a two-times economy, is given by an order complete linear lattice of functions with weak unit, we then prove market completeness with respect to the topology generated by the order relation. We also give an alternative definition for market completeness, in terms of approximate martingale spanning on such linear lattices

Session in Honour of V.P. Godambe's 75th Anniversary

Monday, June 11th, 1:30 –2:00, Halpern

Smooth Simulated Composite Likelihood (SSCL) based inference for Hierarchical Models

Subhash LELE (University of Alberta)

Hierarchical models are useful as models for non-Normal spatial data, as models for data with measurement error and as models in animal breeding, among others. One of the major problems with these models is that the likelihood inference is computationally quite difficult. Composite Likelihood approach has been used for these models as a possible solution to this computational problem. In this paper, we propose an extension of the composite likelihood approach to inference for non-linear, non-Normal State Space models. Applications to non-Normal time series, spatial and spatial-time series ecological situations are discussed.

Monday, June 11th, 2:00 – 2:30, Halpern

Abundance Estimation Based On Daily Pellet Heap Counts: An Estimating Function Approach

M.B. RAJARSHI, (University of Pune)

We analyze the data on daily counts of pellet heaps of deer in a sanctuary. Statistician is required to suggest an estimate of the deer population based on these data. In an earlier work (Gadgil et al, 1986, Biometrical Journal), we had assumed that pellet lifetime distribution is exponential or if it is not exponential, the model viewed as a queue, has an FIFO discipline. Likelihood for a non-exponential distribution is almost non-tractable for estimation of life-time distribution. We propose two estimating functions for estimation of parameters of the life-time distribution. The first one is fairly easy to work out, the other is somewhat complicated but still tractable for computer analysis. We report some real-life data analysis and a few simulation studies. Lastly, an algorithm for estimating functions in the context of incomplete data, is presented. This extends the Imputation-Maximization algorithm in the case of independent observations.

Monday, June 11th, 2:30 –3:00, Halpern

Estimating Functions and Inference for Survey Data

Harold MANTEL and Milorad KOVACEVIC (Statistics Canada)

In this talk we will consider the use of estimating functions for inference with survey data. Estimation and variance estimation procedures will be described for simple descriptive parameters, for multiple linear regression parameters, and for more complex parameters such as logistic regression coefficients. Statistical properties of the resulting estimates will be investigated. Relationships to other methods of estimation will be explored, and the role of optimality will be discussed.

Spatial Statistics

Monday, June 11th, 1:30 –2:00, Images

The Good, The Bad And The Ugly: 30 Years Of Markov Random Fields

Julian BESAG (University of Washington)

Markov random fields (MRF's) are distributions defined via their local characteristics (aka full conditionals). During the 30 years since J. M. Hammersley and P. Clifford established the equivalence between MRF's and Gibbs distributions in physics, their theorem has had an enormous influence on the development of statistical ideas in many different areas, including graphical models, Bayesian image analysis, Markov chain Monte Carlo, multidimensional contingency tables and social networks, quite apart from topics that fall more obviously under the heading of spatial statistics. This talk will describe some of the successes and failures strewn along the way.

Monday, June 11th, 2:00 – 2:30, Images**Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates**

Charmaine DEAN (Simon Fraser University) and Y.C. MACNAB

Generalized additive mixed models are employed for the analysis of geographic and temporal variability of mortality rates. This class of models accommodates random spatial effects and fixed and random temporal components. Spatio-temporal models which use autoregressive local smoothing across the spatial dimension and B-spline smoothing over the temporal dimension are developed. The objective is the identification of temporal trends and the production of a series of smoothed maps from which spatial patterns of mortality risks can be monitored over time. Regions with consistently high rate estimates may be followed for further investigation. The methodology is illustrated by analysis of British Columbia infant mortality data.

Monday, June 11th, 2:30 – 3:00, Images**Accounting For Spatial Autocorrelation In Statistical Tests**

Mark DALE (University of Alberta)

Markov random fields (MRF's) are distributions defined via their local characteristics (aka full conditionals). During the 30 years since J. M. Hammersley and P. Clifford established the equivalence between MRF's and Gibbs distributions in physics, their theorem has had an enormous influence on the development of statistical ideas in many different areas, including graphical models, Bayesian image analysis, Markov chain Monte Carlo, multidimensional contingency tables and social networks, quite apart from topics that fall more obviously under the heading of spatial statistics. This talk will describe some of the successes and failures strewn along the way.

Robust InferenceMonday, June 11th, 1:30 – 2:00, AQ 4150**On The Robustness Of Empirical Likelihood Ratio Confidence Intervals For Location**

Min TSAO (University of Victoria)

In this talk I will examine the robustness of empirical likelihood ratio (ELR) confidence intervals for the mean and $\$M\$$ -estimate. The ELR interval for the mean is shown to have an asymptotic breakdown point of zero. A formula for computing the breakdown point of the ELR interval for $\$M\$$ -estimate is also given. Relative advantages of the ELR interval to the commonly used confidence intervals based on the asymptotic distribution of the $\$M\$$ -estimator are examined through a numerical study.

Monday, June 11th, 2:00 – 2:30, AQ 4150**Robust Sequential Designs for Misspecified Models**

Doug WIENS (University of Alberta)

We will discuss recent work on regression and clinical designs which are intended to be robust against misspecified response functions. An example is that in which the experimenter will fit a Michaelis-Menten model to the responses, but wishes a design which will protect against biases in the predicted values if the true response function has a similar but different shape. Another example is that of assigning subjects to treatments in a clinical trial, when the effects of the prognostic factors on the response variable are incorrectly modeled. One seeks a design which will protect against the ensuing increase in the MSE of the estimates of the treatment contrasts, while maintaining a reasonable level of balance.

Monday, June 11th, 2:30 – 3:00, AQ 4150**On The Role Of Redescending M-Estimates In Computer Vision (And Other Areas)**

David TYLER (Rutgers University)

There is a long history on the application of robust statistical methods in computer vision. Many of the early robust methods used in the vision community were proposed by vision researchers and arose from a need for such methods in vision research. Such methods include the Hough Transform and RANSAC. More recently researchers in computer vision have become aware of parallel developments on robust methods in statistics, in particular M-estimates and least median of squares, and have incorporated these methods into their tool box.

Contributed Session: Survival AnalysisMonday, June 11th, 1:45 – 2:00, AQ 4140**Selection Of Survival Predictors Using A Nonparametric Method**
Marylène TROUPÉ and J. VAILLANT (Université des Antilles-Guyane)

The aim of the study is to select covariables with a high predictive value using mutual information between a survival variable and covariables when parametric models do not fit the observed phenomena. The method, based on information conservation permits the aggregation of survival curves corresponding to different modalities of the same selected predictor. The information rate carried out by covariables is decomposed similarly to the analysis of variance and then is analyzed. An application to breast cancer using data from 1 304 patients who were followed over a period of ten years will be presented.

Monday, June 11th, 1:45 – 2:00, AQ 4140**Checking the Fit of a Discrete-time Multi-level Survival Model**
Jennifer ORLICH and Susan MURPHY (University of Michigan)

Discrete-time multi-level survival analysis is commonly employed to analyze data that is longitudinal, censored, and clustered. While implementing a discrete-time multi-level survival model is straightforward, model checking is not well developed. Two graphical methods for assessing the fit of this model will be provided. The first method verifies that the predicted probability of the event, based on the model, is approximately the estimated probability of this same event based on the data. The second method checks if the predicted probabilities derived from the model and cumulated over time reflect the corresponding cumulated number of events that are observed in the data. We demonstrate and give arguments for how the plot should appear for a well fitting model and provide illustrations of how the methods detect lack of fit. The situation where most of the predicted probabilities are close to zero is of special interest.

Monday, June 11th, 2: 00 – 2: 15, AQ 4140**Efficient Estimation for the Proportional Hazards Model with Left Truncated and "Case1" Interval-Censored Data**
Jong Sung KIM (Portland State University)

The maximum likelihood estimator (MLE) for the proportional hazards model with left-truncated and "Case 1" interval-censored data is studied. Under appropriate regularity conditions, the MLE of the regression parameter is shown to be asymptotically normal with a root-n convergence rate and achieves the information bound, even though the difference between truncation time and censoring time of the MLE of the baseline cumulative hazard function converges only at rate $n^{1/3}$. Two methods to estimate the variance-covariance matrix of the MLE of the regression parameter are considered. One is based on a generalized missing information principle and the other is based on the profile information procedure. Simulation studies show that both proposed methods work well in terms of bias and variance for samples of moderate sizes. An example is provided to illustrate the methods.

Monday, June 11th, 2:15 – 2:30, AQ 4140**Estimating Wildlife Survival Rates With Time Series Of Age-Structure Data**
Mark UDEVITZ (Alaska Biological Science Center)

Age-structure data have a long history of use for estimating survival rates in wildlife populations. Survival rates can be estimated from age class proportions in either the standing population (standing age-structure data) or the natural deaths (ages-at-death data). Most estimators have required the often unreasonable assumption of a stable age structure. Age structures will become stable only if recruitment and survival rates remain constant for a long enough period of time. Recently, a new set of survival rate estimators was developed that use both standing age-structure and ages-at-death data from a single year, thereby eliminating the requirement for age-structure stability. Now, I generalize the likelihoods used to derive those estimators to consider multiple years of age-structure data. Using multiple years of either type of age structure data also allows relaxation of the stability assumption. Standard likelihood-based model selection procedures can be used to test hypotheses about age structure stability and age-related patterns in survival rates and to identify the most appropriate model for estimating those rates.

Monday, June 11th, 2:30 – 2:45, AQ 4140

A Partially Linear Proportional Hazards Model

Xuwen LU (Agriculture and Agri-Food Canada)

The proportional hazards regression model usually assumes that the covariate has a log-linear effect on the hazard function. In this paper, we consider a semiparametric survival model with flexible covariate effects. We suppose the baseline hazard function is parameterized, while the risk function associated with covariates is modeled in a semiparametric way. A "one-step algorithm" is used to estimate the nonparametric components and the parametric components by maximizing the local and the global likelihood, respectively. Using the local linear method, estimates of the unknown parameters and the unknown covariate function are proposed, and their asymptotic distributions obtained. The performance of "one-step algorithm" is compared with that of the fully iterated algorithm and the profile likelihood method. A simulation study illustrates the algorithms and the models.

Monday, June 11th, 2:45 – 3:00, AQ 4140

Using Auxiliary Information In Interval-Censored Survival And Event History Data Analysis

Jaroslav HAREZLAK (Indiana University School of Medicine)

In common survival or event history analysis, whenever an event takes place we observe the time of the event with reasonable accuracy. In certain situations however, we only know that the event has happened between two observation points. A standard analysis in that situation is to treat the event as interval censored. The available interval censoring methodology tends to be less efficient when the intervals are wide. In a study of sexually transmitted diseases (STD), we followed a group of young women who were either infected with an STD or were partners of infected subjects (i.e. contacts). The event of interest was a subsequent STD infection. From the scheduled follow-up visits (at one, three, five and seven months), we only know if the re-infection occurred between these visits. In addition to the follow-up data, we also collected daily diary data on the sexual activities of the study subjects. With the auxiliary information provided by the diaries, we propose to use the times of possible infection (i.e. sexual encounters) to improve the estimation of the time to a subsequent infection. A sample of sexual encounters is drawn at random from the interval of re-infection. Each sample is then analyzed using one of the available methods for the right-censored data, and the estimates are combined in an appropriate manner. Our method is compared with other common methods used for interval-censored data. Extension is also proposed for the multiple events per subject. In our simulations, we utilize non-parametric, semi-parametric and parametric models. Our results indicate a good performance of the proposed method in all considered settings. Analysis of the motivating example indicates the higher risk of an infection in the subjects who were infected at baseline than the contacts.

WNAR Student Paper Presentations

Monday, June 11th, 1:30 – 2:00, AQ 4130

Vector-valued Markov Models for Analyzing Disease History Data with Multi-variate Outcomes and Incomplete Observations

Shenghai ZHANG (University of Waterloo)

A number of dependent multi-state processes arise when several response variables are required to measure the outcome of interest. Each subject is periodically visited and outcome measures have been recorded simultaneously. The information a state the subject is in is incomplete. Although some methods analyzing such data were provided previously, they are only for bivariate irreversible processes and the same time interval of visiting was assumed. The purpose of this paper is to develop a methodology to overcome these difficulties. The methods provided here is to estimate not only a treatment effect but also transition probabilities between states. First, the joint intensity for continuous time vector-valued Markov process is discussed and formulas which describes the relationship between transition probabilities and intensities are established. Based on a family of multi-variate exponential distribution which is given in this paper, the joint intensities can be specified by the marginal intensities which are usually to be obtained easily in practical applications. Moreover, a procedure to estimate interested parameters is described.

Monday, June 11th, 2:00 – 2:30, AQ 4130**Comparison of Statistical Methods for High Throughput Screening Data**

Yuanyuan WANG, Hugh CHIPMAN, and William WELCH (University of Waterloo)

High throughput screening (HTS) is often used in pharmaceutical drug discovery to measure the biological activity of compounds against various biological targets. The data generated by HTS have tens of thousands to hundreds of thousands of compounds. Also, chemists and biologists are interested in predicting the highest activity compounds, which are relatively rare. I will compare the performances of various statistical models on several HTS data sets in terms of ability to predict extreme cases. I will also identify the special features of HTS data that lead to certain types of models doing well.

Monday, June 11th, 2:30 – 3:00, AQ 4130**Multi-List/Capture-Recapture Methods With Incomplete And Partial Stratification Applied To Estimating Forest Fire Incidence**Jason SUTHERLAND, C.J. SCHWARZ, J.G. FALL, B. DORNER, K.P. LERTZMAN
(Simon Fraser University)

Multi-list capture-recapture methods are commonly used to estimate the size of elusive populations. When possible, stratification is a useful way to reduce the biases caused by heterogeneity in the probability of list membership among members of the population. In this paper, we develop a method to deal with cases when not all lists are active in all strata. The motivation for this paper is estimating the number of forest fires that have taken place in recent history, where individual trees function as lists recording fires. This paper presents a generalized approach for more than two lists when not all lists cover all strata and presents an estimator of incidence of events. We allow for list dependencies and differential probabilities of ascertainment in each list using a log-linear modeling framework.

IMS Special Invited Session IMonday, June 11th, 3:30 – 5:00, Halpern**Chain Graph Models and Their Causal Interpretations**

Steffen LAURITZEN and Thomas S. RICHARDSON (Aalborg University, Denmark)

Chain graphs are a natural generalization of directed acyclic graphs (DAGs) and undirected graphs. However, the apparent simplicity of chain graphs belies the subtlety of the conditional independence hypotheses that they represent. There are a number of simple and apparently plausible, but ultimately fallacious interpretations of chain graphs that are often invoked, implicitly or explicitly. These interpretations also lead to flawed methods for applying background knowledge to model selection. We present a valid interpretation by showing how the distribution corresponding to a chain graph may be generated as the equilibrium distribution of dynamic models with feedback. These dynamic interpretations lead to a simple theory of intervention, extending the theory developed for DAGs. Finally, we contrast chain graph models under this interpretation with simultaneous equation models which have traditionally been used to model feedback in econometrics.

Business and Industry Section Special Invited SessionMonday, June 11th, 3:30 – 5:00, Images**Accelerated Testing: A Method for Obtaining Reliability Information Quickly**

Bill MEEKER and Luis A. ESCOBAR (Iowa State University)

Accelerated tests are used to obtain timely information on products reliability. Changes in technology, the calls for rapid product development, and the need to continuously improve product reliability have combined to increase the need for developing improved methods for accelerated testing. Laboratory tests with increased use rates or higher than usual levels of accelerating variables like temperature or voltage are used to accelerate failure mechanism. Then the results are used to make predictions about product life or performance over time at use or design conditions. The predictions involve extrapolation in several dimensions. Interesting statistical problems arise in modeling physical phenomena, use of engineering/physical information, planning accelerated tests, and quantifying uncertainty. This talk reviews the basic physical and statistical models and methods used in accelerated testing. Current research in this area will be outlined and areas for future research will be described.

Gene Ancestry, Identity and Allelic AssociationsMonday, June 11th, 3:30 – 4:00, AQ 4130**A Score Test For The Linkage Analysis Of Qualitative And Quantitative Traits Based On Identity By Descent Data**
Sandrine DUDOIT (MSRI, Berkeley)

Genetic factors contribute to many human diseases by conferring susceptibility or resistance, and by affecting the severity and progression of the disease. Linkage mapping has proved to be a powerful tool for localizing genes for hundreds of rare Mendelian diseases (e.g. Huntington's disease, cystic fibrosis). In contrast, the identification of genes influencing more common diseases such as diabetes or hypertension has proceeded slowly. Such diseases are termed complex for a number of reasons, including diagnostic uncertainties (late age-of-onset), unknown and non-Mendelian modes of inheritance, and the involvement of multiple genes. The public health importance of understanding the genetic basis of complex diseases has prompted increased interest in the development of statistical methods for genetic mapping. This talk presents a unified likelihood-based approach for the genetic mapping of complex human traits, qualitative and quantitative, using identity by descent (IBD) data from small pedigrees. The unified approach considers the likelihood of IBD data conditional on phenotypes, and tests for linkage between a marker locus and a gene influencing the trait using a score test. For a given pedigree type, the form of the linkage score statistic is determined by the second largest eigenvalue and corresponding eigenvector(s) of the transition matrix for IBD configurations along the length of a chromosome. Compared to existing approaches, the linkage score test provides a robust and powerful method for the genetic mapping of complex traits. Applications of microarrays in the context of genetic mapping will also be discussed.

Monday, June 11th, 4:00 – 4:30, AQ 4130**Distinguishing Between Causative And Linked Genetic Loci**
Laura LAZZERONI (Stanford University)

Affected sib-pair designs and transmission-disequilibrium tests are standard methods used to identify and map genetic risk factors. When such tests yield positive results for one or more polymorphisms, it is of interest to determine which, if any, of the tested loci are causally related to the disease phenotype and which are merely linked to a causative locus. This talk will discuss what information the data provide to help make this determination.

Monday, June 11th, 4:30 – 5:00, AQ 4130**A New Statistical Method For Haplotype Reconstruction**
Matthew STEPHENS (University of Washington)

Current routine experimental methods for determining the genetic types of individuals in a population ("genotyping") typically do not provide haplotype information - i.e. information on which of the two alleles at each locus was inherited from the individual's mother, and which was inherited from the father. Haplotype information is important for many analyses of fine-scale molecular genetics data, including disease mapping and inferring population histories. Haplotypes can be obtained, at considerable cost, experimentally, or (partially) through genotyping additional family members. Alternatively, a statistical method can be used. We describe two existing statistical approaches to the problem, and propose a new statistical method, that improves substantially on current algorithms by incorporating ideas from population genetics theory. (Joint work with Nicholas Smith and Peter Donnelly)

Survey Methods Contributed Session: Imputation and Variance EstimationMonday, June 11th, 3:30 – 3:50, AQ 4150**Inference for Regression Parameters under Imputation for Missing Data**

David HAZIZA (Statistics Canada)

A population mean can be estimated unbiasedly under mean imputation for missing data and uniform response, but the imputed estimator of a regression parameter is generally biased. Following the approach of Skinner and Rao (1999), we obtain a bias-adjusted estimator of a regression parameter under an arbitrary design. We derive consistent variance estimators, using a method introduced by Fay (1991) in which the usual sample-response path is reversed. We also consider the model-assisted framework in which case the imputed estimator is design-model unbiased and derive consistent variance estimators.

Monday, June 11th, 3:50 – 4:10, AQ 4150**Variance Estimation for Estimating Equations in the Presence of Missing Values**

Wesley YUNG, Mike HIDIROGLOU and J.N.K. RAO (Statistics Canada)

Population parameters of interest can be expressed as solutions to appropriate population estimating equations. Sample estimates of these parameters can be obtained by solving suitable sample estimating equations. Hidiroglou, Rao and Yung (1999) showed how variance estimates can be obtained for solutions of estimating equations which incorporates weighting adjustments due to auxiliary information, such as Generalized Regression (GREG) estimation weights. In this paper, we extend the results of Hidiroglou, et al. (1999) to the case where adjustments are also made for nonresponse or imputation methods are applied to account for missing values.

Monday, June 11th, 4:10 – 4:30, AQ 4150**Variance Estimation for the Annual Survey of Manufactures**

Wisner JOCELYN and Garriguet DIDIER (Statistics Canada)

The annual Survey of Manufactures collects financial and commodity information. It is being currently redesigned. One of the requirements is to provide commodity details on a census basis. A simultaneous two-phase stratified design is used. Limited commodity details are collected through the first phase sample where extensive commodity details are collected for the sub-sample. The second phase detailed commodity information collected is combined with the limited first-phase commodity data and financial auxiliary information available for the whole population to derive regression models. The models are used to produce commodity estimates for the non sampled units in order to satisfy the census requirement. In this paper we discuss the variance estimation process for this particular two-phase design where we use the sampled data to randomly assign commodities to each non sampled units. We compare the efficiency of such a scheme to Donor imputation for each non sampled unit, via a Monte-Carlo simulation. Some of the results along with recommendations for the implementation of the strategy will be discussed.

Monday, June 11th, 4:30 – 4:50, AQ 4150**Variance Estimation in the Presence of Imputation**

Charlie ARCARO and Wesley YUNG (Statistics Canada)

The problem of missing or incomplete survey data is a common problem faced by statistical agencies. In many surveys conducted by Statistics Canada, missing data is dealt with by replacing (or imputing) them with values derived from current or historical data. While imputation methods can produce unbiased point estimates, conventional variance estimators can lead to underestimation of the true variance if the imputed values are treated as true values. Several methods have been proposed to compute the variance due to imputation. We study two of them; one based on an adjusted jackknife and the other based on a model assisted approach. We compare the properties of these two methods using theoretical and empirical arguments.

Contributed Session: Optimal DesignsMonday, June 11th, 3:30 – 3:45, AQ 4140**A Two-Stage Phase II Design with a Multinomial Endpoint**

Alex SMITH, Katherine S. PANAGEAS, Mithat GONEN, Paul B. CHAPMAN (Memorial Sloan-Kettering Cancer Center)

Phase II clinical trials in oncology are performed to evaluate the therapeutic efficacy of a new treatment regimen. A common measure of efficacy for these trials is the proportion of patients who obtain a response measured by tumor shrinkage. It is standard practice to classify this response into the following categories: (1) Complete Response (CR); (2) Partial Response (PR); (3) Stable Disease (SD); and (4) Progression of Disease (PD). Tumor response is then treated as a binary variable whereby patients who achieve either a CR or a PR are considered responders and all others non-responders. Two-stage designs that allow for early termination of the trial if the treatment shows little efficacy such as those of Gehan or Simon give equal weight to a CR and a PR. However, a CR is defined as complete disappearance of the tumor and is more likely to result in a survival advantage. We argue that a CR should be given more weight and hence, we propose a two-stage design with a multi-level endpoint, i.e., CR, PR, or non-responder. This design is an extension of Simon's optimal two-stage design and is based on a multinomial model. Comparisons are made with Simon's optimal two-stage design and new designs for common parameter sets are tabulated

Monday, June 11th, 3:45 – 4:00, AQ 4140**Constructing Optimal Designs With Constraints**

Saumendranath MANDAL, K.C. CARRIERE (University of Alberta)

There are various classes of optimization problems in which we wish to find optimizing probability distributions. Optimal regression designs are a particular example. In this work, we focus on the problem of finding constrained 'approximate' designs maximizing a criterion subject to constraints. The constraints are zero covariances (or correlations) of the estimates of the parametric functions of interest. In many regression designs it is desired to estimate parameters independently of others, by making covariances or correlations between relevant parameter estimates as small as possible. We approach this problem by initially formulating the Lagrangian with two constraints, but removing the Lagrange parameters through substitution, transforming the constrained optimization problem to that of maximizing three functions of the design weights simultaneously. They have a common maximum of zero that is simultaneously attained at the constrained optimal design weights. In the case of one constraint, the approach transforms to a problem of maximizing two functions simultaneously. To find the optimal design, we consider a class of multiplicative algorithms, indexed by a function $f(\cdot)$, satisfying certain conditions (positive and strictly increasing) and may depend on one or more free parameters. These iterations neatly satisfy the constraints of the problem. We also consider techniques for improving the convergence rates with some suitable choices of the function $f(\cdot)$. The approach can be extended to the case of more than two constraints.

Monday, June 11th, 4:00 – 4:15, AQ 4140**Trial Design For Negative Binomial Responses Based On ANCOVA**

Wei WEI and R.J.COOK (University of Waterloo)

Randomized trials are more frequently being designed based on count data. Between patient variability in the event rates means that mixed models, such as the negative binomial model, are suitable for the analysis. When baseline counts are available analyses should be conditional on these counts leading to the count data analogue of ANCOVA. We develop design criteria for such settings, as well as for settings where patients are selected for inclusion based on the baseline count. Sensitivity of the design criteria to model misspecification is also investigated. We contrast our design methods with those of McMahon et al (1994).

Monday, June 11th, 4:15 – 4:30, AQ 4140

Two-Stage Designs For Genome Scanning

David VERBEL, Jaya M. SATAGOPAN, E.S. VENKATRAMAN, Colin B. BEGG (Memorial Sloan-Kettering Cancer Center)

The goal of this article is to describe a two-stage design which maximizes the power to detect gene-disease associations using genome scanning where the principal design constraint is the total cost, represented by the total number of gene evaluations rather than the total number of patients. In the first stage, all genes of interest are evaluated on a subset of patients. The most promising genes are then evaluated on additional subjects in the second stage. This will eliminate wastage of resources on genes unlikely to be associated with disease based upon the results of the first stage. We consider the case where the genes are correlated and the case where the genes are independent. Using simulation results it is shown that, as a general guideline when the genes are independent or when the correlation is small, screening all the genes on approximately 25% of the patients and then evaluating the most promising 10% of the genes on the remaining patients provides near-optimal power for a broad range of parametric configurations.

Monday, June 11th, 4:30 – 4:45, AQ 4140

On Modeling of Exceedances of Riverflows Over Thresholds by Generalized Pareto Distribution

Vartan CHOULAKIAN (Université de Moncton), M. A. Stephens (Simon Fraser University)

The Generalized Pareto distribution is used in hydrology to model exceedances in flood control; exceedances are heights of rivers above a threshold T , and fitting the GPD has been suggested as a method of deciding T . For this purpose one needs to be able to test the GPD. Tests will be given and the method investigated for exceedances of 238 rivers in Canada.

Monday, June 11th, 4:45 – 5:00, AQ 4140

Event History Analysis and Ignorable Sampling

Christian BOUDREAU (University of Waterloo)

A considerable amount of event history analysis data is collected through longitudinal panel, or cohort studies. The goal is to understand the different events that individuals experience over time. Examples of such events include marriage, divorce, fertility, unemployment, etc. Event history analysis deals with the modeling and the analysis of such processes. Note that survival analysis is both a special case and a cornerstone of event history analysis. However, longitudinal studies generally involve the use of complex survey designs such as clustering and stratification. These features of complex surveys complicate the statistical analysis and will be the topic of this talk. In particular we explore the case of ignorable sampling designs. With such designs it is possible to simplify the statistical inference by carrying unweighted analyses. However, dependence between observations introduced by clustering and other sampling techniques must be taken into account when computing variance estimates of event history quantities. This is achieved by using the theory of estimating equations.

Measurement ErrorTuesday, June 12th, 8:30 – 9:00, Halpern**Extended Partial Likelihood Methods For Failure Time Models, With Measurement Error In A Time-Varying Covariate**

Donna SPIEGELMAN and R. LOGAN (Harvard School of Public Health)

Occupational and environmental epidemiologists often assess prospective associations of cumulative exposure variables with chronic disease outcomes. These variables are usually measured with moderate to considerable error. An empirically verified model for the error in the point exposures at single instants of time is identified from validation data, and the model for the measurement error in the cumulative exposure variable is derived from this model for the point exposures. An extended partial likelihood is then constructed and maximized, and estimates are obtained which fully account for the measurement error process, including uncertainty about it, assuming proportional hazards for the perfectly measured data. Identifiability, consistency and asymptotic normality of the estimator of the parameter of interest are considered for several common study designs. This method is compared to other methods currently proposed for addressing related problems, models and data structures. A detailed example illustrates these methods, examining mortality from lung cancer and cumulative exposure to radon progeny among underground miners in New Mexico, USA. A gamma measurement error model fit the annual individual exposure data well. For a 500 WLM increase in cumulative radon exposure, the estimated lung cancer mortality rate ratios, uncorrected and corrected for exposure measurement error, were 5.8 (95% confidence interval (CI) 3.0-11.1) and 16.7 (95% CI 5.2, 53.9), respectively.

Tuesday, June 12th, 9:00 – 9:30, Halpern**Measurement Error Regression And Nonparametric Maximum Likelihood**

Dan SCHAFFER (Oregon State University)

Addressing statisticians and biostatisticians who are somewhat familiar with this problem, I shall provide a review of the statistical issues associated with explanatory variable imprecision, my perception of the current state of useful statistical methods, and a fairly casual discussion of nonparametric maximum likelihood methodology for various types of regression with regressor measurement error. In many epidemiological applications the most crucial part of the analysis is modeling the dose or exposure uncertainties. Some discussion will be provided about this, including a checklist for classical and Berkson modeling.

Tuesday, June 12th, 9:30 – 10:00, Halpern**Case-Control Analysis with Partial Knowledge of Exposure Misclassification Probabilities**

Paul GUSTAFSON (University of British Columbia)

Consider case-control analysis with a dichotomous exposure variable that is subject to misclassification. If the classification probabilities are known, then methods are available to adjust odds-ratio estimates in light of the misclassification. We study the realistic scenario where reasonable guesses, but not exact values, are available for the classification probabilities. If the analysis proceeds by simply treating the guesses as exact, then even small discrepancies between the guesses and the actual probabilities can seriously degrade odds-ratio estimates. We show that this problem is mitigated by a Bayes analysis which incorporates uncertainty about the classification probabilities as prior information.

Survey Methods Contributed Session: Survey Applications

Tuesday, June 12th, 8:30 – 8:50, AQ 4150

The Redesign of the Monthly Wholesale and Retail Trade Survey of Statistics Canada

Hélène BÉRARD (Statistics Canada)

Statistics Canada conducts a major survey known as the Monthly Wholesale and Retail Trade Survey (MWRTS). The survey produces estimates based on monthly data collected for sales and inventories at various province and industry levels. The estimates derived from the survey form a substantial portion of the monthly estimates for the Gross Domestic Product (GDP). The sales trend represents an important economic indicator. The monthly estimates from the Retail sector are used to calculate the equalization payments (transfer of funds from the federal to the provinces). Furthermore, MWRTS is the frame for the Quarterly Retail Commodity Survey (QRCS). The current MWRTS uses a stratified sample design that was elaborated in 1988. Many improvements since then have strengthened and maintained data quality, however, there was a need for a redesign. A redesign is required to provide estimates for the new North American Industry Classification System (NAICS), and to take full advantage of the availability of administrative data from the Goods and Services Tax program. The redesign also addresses the need to reduce cost and respondent burden, update computer systems, and to harmonize various concepts used by other annual surveys that have been redesigned. This talk presents the challenges associated with redesigning a monthly survey for the wholesale and retail trade. I will discuss the sampling plan, the stratification procedure using administrative data, the sample selection, and the sample update mechanisms. I will also explain the additional constraints driven by the QRCS.

Tuesday, June 12th, 8:50 – 9:10, AQ 4150

Challenges in Measuring Electronic Commerce

Patricia WHITRIDGE (Statistics Canada)

Interest in electronic commerce has been increasing over the last few years. Statistics Canada undertook a survey to measure the amount of sales taking place over the Internet for reference year 1999. It was decided to take advantage of the sample design of an existing economy wide establishment survey. A questionnaire was developed with questions about the use of technology in businesses. Information was collected on the availability of computers, the use of e-mail and access to the Internet. Businesses were also asked if they had a Web page, if it could be used for sales, and the amount of those sales. The data were released in the summer of 2000. Now, many issues are being discussed concerning the most effective methods to use to estimate the total sales over the Internet. For example, should we be asking questions about Web pages to establishments, or rather to the head offices? Is a general survey of businesses the best method? Could we draw a sample of businesses and then research the Web pages ourselves? The presentation will begin with some background information about the 1999 survey, including its challenges. Then, current plans for the 2000 survey, which is underway, will be discussed. Some issues that have already surfaced will be identified, and different scenarios for the longer-term future will be examined. The presentation will conclude with some general observations from Statistics Canada's experience measuring electronic commerce.

Tuesday, June 12th, 9:10 – 9:30, AQ 4150

Sample Design of the Canadian Mental Health Survey

Yves BÉLAND (Statistics Canada)

As part of the Canadian Community Health Survey (CCHS) biennial strategy, the provincial survey component of the CCHS first cycle will focus at different aspects of the mental health and well-being of Canadians living in private occupied dwellings. Moreover, this new survey will collect data on both positive and negative factors affecting mental health, the utilization of mental health care services as well as data on social impacts and the costs associated with mental health. All this will be rounded out with the collection of a series of social and demographic characteristics. The sample planned for this survey will consist of 30,000 respondents selected from an area probability frame under a multistage cluster design where the dwelling is the final sampling unit. One person aged 15 years old or over will be selected at random in the sampled households. Plans are to oversample people in the 15-24 age group by increasing their probability of selection in order to improve their representativity in the sample. Data collection will begin in January 2002 and will extend over 12 months to eliminate possible seasonal effects on some key characteristics. All interviews will be conducted face-to-face using a computer-assisted application. (However some sensitive topics such as suicidal thoughts will be collected through a self-completed paper questionnaire.) The averaged interview length is expected to be 90 minutes. This paper will describe several key aspects of the sample design of this new and exciting survey as well as some methodological challenges encountered in the testing and validation of the questionnaire. In parallel, a brief overview of the sample design of a special survey on the mental health and well-being of the Canadian Armed Force members will be given.

Tuesday, June 12th, 9:30 – 9:50, AQ 4150**What's New for the 1995-1997 Canada Life Tables**Brad W. THOMAS, Doreen DUCHESNE (STC), Marek WYSOCKI (STC), Robert BOURBEAU (University of Montréal)
(Statistics Canada)

A life table shows expected mortality experience for a hypothetical cohort of 100,000 persons born at the same time and subjected to the age-sex-specific mortality rates which an actual population experienced. The last set of detailed life tables for Canada and the provinces was for the 1990-92 period (Statistics Canada Cat. No. 84-537), based on the 1991 Census of Population (CEPOP) and mortality data for 1990 through 1992. The availability of 1996 CEPOP and 1997 mortality data now allows us to produce detailed life tables for the 1995-97 period. The previously cited Statistics Canada publication describes in detail the methodology of the 1990-92 life tables, to which we made a number of improvements for the 1995-97 life tables. The focus of this presentation will be on the most significant improvements, namely the use of a model for estimating probabilities of dying at ages over 85 (Coale and Kisker, 1990), and a new methodology for estimating probabilities of dying for ages 0-4 which uses 1 January population estimates for 1995-98 from Statistics Canada followed by a new method of ensuring a monotonic decreasing trend in the probability at the provincial level.

WNAR Student PresentationsTuesday, June 12th, 8:30 – 8:50, AQ 4130**Detection of Genetic Heterogeneity Among Families with Affected Relatives**

Lucia MIREA and Shelley B. BULL (Samuel Lunenfeld Research Institute, University of Toronto)

Current linkage studies aim to identify chromosomal sites that may harbor susceptibility genes for complex diseases characterized by variability in symptoms, prevalence, and/or mode of inheritance. Statistical methods assess whether the observed allele-sharing distribution, measuring genetic similarity among affected family members, is different from that expected by chance alone. These methods suffer from a loss of statistical power when systematic heterogeneity is present among the allele-sharing distribution of sampled families. To detect heterogeneity, at a given chromosomal position, we propose two approaches that use binary family-level covariates. Clinical sub-diagnosis, ethnicity and age of disease onset are examples of covariates that may indicate genetic differences among families. The first approach is a likelihood ratio test (LRT), based on an extension of the Kong and Cox linkage likelihood (1997, *Am J Hum Genet* 61:1179-1188) that incorporates the effects of a binary covariate. The LRT method compares the model with covariates to the null model assuming no covariate effect. In the second approach (PERM), permutation methods are applied to test for differences in mean linkage scores between covariate-defined subgroups. The LRT involves computation of constrained maximum likelihood estimates and may not have the expected asymptotic distribution. In contrast, the PERM test makes no distributional assumptions and does not involve explicit parameter estimation. Computational demands of both tests are similar. To compare the properties of these methods, we have designed and conducted a simulation study to estimate type I error and power in family studies with affected sibling and cousin pairs. A range of scenarios were evaluated including dominant and recessive models of inheritance with various levels (20%, 40% and 50%) of locus heterogeneity.

Tuesday, June 12th, 9:00 – 9:30, AQ 4130**Uncertainty In Outcomes In Discrete Time Proportional Hazards Modeling**

Amalia MEIER and Barbra RICHARDSON (University of Washington)

Estimation of failure time has typically been done when failure is observed with certainty. However, in a context in which false positive and false negative test results occur, a time to event analysis is also of interest. A procedure is desired that addresses information loss due to imperfect sensitivity/specificity in the outcome. Richardson and Hughes (2000) provide an algorithm for estimating the product limit of discrete time data with uncertain outcomes. Following their work, a method is developed that permits the estimation of both the survival curve and covariate effects. Of particular interest are differences in survival between treatment groups, under the assumption of proportional hazards. It is shown that the greatest bias in both hazard and covariate effect estimation stems from failure to account for false positive tests. However, simulations show that correct specification of these false testing rates permit (nearly) unbiased estimation of both the baseline hazard and covariate effects. A graphical comparison to Cox models is made, evaluating the robustness of the estimates with regard to misspecification of the false testing rates. The method is most applicable to research in the developing world, where the accuracy of tests can be limited by cost and technological difficulty.

Tuesday, June 12th, 9:30 – 10:30, AQ 4130**Models For Zero-Inflated Count Data Using The Neyman Type A Distribution**

Melissa DOBBIE and A.H. WELSH (Australian National University)

In animal abundance studies, where, for example, the estimated abundance of a rare animal may be of primary importance, often a large proportion of the data is zero, leading to an excess of zeros in relation to a Poisson or negative binomial distribution. We refer to this phenomenon as zero-inflation. Ignoring zero-inflation implies that the underlying distributional assumptions will not be met. This can affect the results of an analysis, and hence lead to incorrect conclusions concerning the data. Ridout et al (Proc. XIXth IBC Cape Town, 1998, 179-190) review and compare several approaches to modeling count data with an excess of zeros in relation to the Poisson distribution. One approach models such data by a mixture of Poisson distributions; the most well-known and commonly used mixture is the negative binomial distribution. However, zero-inflated count data typically exhibit either a half-mode (at zero) or one-and-a-half modes (at zero and at a count greater than one) implying that the unimodal negative binomial distribution may be unsuitable. The Neyman type A distribution (Neyman, Ann. Math. Stat., 1939, 10, 35-57) is also generated by a mixture of Poisson distributions but is flexible enough to model multimodal data (Barton, Trab. Estad., 1957, 8, 13-22). We explore the possibility of modeling zero-inflated count data using the Neyman type A distribution. Three parameterizations of the distribution are extended to allow their parameters to depend on covariates. To illustrate the methodology, we develop models for counts of Leadbeater's possum (*Gymnobelideus leadbeateri*) of which 63% are zeros. Using the method of Cox (J. Royal Stat. Soc. B, 1962, 24, 406-424) to test non-nested hypotheses, we then formally compare the Neyman type A models and also compare them to a model from a competing family.

Small Area EstimationTuesday, June 12th, 10:30 – 11:00, AQ 4150**Second Order Asymptotics in Small Area Estimation: Accurate**

David SMITH, G.S. DATTA and J.N.K. RAO (University of Georgia)

A simple area level linear mixed model, popularly known as the Fay-Herriot model, is widely used to produce model-based small area estimators; in particular, empirical best linear unbiased prediction (EBLUP) estimators of small area means. The EBLUP estimators depend on the estimator of the between area variance component. Fay and Herriot proposed a simple iterative method to estimate this variance component, not requiring normality. In this paper we obtain a second order approximation to the MSE of EBLUP estimators for the Fay-Herriot iterative method and an estimator of MSE also correct to second order, and compare our results to previous second order approximations based on the method of moments, maximum likelihood (ML) and residual maximum likelihood (REML). We also obtain a correction to normal theory confidence intervals, based on EBLUP estimator and its estimator of MSE, to achieve the target coverage probability asymptotically as the number of small areas increases.

Tuesday, June 12th, 11:00 – 11:30, AQ 4150**Small Area Estimation using Multiple Surveys**

Narasimha PRASAD and Sharon LOHR (University of Alberta)

In this paper, we provide a general method for small area estimation when information is derived from multiple surveys. Fuller and Harter (1987) and Battese, Harter and Fuller (1988) proposed nested-error regression models for small area estimation. We adopt a multivariate approach to extend these models for combining information from several surveys, with possibly different sampling designs, in order to improve the accuracy of small area estimates. Computational algorithms are discussed, and a simulation study indicates that if quantities in the different surveys are sufficiently correlated, gains in efficiency can be achieved.

Tuesday, June 12th, 11:30 – 12:00, AQ 4150

Benchmarking of Small Area Estimators in a Bayesian Framework

Avi SINGH and Ralph FOLSOM (Research Triangle Institute)

It is often desirable in practice to benchmark the small area estimates in that they sum to the direct large area estimates. This provides robust estimates that protect against possible model mis-specification as well as possible overshrinkage. Moreover, benchmarking provides internal consistency of small area estimates with direct aggregate level estimates which are often published before the small area estimates. For random benchmarks, it is not possible to incorporate these constraints in the (Hierarchical) Bayes theory, because the benchmarks are based on the data used in the likelihood, and are not part of the prior. However, the class of estimators can be constrained to satisfy the benchmarks. To this end, we propose a generalization of the constrained Bayes theory of Ghosh (1992) in that benchmarks need not be posterior means of moments of the histogram of small area parameters. For nonrandom benchmarks, the parameter space becomes constrained, and for this purpose we propose, for the Metropolized MCMC for HB estimation, a method of construction of the proposal distribution such that benchmarks are exactly satisfied by the candidate sample realizations. Computation of MSE of optimal estimates under both random and nonrandom benchmarks is discussed. Some illustrative examples are also presented.

Likelihood

Tuesday, June 12th, 10:30 – 11:00, Halpern

Likelihood Inference, High Dimensional Information and Missing at Random

Susan Murphy (University of Michigan)

We consider two problems in which "Missing at Random" is plausible only when one conditions on a large number of auxiliary covariates. Yet in both problems the desired estimand is marginal with respect to the auxiliary covariates. Because the estimand is marginal, demanding that a conditional (on the auxiliary variables) model is correct is undesirable. We propose a method that depends on two guesses, each for different conditional distributions, and if either guess is correct, we show that the estimators are consistent. Furthermore if both modeling guesses are correct the estimator is locally efficient. This type of estimator is called doubly robust by J. Robins. The first problem concerns the estimation of a marginal survival function in the case in which the failure time and the censoring time are independent conditionally on many auxiliary covariates. The second problem is the estimation of the regression coefficient in a proportional hazards model for the failure time distribution given a covariate (e.g. a treatment indicator). However as in the previous problem, the failure time and the censoring time are independent only conditionally on many additional auxiliary covariates.

This is joint work with Donglin Zeng

Tuesday, June 12th, 11:00 – 11:30, Halpern

Likelihood Models for Assessing Familial Aggregation

Nan LAIRD (Harvard School of Public Health)

A first step in exploring the genetic basis of disease is the demonstration that disease is clustered, or aggregated within families. Case-control designs are often used to gather disease outcome data on first degree relatives of cases and controls. We discuss some likelihood models for the analysis of family data from case control studies and compare likelihood-based methods with those based on logistic regression. We discuss the extension of the single disease approaches to study the co-aggregation of more than one disorder.

Tuesday, June 12th, 11:30 – 12:00, Halpern

Adjustments To Profile Likelihood

Nancy REID (University of Toronto)

In many parametric models nuisance parameters are introduced to make the model more realistic, but the main focus of the investigation is inference on a small number of parameters of interest. Often a relatively simple adjustment to the profile likelihood is effective in accommodating these nuisance parameters. The adjustment is typically derived by an argument using higher order asymptotic theory. Some recent work on this topic will be presented, with emphasis on the similarities between Bayesian and non-Bayesian approaches.

Statistics in Industry and Technology

Tuesday, June 12th, 10:30 – 11:00, AQ 4130

Some Simple Research Problems from Industrial Consulting

Jock MACKAY and Stefan STEINER (University of Waterloo)

Routine consulting in an industrial setting can generate many interesting research questions for a perceptive observer. These questions tend to be directly related to application or at most one step removed, and hence are usually simple to state. It is our belief that these problems are best discovered by "being there" since they often arise separately from the purpose of the immediate consultation. In this talk, we present several examples of such problems in an attempt to convince academic researchers of the value of seeing industrial processes and their control systems in action. Reality is often stranger than the best imagination.

Tuesday, June 12th, 11:00 – 11:30, AQ 4130

Statistical Investigation of Resistance Spot Welding Quality Using a Two-Stage, Sliding Level Experiment

Wei LI, Shaowei CHENG, Jack HU and Justin SHRIVER (University of Washington)

Quality variation in resistance spot welding is a major concern in industry. However, the relationship between weld quality and various process conditions, including abnormal process conditions, has not yet been systematically studied. This talk discusses this relationship using a newly developed two-stage, sliding-level experiment. In the experiment, welding current is treated as a "slid factor" and is determined based on the settings of other process variables. Engineering knowledge is applied in statistical model selection. From the analysis, it is found that abnormal process conditions, such as axial misalignment, angular misalignment, poor fitup, edge weld, and electrode wear, significantly affect weld size and thus cause large variations in weld quality. Although they may help increase the weld size, abnormal process conditions generally lead to less robustness of the process. In order to minimize the effects of the abnormal process conditions, a robust parameter design is formulated using the statistical models developed from the experimental data. The analysis suggests that high current and large electrodes should be used in the process. Developed for this study, the new experimental design and analysis procedures are generic for manufacturing processes where process variables are interdependent.

Tuesday, June 12th, 11:30 – 12:00, AQ 4130

Non-Regular Robust Parameter Design

Derek BINGHAM and William LI (University of Michigan)

Robust parameter designs performed using a combined experimental plan for the noise and control factors often provides cost-savings and increased estimation capacity for the effects of interest when compared to Taguchi's crossed arrays. Selecting the "best" fractional factorial design in this setting has received much consideration in recent years. In this work, we introduce methodology for finding optimal non-regular robust parameter designs and demonstrate that these designs often have better estimation capacity than fractional factorial robust parameter designs.

Contributed Session: TimeseriesTuesday, June 12th, 10:30 – 10:45, AQ 4140**Generalized Signed Rank Autoregression**

Jeff TERPSTRA (North Dakota State University)

Signed-rank estimates are known to achieve a good compromise between robustness and efficiency at normal errors. However, signed-rank estimates are only robust for the response, and are quite sensitive to outliers in the design. Because observations in an autoregressive time series play a dual role as both a response and explanatory variable, a signed-rank-based analysis of autoregressive models may not be totally robust. In this paper, a weighted signed-rank estimate which reduces to the Wilcoxon signed-rank estimate when the weights are constant is considered. The generalization is motivated from an algebraic identity for the Wilcoxon signed-rank norm and is similar to the relationship between the GR-estimate and the Wilcoxon R-estimate of Hettmansperger and McKean (1998). The use of weights induces a large class of estimates which range from very robust to highly efficient. Under a finite second moment assumption, asymptotic linearity properties are derived for these so called GSR-estimates. Furthermore, assuming a symmetric error distribution, the GSR-estimates are shown to be asymptotically normal with a square root of n rate. The theory of U-Statistics along with a characterization of weak dependence that is inherent in stationary AR(p) models are the primary tools used to obtain the results. Tests of hypotheses as well as standard errors for confidence interval procedures can be based on the asymptotic theory. The GSR and GR-estimates are compared via an example and efficiency study. The results indicate that both the GSR and GR-estimates are robust for autoregressive time series estimation. However, for an appropriately chosen set of weights the GSR-estimate is more efficient than the GR-estimate. Lastly, the GSR-estimate has an added advantage in that an intercept term can be estimated simultaneously; unlike the GR-estimate where the intercept term is based on a location estimate taken from the initial set of residuals.

Tuesday, June 12th, 10:45 – 11:00, AQ 4140**The Exact Maximum Likelihood Estimate For Stationary Normal AR(1) Processes and Its Applications in Unit Root Test**

Ying ZHANG and A.I. McLEOD

The closed form solution for the exact maximum likelihood estimate for stationary normal AR(1) processes is given. Its asymptotic moments are derived by power notion. Consequently, we develop an algorithm for the high order asymptotic moments of the exact maximum likelihood estimate. Using the exact maximum likelihood estimate, we propose one-tail unit root tests by the technique of mis-specified model. We establish their consistency. We also present a Monte Carlo power study comparing these tests with existing alternative methods. The power of our unit root tests slightly outperforms that of previous proposals.

Tuesday, June 12th, 11:00 – 11:15, AQ 4140**Applying Seasonal Adjustment Methods to Difficult Time Series: Croatian Case**

Ante ROZGA (University of Split)

In recent years we have been witnessing strong development of seasonal adjustment methods. This is particularly due to demand for higher data quality by economic analysts, and also seeking for more theoretically sound methods. Model-based methods such as TRAMO/SEATS are confronted to traditionally used empirical methods like X-11 family. We have tested behavior of TRAMO/SEATS and X-12-ARIMA on 470 Croatian economic time series and have found substantial difference. Here we concentrated our analysis on difficult time series. Croatian economic time series are more "difficult" compared with EU or US for obvious reasons of war and transition process. We used automatic procedure with DEMETRA interface developed by EUROSTAT, Luxembourg. There were 77.45% of all series found difficult by X-12-ARIMA compared with 46.17% by TRAMO/SEATS. The reason for so high percentage of difficult time series is in high level of their desegregation. If we used more aggregated time series this percentage is much lower. This is due to erratic behavior of economically insignificant time series. We also investigated reasons for the series to be regarded as difficult. With user intervention the number of difficult time series could be reduced, particularly when using X-12-ARIMA. More generally, TRAMO/SEATS needs less user intervention and is more convenient when modeling many time series. To avoid difficulties series could be divided in two parts, which could be modeled separately. In Croatian case it is the period which is including the war and the other period after the war. This approach has resulted in better performance of both approaches. Altogether, we found TRAMO/SEATS more flexible to majority of Croatian economic time series.

Tuesday, June 12th, 11:15 – 11:30, AQ 4140

Financial Ups and Downs: Some properties of the Extremes of Stochastic Processes

Don MCLEISH (University of Waterloo)

Observations on security prices, currency exchange rates, interest rates, and other financial time series usually include not only an open and close, but also a high and low price for the period. Suppose we observe stochastic processes $X(t)$ over the interval $[0, T]$. We record the open $X(0)$, the close $X(T)$, the high $H = \max \{X(t); t \text{ in } [0, T]\}$ and the low $L = \min \{X(t); t \text{ in } [0, T]\}$. Properties of the joint distribution of highs, lows and close conditional on the open are well-known in certain diffusion processes, for example when $X(t)$ is a Brownian motion. We consider some simple properties of the joint distribution of $(X(T), H, L)$ and their application to simulation, the valuation of derivatives, estimation of parameters and survivorship bias.

Tuesday, June 12th, 11:30 – 11:45, AQ 4140

Detecting Duration Clustering and Adequacy of Autoregressive Conditional Duration Models Using Wavelets

Pierre DUCHESNE and Yongmiao HONG (Ecole des Hautes Etudes Commerciales)

There has been an increasing interest recently in the analysis of financial data that arrives at irregular intervals. An important class of models is the Autoregressive Conditional Duration (ACD) model introduced by Engle and Russell (1998), and its various generalizations. These models have been used to model duration clustering for financial data such as the arrival times of trades and price changes. However, very few evaluation procedures for adequacy of ACD models are available in the literature. A commonly used diagnostic test is the Box-Pierce/Ljung-Box statistic adapted to the estimated standardized duration residuals, but its asymptotic distribution is unknown and perhaps has to be adjusted. In this talk we propose a new test for duration clustering and a diagnostic test for ACD models using a wavelet spectral density estimator of the duration. The first test exploits the one-sided nature of duration clustering. An ACD process is positively autocorrelated at all lags, resulting in a spectral mode at frequency zero. In particular, it has a spectral peak at frequency zero when duration clustering is persistent or when duration clustering is small at each individual lag but carries over a long distributional lag. As a joint time-frequency decomposition method, wavelets can effectively capture spectral peaks and thus are expected to be powerful. Our second test checks the adequacy of an ACD model by using a wavelet spectral density of the standardized estimated duration residuals over the whole frequency. The proposed diagnostic test has a convenient asymptotic "nuisance parameter-free" property -- parameter estimation uncertainty has no impact on the asymptotic distribution of the test statistic. Moreover, it can check a wide range of alternatives and is particularly powerful when the spectrum of the estimated duration residuals is non-smooth, as can arise due to neglected persistent duration clustering, seasonality, calendar effect and business cycles. We propose and justify suitable data-driven methods to choose the finest scales ---the smoothing parameter in wavelet estimation. This makes the proposed methods fully operational in practice. We present a simulation study, illustrating the merits of the proposed procedures.

Session in Honour of J.N.K. Rao's Retirement

Tuesday, June 12th, 1:30 – 2:00, Images

Jon Rao: an Appreciation of His Research Contributions

David BELLHOUSE (University of Western Ontario)

Jon Rao published his first research paper in sampling at the age of twenty, four years before he received his Ph.D. in 1961. Very quickly Jon became one of the leading researchers in survey sampling and has remained one of the field's leaders to this day. Over this forty-year time period, research in sampling has changed dramatically and Jon has often led the way. Jon's research publications are reviewed in relation to the evolution of modern sampling theory and methods.

Tuesday, June 12th, 2:00 – 2:30, Images

Resampling Methods in Complex Surveys: An Overview

Randy SITTER and Jeff WU (Simon Fraser University)

This presentation will attempt to review some of the developments in resampling methods in the context of sample surveys. This will include the jackknife, balanced repeated replications and the bootstrap, and will focus primarily on variance estimation. Complex sampling structures and issues such as stratification, multistage cluster sampling, without replacement sampling, missing data and imputation etc., will be considered and the resulting impact these have on the structure and complexity of resampling methods will be discussed. We will finish with discussion of some recent developments, fundamental limitations and challenges for the future.

Tuesday, June 12th, 2:30 – 3:00, Images**Contributions To Survey Methodology At Statistics Canada By J.N.K. Rao**

Mike HIDIROGLOU (Statistics Canada)

J.N.K. Rao has been a consultant in survey sampling at Statistics Canada since 1973. The paper focuses on Jons direct and indirect contributions to both the research and applied methodology program at Statistics Canada. His most direct contributions include the various articles that he co-authored with a number of Statistics Canada methodologists, his long and active participation on the Advisory Committee on Statistical Methods at Statistics Canada, his work as associated editor of Survey Methodology, and the application of his theoretical developments in survey sampling in several of Statistics Canada's surveys. His indirect contributions include the mere fact that Statistics Canada was fortunate to have someone of Jons stature as a consultant. He has provided valuable statistical advice to many methodologists at Statistics Canada, and a wide range of survey problems. His continuous encouragement of research, and his interactions with the more oriented research methodologists has clearly been one of the contributing factors to Statistics Canada leadership in survey methodology. Last but not least are Jons genuine care for people and his love for survey method.

Bioinformatics Tutorial ITuesday, June 12th, 1:30 – 3:00, Halpern**Inference In High-Throughput Molecular Biology**

Larry HUNTER (University of Colorado)

Biomedicine is entering a new era of high-throughput data production. The macromolecular sequence databases are doubling in size every 18 months or so, and now contains more than 7 million sequences representing more than 9 billion nucleotides. Gene chips allow for the simultaneous assaying of the expression levels of thousands of genes at a time. Related technologies allow the identification of millions of point mutations in particular individuals, the simultaneous screening of tens of thousands of compounds for drug-like binding affinity, and the relatively rapid determination of three-dimensional macromolecular structures. The challenges inherent in analyzing this onslaught of extraordinarily interesting data are defining the new field of bioinformatics. Making biologically relevant inferences from massive data sets is a key aspect of the field, drawing on techniques from statistics, pattern recognition, and machine learning. In this overview, several representative problems and solution techniques will be presented for a statistical audience. Techniques illustrated include hidden Markov models, support vector machines and Bayesian networks.

Discrete Probability and CombinatoricsTuesday, June 12th, 1:30 – 2:15, AQ 4150**Exponential Decay And Mixing Properties For Lattice System in Finite Volumes**

Ken ALEXANDER (University of Southern California)

For spin systems (such as the Ising or Potts models) and dependent percolation models (such as the FK model), we examine some questions related to the degree to which the influence of an event occurring in one region inside a finite volume can propagate so as to alter the probabilities of events occurring in distant regions inside the volume. These questions include the following: if the model exhibits exponential decay of correlations or connectivities in infinite volume, is there an analogous exponential decay in finite volumes which is uniform in the choice of volume and boundary condition? How is such uniform exponential decay of the two-point function (connectivity or correlation) related to uniform versions of mixing properties? Typically mixing properties for lattice models (weak and strong mixing) are formulated in terms of the maximum influence of an event occurring in some region Γ on another event occurring in some distant region Δ , with Γ and Δ deterministic. To what extent do such mixing properties hold when Γ and Δ are allowed to be random?

Tuesday, June 12th, 2:15 – 3:00, AQ 4150**Critical Two-Point Functions For Self-Avoiding Walk, Percolation, Lattice Trees And Lattice Animals**

Gordon SLADE (University of British Columbia)

It is well-known that the random walk Green function decays like $|x|^{-d}$ in dimensions $d > 2$. This talk discusses a recent proof of this same power law decay for the critical two-point functions of self-avoiding walk, percolation, and lattice trees and lattice animals, for sufficiently spread-out models in dimensions $d > 4$ (self-avoiding walk), $d > 6$ (percolation) and $d > 8$ (trees and animals). The proof uses a new unified approach to the lace expansion, and is joint work with Takashi Hara and Remco van der Hofstad.

IMS Special Invited Session IITuesday, June 12th, 3:30 – 5:00, Images**Interpreting Genomic Polymorphism Data: What History Has To Tell Simon**

TAVARE (University of Southern California)

Technological advances in molecular biology have made it possible to survey genome-wide DNA sequence variation in natural populations. These data include restriction fragment length polymorphisms, micro satellite repeats, single nucleotide polymorphisms (SNPs) and complete DNA sequences of particular loci. The analysis and interpretation of the patterns of variation seen in such data is complicated by the fact that the sampled chromosomes share a common ancestry, thus making the data highly dependent. To make matters worse, the nature of this common ancestry is not known precisely and therefore needs to be modeled. Since the early 80s, population geneticists have used the coalescent as a stochastic description of the ancestry of a sample of chromosomes, and there is now an extensive literature on inference and estimation for such processes. In this talk I will give an overview of coalescent methods, touching on a number of applications including inference about demographic history, selective sweeps, the age of mutations, and the hunt for disease genes using linkage disequilibrium mapping. A number of open problems will be discussed.

Bioinformatics Tutorial IITuesday, June 12th, 3:30 – 5:00, Halpern**Predicting Metabolic Networks**

Imran SHAH (University of Colorado)

High-throughput molecular biology is yielding an unprecedented amount of data that bears indirectly on our understanding of biological entities and processes. This data can aid in unraveling the intertwined network of biochemical interactions known as metabolism. Understanding the metabolism of living organisms will have a broad impact on research in the life sciences. Bioinformatics approaches aimed at tackling this problem face a number of computational challenges. This talk will describe a bioinformatics strategy for predicting metabolic networks highlighting the important statistical and computational challenges of the approach.

Canadian Initiative on Social StatisticsTuesday, June 12th, 3:30 – 5:00, AQ 4150**Perspectives on the Canadian Initiative on Social Statistics**

Doug NORRIS (Statistics Canada), Paul BERNARD (Université de Montréal), Mary THOMPSON (University of Waterloo)

The Canadian Initiative on Social Statistics (CISS) is a cooperative project of SSHRC and Statistics Canada. Its aim is to promote research and training in the application of social statistics. Once CISS has implemented the recommendations of a joint National Task Force of leading Canadian researchers and statisticians, it is anticipated that there will be a great increase in Canada's capacity to undertake major, innovative, social research based on quantitative evidence. The purpose of this SSC session is to give various perspectives on the kinds of research the CISS will help to bring about. The speakers will tackle some provocative topics such as (i) what social statistics are and are not; and (ii) what the cooperation should be between statisticians and social scientists involved in social statistics research.

SSC Presidential Invited Address

Wednesday, June 13th, 8:30 – 10:00, Images

The Geometry Of Random Images In Astrophysics And Brain Mapping

Keith WORLSEY (McGill University).

The geometry in the title is not the geometry of lines and angles but the geometry of topology, shape and knots. For example, galaxies are not distributed randomly in the universe, but they tend to form clusters, or sometimes strings, or even sheets of high galaxy density. How can this be handled statistically? The Euler characteristic (EC) of the set of high density regions has been used to measure the topology of such shapes; it counts the number of connected components of the set, minus the number of 'holes,' plus the number of 'hollows'. Despite its complex definition, the exact expectation of the EC can be found for some simple models, so that observed EC can be compared with expected EC to check the model. A similar problem arises in human brain mapping, where the EC is used to detect local increases in blood flow (positron emission tomography) or oxygenation (functional magnetic resonance imaging) due to an external stimulus. This allows researchers to find the regions of the brain that are 'activated' by the stimulus. Recent work has extended these ideas to manifolds so that we can detect changes in brain shape via structure masking, surface extraction, and 3D deformation fields. Finally, in organic physical chemistry, vortices in excitable media are low density sets that form random closed loops. These are stable if they are linked or knotted, so the question is whether this can actually happen.

Statistics in Fisheries I

Wednesday, June 13th, 10:30 – 11:00, Images

Defining Conservation Units Based On Capacity For Local Adaptation: Testing Predictions Of An Heuristic Mode

Chris WOOD (Fisheries and Oceans Canada)

Conservation biologists have emphasized the protection of rare and endangered species to maintain biodiversity, ecological complexity, and the capacity for evolution. This emphasis is reasonable because species typically have evolved over many thousands or millions of years and now exist as closed, self-perpetuating pools of genetic diversity that once extinct, cannot be replaced within human timescales. However, species are abstract entities resulting from the evolutionary processes of mutation, random genetic drift, migration and natural selection that operate within populations. Genetic diversity must also be conserved within species by identifying and prioritizing population units that warrant special protection or separate management. I suggest that our goal should be to conserve capacity for local adaptation and I review the theoretical conditions under which local adaptation could be expected to occur. I also present an heuristic model for defining conservation units based on genetic survey and census data. Although this simple equilibrium model includes several unrealistic assumptions, its predictions are shown to be qualitatively reasonable when tested in well-studied populations of sockeye salmon.

Wednesday, June 13th, 11:00 – 11:30, Images

A Hierarchical Statistical Model To Link Spatially Explicit Salmon Abundance Records To Habitat Characteristics In The Salmon, Snohomish, And Willamette River Basins

Ashley STEEL Blake E. FEIST, George R. PESS, and Robert E. BILBY (U.S. National Marine Fisheries Service)

Within the Willamette, Snohomish, and Salmon River basins, a small proportion of stream reaches consistently receive the majority of salmon spawners despite annual fluctuations in total population size. We used the annual consistency in fish distribution to identify relationships between coarse-scale habitat characteristics and spawner abundance. We linked redd counts to multiple layers of habitat data available at broad spatial scales (geology, mean annual air temperature, road density, and forest cover); habitat layers were chosen for analysis based on expected species-specific relationships between in-stream habitat characteristics and abundance at multiple life stages. To test for statistical significance of observed patterns, we used a hierarchical linear model (HLM) in combination with a randomization test. Our method identified the suite of habitat characteristics associated with the most productive stream reaches and predicted locations within the basin that might have in-stream habitat characteristics suitable for supporting high numbers of fish.

Wednesday, June 13th, 11:30 – 12:00, Images

Bayesian Methods For Inference In Populations Of Hybridizing Fish

Eric ANDERSON (University of Washington)

A pervasive concern in the conservation of salmonids is hybridization between different populations and species. For example, small, wild populations might encounter substantial gene flow from fish straying from hatcheries. Also, the use of genetic data in helping to delineate "evolutionarily significant units" may be complicated by naturally-occurring hybridization between closely related species. However, ongoing hybridization between genetically dissimilar populations leads to predictable changes in the frequency of multilocus genotypes, and so may be investigated using genetic marker data. I will review recent advances in Bayesian methods for analyzing genetic data from populations containing hybrids. I illustrate these versatile methods by using them to estimate allele frequencies and to identify coastal cutthroat trout, steelhead trout, and their hybrids in samples of juvenile fish taken from Oregon and Washington

Nonparametric Econometrics

Wednesday, June 13th, 10:30 – 11:15, AQ 4130

Nonparametric Regression Estimation Using Weak Separability

Joris PINKSE (University of British Columbia)

In this paper I propose a nonparametric kernel estimator of weakly separable regression functions for independent and identically distributed (i.i.d.) data. Weak separability nests a number of other separability conditions, including the additively separable model used by among others Linton and Nielsen (1995) and the generalized additive model of Linton and Hardle (1996) and Horowitz (1999). Like other papers my method uses marginal integration. A desirable and unusual feature of my estimation method is that the integration region can depend on the value of the sub--vector that is not being integrated and can be infinite. Additionally, support of the joint density of regressors can be any open, possibly infinite, convex set. Finally, in most cases the estimator can be computed without numerical integration and computation is hence straightforward.

Wednesday, June 13th, 11:15 – 12:00, AQ 4130

Economic Development and the Return to Human Capital

Thanasis STENGOS (University of Guelph)

The paper looks at the impact of human capital accumulation on the growth of output by allowing the contribution of traditional inputs (capital and labor) as well as that of human capital to vary both across countries and time. The former is accomplished by constructing an index of (Total Factor Productivity) TFP growth for traditional inputs, while human capital effects are estimated by semiparametric local polynomial regression methods using kernels. We find that the output elasticity of human capital is on average 0.25 for high income countries and is lower and insignificant for low- and middle-income economies. We also present social rates of return to human capital that are based on aggregate data. For high income countries these are consistent with estimates from micro studies.

Contributed Session: Longitudinal and Correlated Data Analysis

Wednesday, June 13th, 10:30 – 10:45, AQ 4140

Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates

Ying MACNAB (University of British Columbia)

Within the context of disease mapping and surveillance, there have been a few recent literature on spatio-temporal modeling of disease rate; those publications concern mainly with methodological developments to accommodate temporal effects of various forms. In this presentation I develop generalized additive mixed spline (GAMS) models for an additive modeling of spatial effects, temporal effects and age effects, presenting a spline-based methodology for flexible fitting of nonlinear fixed and random temporal effects while permitting spatial dependence. I propose the use of B-splines, in particular, the use of random B-splines for temporal smoothing and show that the GAMS inference can be implemented using existing methods that are within the generalized linear mixed model (GLMM) framework. I describe an analytic strategy that addresses a variety of issues, including random splines and model invariance, knot selection, computation, interpretation of temporal trends of moderate nonlinearity and production of smoothed maps. The methodology is illustrated by two real-life examples, one within the context of disease mapping and the other of spatio-temporal disease surveillance.

Wednesday, June 13th, 10:45 – 11:00, AQ 4140

Inverse Probability Weighted Second Order GEE for Clustered Data with Intermittently Missing Observations

Grace Yun YI and Richard J. COOK (University of Waterloo)

Generalized estimating equations have been widely used for the analysis of clustered data from longitudinal studies for estimating parameters in regression models for the marginal means. Inverse probability weighted generalized estimating equations (Robins, Rotnitzky, and Zhao 1995) have been developed to deal with incomplete data which arise from a missing at random (MAR) mechanism. In many cases, however, the repeated measurements themselves may arise in clusters, which leads to both a cross-sectional and a longitudinal correlation structure. In some applications the correlation structure may become of interest itself. We describe inferential procedures which account for intermittently missing at random data based on weighted second order generalized estimating equations for the coefficients in regression models for the marginal means, as well as association parameters. The methods are applied to data from the Waterloo Smoking Prevention Project in which the binary responses indicate smoking status.

Wednesday, June 13th, 11:00 – 11:15, AQ 4140

Maximizing Spatial Likelihood And Its Applications

Hongtu ZHU, Minggao GU, Julie ZHOU, Fanliang HE (University of Victoria)

We propose a two-stage algorithm for computing maximum likelihood estimates for a class of spatial models. The algorithm combines Markov chain Monte-Carlo methods such as the Metropolis-Hastings-Green algorithm and the Gibbs sampler, and stochastic approximation methods such as the off-line average (Polyak and Juditsky, 1992) and adaptive search direction (Gu and Kong, 1998). A new criterion is built into the algorithm so stopping is automatic once the desired precision is set. Simulation studies have been conducted for two spatial models. Applications are provided to model the distribution of two vegetation types in British Columbia, Canada.

Wednesday, June 13th, 11:15 – 11:30, AQ 4140

Analysis of Repeated Measures Data with Clumping at Zero

Janet TOOZE, Gary K. GRUNWALD, Richard H. JONES (National Cancer Institute)

Longitudinal or repeated measures data with clumping at zero occur in many applications in biometrics, including health policy research, epidemiology, nutrition, and meteorology. These data exhibit correlation because they are measured on the same subject over time or because subjects may be considered repeated measures within a larger unit such as a family. They present special challenges because of the extreme non-normality of the distributions involved. A model for repeated measures data with clumping at zero, using a mixed-effects mixture model with correlated random effects, is presented. The model contains components to model the probability of a non-zero value and the mean of non-zero values, allowing for repeated measurements using random effects and allowing for correlation between the two components. Methods for describing the effect of predictor variables on the probability of non-zero values, on the mean of non-zero values, and on the overall mean amount are given. This interpretation also applies to the mixture model for cross-sectional data. The proposed methods are illustrated with analyses of effects of several covariates on medical expenditures in 1996 for subjects clustered within households using data from the Medical Expenditure Panel Survey.

Wednesday, June 13th, 11:30 – 11:45, AQ 4140

Simplex Mixed Models For Longitudinal Proportional Data

Zhenguo QIU, Peter X.-K. SONG and Ming TAN (York University)

We will present a generalized linear mixed model (GLMM) based on the simplex distribution of Barndorff-Nielsen and Jorgensen (1991), suitable for modeling longitudinal continuous proportional data. The parameter estimation for regression coefficients is carried out by the penalized quasi-likelihood (PQL) proposed by Breslow and Clayton (1993). The REML (Restricted maximum likelihood) method is then used to obtain the estimates of variance parameters for random effects. The result of simulation study is given to investigate the estimates of variance parameters chosen at several different levels of dispersion parameters. The proposed method is illustrated by analyzing an eye surgery data.

Wednesday, June 13th, 11:45 – 12:00, AQ 4140

State Space Mixed Models for Binomial Longitudinal Observations

Claudia CZADO (Munich University of Technology) and Peter X.-K. Song (York University)

We propose a new class of state space models for longitudinal discrete response data, where the observation equation is specified in an additive form involving both deterministic and dynamic components. These models allow us to explicitly address the effects of trend, seasonal or other time-varying covariates while preserving the power of state space models in modeling dynamic pattern of data. We develop different MCMC algorithms and compare their mixing and convergence properties in a simulation experiment. Finally, the applicability of the proposed models is illustrated for the Tokyo rainfall data (Kitagawa, 1987)

Contributed Session: Inference

Wednesday, June 13th, 10:30 – 10:45, AQ 4120

Sequentail Testing Of Composite Hypotheses

Edit GOMBAY (University of Alberta)

We briefly review sequential tests that are used in applied statistics. We analyze one type that is likelihood based, in the presence of nuisance parameters. Some new procedures are suggested that have optimal power properties.

Wednesday, June 13th, 10:45 – 11:00, AQ 4120

Consistent Estimation Of Distributions With Type II Bias

Hari MUKERJEE and H. El BARMi (Wichita State University)

A random variable X is said to have a Type II positive bias if $P(X \text{ in } [0, x])$ is greater than or equal to $P(X \text{ in } [-x, 0])$ for all x greater than or equal to 0. It is known that the nonparametric maximum likelihood estimator of the distribution function of X is inconsistent when the distribution is continuous. We provide a projection type estimator in this case that is strongly uniformly consistent. We derive its weak convergence properties and provide asymptotic inference procedures. It is shown that the same mathematical technique can be used to estimate two cumulative incidence functions in a competing risks problem when their cause specific hazard rates are ordered. All results have also been extended to the censoring case.

Wednesday, June 13th, 11:00 – 11:15, AQ 4120

On Semiparametric Estimates of Population Density in Line Transect Sampling

R. KARUNAMUNI and S. ZHANG (University of Alberta)

This paper investigates some potentially useful and possibly relevant semiparametric methods in line transect sampling. We obtain expressions of semiparametric Bayesian and non-Bayesian estimates for the density of biological populations. The proposed estimates are compared with the traditional kernel and maximum likelihood counterparts. Some examples of line transect sampling surveys are also analyzed.

Wednesday, June 13th, 11:15 – 11:30, AQ 4120

Prior Density Estimation via Haar Deconvolution

Alexandre LEBLANC and Jean-François Angers (Université de Montréal)

The problem of prior elicitation often arises when one is interested in doing inference in a Bayesian setting. Common solutions to this problem consist of hierarchical modeling, noninformative priors, different empirical Bayes techniques, etc. An alternative to these solutions might be to estimate the unknown prior directly from the available observations. In the case of the prior distribution of a location parameter, we approach the problem by deriving a nonparametric estimate of the unknown density. The proposed technique, called Haar deconvolution, is similar to the well-known Fourier deconvolution which can be used to obtain estimates of mixing densities. The resulting estimated density is a piecewise constant function which is easy to use to perform both numerical integration and statistical simulation.

Wednesday, June 13th, 11:30 – 11:45, AQ 4120

A Statistical Test for Comparing Poisson, Negative Binomial, and Binomial Populations

Humberto VAQUERA, Jose VILLASEÑOR and B.C. ARNOLD (ITESM-CP)

In some research areas as Biology, Insurance, etc., is frequent the use of discrete variables. The most used discrete probability models are Poisson, Binomial, and Negative Binomial. In this study a statistical procedure is proposed for testing the null hypothesis H_0 : Poisson against H_a : Negative Binomial or Binomial. The test is based on the Fisher's dispersion index. The distribution of the proposed test statistic is obtained asymptotically. For small samples we use computer simulation for investigating the distribution of the test statistic under the null hypothesis. A power study is also performed for comparing the proposed test with another methods. Two data sets are used to illustrate the method.

Wednesday, June 13th, 11:45 – 12:00, AQ 4120

Testing The Number Of Components Of Finite Mixture Models With Modified Likelihood

Jiahua CHEN, Hanfeng CHEN, Jack KALBFLEISCH (University of Waterloo)

Finite mixture models belong to a class of non-regular probability models. Standard statistical procedures are often not suitable. It is particularly difficult to test the number of components of the finite mixture models. In this presentation, we discuss the utility of using the modified likelihood in constructing hypothesis test of the mixture of two against the possibility of three or more. We show that the proposed test statistic has relatively simple limiting distribution. Unlike our previous results in testing homogeneity, however, the null limiting distribution is dependent on the true distribution. We find that the null limiting distribution can be determined by replacing unknown parameters by their estimates. Consequently, the test can be done at the asymptotically correct significance level. Our preliminary simulation results reveal that the asymptotic results are very useful when sample size is at a level of 500.

WNAR Student Presentations

Wednesday, June 13th, 10:30 – 11:00, AQ 4150

Mammographic Computer-Aided Diagnosis Using Bootstrap Aggregation

Mireya DIAZ and J. Sunil RAO (Case Western Reserve University)

Mammography provides an effective tool for the early detection of possibly irregular breast masses. Computer-aided diagnosis (CAD) using digital mammograms can be a very effective complement to a visual diagnosis. One approach classifies each pixel indirectly into mass or non-mass via a one-split classification stump derived from the pixel intensity profile from a point of origin. A bootstrap ensemble method (bagging) combining these simple circle approximations can be an effective method for detecting more complex shapes. Two data resampling schemes are implemented: resampling broken line residuals and resampling candidate pixel centers. Both approaches were tested on simulated "masses" of circular, and elliptical shapes under different noise conditions. Results follow the theory on bagging with bootstrapping residuals having a beneficial effect for generally circular masses, and bootstrapping centers providing large decreases in misclassification for more general shapes. The techniques will be detailed and illustrated. In addition, extensions to tracking temporal shape profiles discussed.

Wednesday, June 13th, 11:00 – 11:30, AQ 4150

Perfect Simulation of Stochastic Differential Equations

Xiaoqiang LI and Duncan MURDOCH (University of Western Ontario)

Propp and Wilson (1996, Random Structures & Algorithms) proposed an algorithm called coupling from the past (CFTP). Their approach permits finite simulation to be used to produce an exact sample from the limiting distribution of a Markov chain. We describe our work applying the underlying ideas of CFTP to the simulation of stochastic differential equations (SDEs), e.g. Sdx

Wednesday, June 13th, 11:30 – 12:00, AQ 4150

Responder Cell Frequency Estimation and Binomial Three-level Nonlinear Mixed Effects Model in Limiting Dilution Assays

Misoo ELLISON (University of Colorado Health Sciences Center/National Jewish Medical and Research Center)

Responder cell frequencies (RCF), which describe boosted immune responses in Herpes Zoster (HZ) prevention studies, are estimated by using Limiting Dilution Assays (LDA). The theoretical linearity assumption, between the logarithm of the proportion of nonresponding wells (S) and the cell concentration (N) (or dilution level) in LDA, based on the single-hit Poisson model is often violated with observed data resulting in biased estimates of RCF. This Poisson assumption will be modified by applying a mixture of Poisson and Gamma distributions, resulting in a Negative Binomial assumption, which presents a better fit between S and N. In LDA for HZ prevention studies, the responses (nonresponding wells) are repeatedly measured at different cell concentrations and over time. I propose a model, where the response variable is binomial and the probability (S) has a nonlinear functional relationship with covariates, cell concentrations and times. The nonlinear functional relation, between S, N, and time, is based on the negative binomial assumption. This binomial three-level nonlinear mixed effects model accounts for correlation between responses in dilution levels from the same individual and the correlation between repeated assays of an individual over time, simultaneously.

Statistics in Forestry

Wednesday, June 13th, 1:30 – 2:00, Halpern

Bayesian Statistics To Make Complex Decisions In Natural Resources

Gay BRADSHAW and Jeffrey BORCHERS (USDA Forest Service)

Recognition of social and ecological complexity involved in conservation plans have encouraged the development of collaborative efforts between scientists, the public, and land agencies. However, because these three sectors are governed by different constraints and cultures, communication and achieving success collaboration is often difficult. For instance, perceptions of risk and levels of confidence in scientific uncertainty are generally different for scientists and agency specialists than for the public. To facilitate in the collaboration process, a coupled Bayesian and objectives hierarchy model has been developed to represent values, objectives, and scientific information used in the decision making process for conservation plan development. The model and process was developed for the Little Applegate watershed in an Adaptive Management Area, southern Oregon, USA. The integration of modeling and the collaborative process has increased trust among the three sets of collaborators because it makes the decision making process transparent. It also has been used to develop a monitoring strategy and was used to assess the feasibility of projected land management activities given infrastructure constraints.

Wednesday, June 13th, 2:00 – 2:30, Halpern

The Size Distribution Of Wildfires

William REED and Kevin S. MCKELVEY (University of Victoria)

Examination of extensive records on areas burned by wildfires in six regions in North America reveal that there are some commonalities in the size distributions across the regions. However a simple power-law distribution for size, as has been suggested by several authors, is too simple to describe the size-distribution over its full range. A simple stochastic model for the spread and extinguishment of fires is used to examine conditions for power-law behavior and departures from it. Some alternative parametric models for the size distribution are considered with a 'competing hazards' model providing the best fit.

Wednesday, June 13th, 2:30 – 3:00, Halpern

Silviculture Survey System: Statistical Properties and Anticipated Yields

Wendy BERGERUD (B.C. Ministry of Forests)

The BC Ministry of Forests has a silviculture survey system designed to assess whether recently established stands have reached "free-growing". Forest companies are required to establish free-growing stands on areas they have harvested. When they have done this the responsibility for managing the stand moves from the forest company to the Crown. While the current silviculture survey can be viewed as a simple random sample, the methodology includes some unusual parameters that don't allow the use of traditional statistical formulae. The acceptance sampling framework and simulation methods were used to examine how changing the parameters of the survey effect the chances of mis-identifying stands that were are not actually free-growing. Further, since the long-term goal of the survey system is to identify stands that will not provide sufficient yield at rotation, the Ministry's tree-growth model TASS was used to project the expected yields. Preliminary results from this research will be presented.

Statistics in Fisheries II

Wednesday, June 13th, 1:30 – 2:00, Images

Ideal Survey Designs for Various Spatial Distributions of Benthic Invertebrate Populations

Zane Z. ZHANG (Department Fisheries & Oceans, Nanaimo)

Populations of sedentary benthic invertebrates, such as sea urchins, are often surveyed along transects laid perpendicular to the shore. Transects may be placed in a systematic, random, or stratified random manner. Along each transect, quadrats are placed either consecutively, systematically, randomly or stratified-randomly. Animals within each quadrat are counted and/or sampled to estimate population parameters. To examine the relative accuracy and precision of these various survey designs, a computer simulation model has been built, which evaluates various sampling strategies at differing degrees of animal aggregations. It allows us to search for ideal designs for various animal spatial distributions.

Wednesday, June 13th, 2:00 – 2:30, Images

Adaptive Cluster Sampling of Gulf of Alaska Rockfish

Terry QUINN (University of Alaska, Juneau)

Adaptive cluster sampling (ACS) has stirred up significant interest for sampling clustered populations since Thompson's initial publication in 1990. While many papers have been published about ACS, practical field experiments with fish populations are rare. In 1998 and 1999, two sampling cruises tested adaptive cluster sampling for rockfish, specifically Pacific ocean perch (POP) and shorttraker and rougheye rockfish (SR/RE). POP are thought to be more aggregated than SR/RE. The variable of interest was rockfish catch in a trawl haul of 15 to 30 minutes duration. Adaptive sampling was conducted around each haul that exceeded the fixed criterion value. A stopping rule prevented the sampling from continuing indefinitely. Our working hypotheses were (1) that ACS would outperform simple random sampling (SRS) for POP, and (2) ACS would be more applicable for POP than for SR/RE, because in 1998, the sampling strategy was ACS using order statistics for the criterion value. This approach necessitated the taking of simple random samples of 12-15 stations across 6 strata. In 1999, we used a larger study area and considered four alternatives for choosing a criterion value before the survey took place. The alternative that yielded the lowest criterion value was used in the field, and the higher criterion values were simulated after the survey. Regarding the first hypothesis, the performance of ACS was sometimes better and sometimes worse than SRS with a similar amount of sampling effort. When the study population was moderately aggregated (as judged from variogram analysis), ACS had better absolute precision and time-efficiency but relative precision was sometimes no better or even worse. Curiously, the SRS estimator was often substantially higher than the ACS estimators, even though all estimators are unbiased. This was a result of a large catch per haul in the initial sample that was mediated through adaptive sampling. Regarding the second hypothesis, ACS did prove to be less effective for SR/RE due to its lower aggregation. Simulation showed that carefully-chosen, high criterion values result in substantially less adaptive sampling with little tradeoff in precision. A substantial problem with ACS in the field is that edge units contribute little to the efficiency of an estimator but may take substantial time to carry out. Overall, our ACS results were not as optimistic as those published in the literature but optimistic enough to encourage further application.

Wednesday, June 13th, 2:30 – 3:00, Images

Assessing Fish Population With Capture-Recapture: A How-To-Guide

Carl SCHWARZ (Simon Fraser University)

Capture-recapture methods are a well-known tool in fisheries management. Methods range from the simple Petersen to more complex multiple capture multiple strata models. In this talk, we provide an overview of the many types of capture-recapture models used in practice and review their success and failures.

Innovative Applications of Statistics in BusinessWednesday, June 13th, 1:30 – 2:00, AQ 4150**Early Detection of High Risk Claims at the Workers Compensation Board of British Columbia**

Ernest URBANOVICH, Ella E. YOUNG, Martin L. PUTERMAN, Sidney O. FATTEDAD (Workman's Compensation Board)

The Workers Compensation Board of British Columbia (WCB) annually accepts more than 74,000 short term disability claims for work place injuries and pays out \$900 million dollars in benefits. A small fraction (4%) of these claims, which we refer to as high-risk, require vocational rehabilitation and/or long-term disability payments in addition to initial short-term disability payments. These high-risk claims incur average benefit payments over 40 times greater than the average for claims that do not require these benefits. For the 321,000 claims received between 1989 and 1992, high-risk claims incurred \$1.1 billion (64%) of the total payment of \$1.8 billion for all claims. The purpose of this study was to develop a computer implementable approach for early detection of these costly claims so that more intensive case management may be employed early in the claim process. Our approach combined decision analysis and logistic regression and was applied to data on the 321,000 claims received between 1989 and 1992. Data was stratified by injury type and the risk factors age of worker and number of days of initial payments were found to be highly predictive of high risk status. A decision theoretic approach was used to develop a classification rule that was shown to have high out-of-sample predictive power. The model is now being incorporated in a claims risk scorecard that will be used to determine which claims need early intervention. We estimate annual savings of approximately \$5 million dollars as a result of this targeted early intervention.

Wednesday, June 13th, 2:00 – 2:30, AQ 4150**Analyzing Telecommunications Network Call Traffic Flow at Telus**

Jason GOTO, Isabelle SMITH, Kelly CHUNG and Hong CHEN (Center for Operations Excellence)

Call traffic data detailing the start time, duration and route of individual calls across a telecommunication switching network was analyzed to identify clusters of network sectors with similar time of day demand patterns. The study revealed opportunities to increase network utilization through alternate call routing in adjacent network sectors with complementary time of day demand patterns.

Wednesday, June 13th, 2:30 – 3:00, AQ 4150**Dynamic Linear Models for Motion Pictures Box-Office Forecasting**

Remi DESMEULES (University of British Columbia)

I use a Dynamic Linear Models to predict the gross box-office of movies from week to week, particularly trying to replicate the Top 10 box-office results. I use multiple linear regression to obtain a prior for the parameters of the exponential decay model and then use Dynamic Linear Models to dynamically update the forecast as data is gained on the stream of revenue of the movie. I also use an "attraction model" to fine tune my Top 10 predictions and account for seasonality and competition.

Recent Results on Normal and Related ApproximationsWednesday, June 13th, 1:30 – 2:00, AQ 4140**An Inequality For Tail Probabilities Of Martingales With Bounded Differences**

V. BENTKUS (Vilnius Institute of Mathematics)

We provide precise bounds for tail probabilities $P\{M_n \geq x\}$ of martingales with bounded differences. The bounds are expressed through tail probabilities of sums of i.i.d. Bernoulli random variables. In other words, we show that the tails are sub-Bernoullian. Sub-Bernoullian tails are strictly dominated by Gaussian tails, since their ratio Gaussian/sub-Bernoullian tends to infinity exponentially fast as $x \rightarrow \infty$. Possible extensions of the methods are discussed.

Wednesday, June 13th, 2:00 – 2:30, AQ 4140**Stein's Method For Computing Error Bounds**

Susan HOLMES and Gesine REINERT (Stanford University)

We present new proofs for known results about the convergence in law of the bootstrap distribution to the true distribution of smooth statistics, whether the samples studied come from independent realizations of a random variable or dependent realizations with weak dependence. The techniques employed are based on Stein's method for empirical processes as developed by Gesine Reinert in her Ph D thesis, with whom this is joint work.

Wednesday, June 13th, 2:30 – 3:00, AQ 4140

Asymptotic Expansions For Studentized Statistics Under Weak Dependence

Soumendra N. LAHIRI (Iowa State University)

Let X_1, X_2, \dots be a sequence of stationary strongly mixing random vectors with mean vector μ . Let $\hat{\theta}_n = H(\bar{X}_n)$ be an estimator of a parameter of interest $\theta = H(\mu)$, where H is a smooth function. In this talk, we consider asymptotic expansions for Studentized statistics of the form $T_n = n^{1/2}(\hat{\theta}_n - \theta) / s_n$ for a class of Studentizing factors s_n that estimate an unbounded number of lagged autocovariances of the sequence $\{X_n\}$. Some applications of the expansions to block bootstrap methods are also indicated.

Block Designs: Directions and Inspirations

Wednesday, June 13th, 1:30 – 1:50, AQ 4130

A- And MV-Efficient Designs For Comparing A Set Of Test Treatments To A Set Of Controls

M. JACROUX (Washington State University)

The problem of comparing t test treatments to s controls using experimental units arranged in b blocks of size k is considered. Methods are given for deriving highly A- and MV-efficient designs for such situations. The methods given combine optimality results obtained for one-way classification models with the optimal augmentation of blocks with controls of block designs which are known to be efficient for making comparisons among the test treatments. Examples are given to show how the results obtained can be applied.

Wednesday, June 13th, 1:50 – 2:10, AQ 4130

Design Of Computer Experiments To Optimize The Mean Of A Response Subject To Constraints

W. NOTZ (Ohio State University)

For purposes of this talk, a computer experiment is assumed to consist of computer code that produces deterministic responses for a given set of inputs. We consider the situation in which there are two types of inputs; manufacturing variables and noise variables. Manufacturing variables are those that can be controlled by the product designer. Noise variables cannot be controlled but have values that follow some probability distribution. We are interested in the means of the responses over the distribution of the noise variables, and we seek the values of the manufacturing variables that optimize the mean of one of the responses, subject to constraints on the mean of another response. The approach is Bayesian; in the case of two responses, the prior information is that the responses are draws from a covariance stationary bivariate Gaussian stochastic process with correlation functions belonging to a parametric family with unknown parameters. A latin hypercube design is used to provide an initial fit of this model. Subsequently, a sequential strategy is used to select values of the inputs at which to observe the response. This strategy involves computing, for each unobserved set of values of the inputs, the posterior expected "improvement" over the current best guess at the optimum and the next set of inputs to observe are those that maximize this expected improvement. The strategy is illustrated with examples.

Wednesday, June 13th, 2:10 – 2:30, AQ 4130

Block Design Optimality: Results for Combinatorially Problematic Settings

Brian RECK (Old Dominion University) and J. P. Morgan (Virginia Tech)

Given v treatments to compare, and having available b blocks of k experimental units each, the thoughtful statistician asks, "What is the optimal allocation of the treatments to the units?" This is the basic block design problem. Let n_{ij} be the number of times treatment i is used in block j and let N be the $v \times b$ matrix $N = (n_{ij})$. There is now a considerable body of optimality theory for block design settings where binarity (all n_{ij} in $\{0,1\}$), and symmetry or near-symmetry of the concurrence matrix NN' , are simultaneously achievable. Typically the same classes of designs are found to be best using any of the standard optimality criteria. Among these are the balanced incomplete block designs (BIBDs), many species of two-class partially balanced incomplete block designs, and regular graph designs. However, there are triples (v,b,k) in which binarity precludes near-symmetry. For these combinatorially problematic settings, recent explorations have resulted in new optimality results and insight into the combinatorial issues involved. Of particular interest are the irregular BIBD settings, that is, triples (v,b,k) where the necessary conditions for a BIBD are fulfilled but no such design exists. A thorough study of the smallest such setting, $(15,21,5)$, has produced some surprising optimal designs.

Wednesday, June 13th, 2:30 – 2:50, AQ 4130

Polygonal Designs

John STUFKEN (Iowa State University)

It is well known that there are interesting relationships between certain concepts in design of experiments and survey sampling. In particular, there is a close relationship between block designs and sampling plans. Motivated by the desire to avoid the selection of contiguous units in a sample, A. Hedayat, C.R. Rao and J. Stufken (Journal of Statistical Planning and Inference 19, 159-170, 1988) initiated the study of a class of block designs now known as polygonal designs. We will discuss recent results on polygonal designs and related block designs, and provide directions for future research motivated by the sampling application.

Survey Methods Contributed Session: Estimation 1

Wednesday, June 13th, 1:30 – 1:50, AQ 4120

A Test for Survival Distributions using Data from Complex Samples

Susanna Rubin BLEUER (Statistics Canada)

The distribution of spell durations and other lifetime distributions may be estimated by applying survival analysis methods. Survival analysis methods were developed for independent identically distributed data (i.i.d.) from a super-population. When data is obtained from complex surveys the i.i.d. assumption is not valid and standard results may not apply. Examples of Statistics Canada lifetime data from complex surveys are lengths of spells of low income, which can be obtained from the Longitudinal Administrative Database (LAD). Data on spells of unemployment can be obtained from the Survey of Labour and Income Dynamics (SLID). In this paper, we concentrate in a non-parametric significance test to assess observed differences in empirical survival curves. We develop a test, based on the Gehan-Wilcoxon Statistic, that can be applied to data from p.p.s. design. The necessary results require the use of the structure of a product probability space which includes both the sampling space and the super-population.

Wednesday, June 13th, 1:50 – 2:10, AQ 4120

Complexities in Analysis of Duration Data from Longitudinal Surveys - with Application to the Survey of Labor and Income Dynamics

Georgia R. ROBERTS (Statistics Canada)

The distribution of spell durations and other lifetime distributions may be estimated by applying survival analysis methods. Survival analysis methods were developed for independent identically distributed data (i.i.d.) from a super-population. When data is obtained from complex surveys the i.i.d. assumption is not valid and standard results may not apply. Examples of Statistics Canada lifetime data from complex surveys are lengths of spells of low income, which can be obtained from the Longitudinal Administrative Database (LAD). Data on spells of unemployment can be obtained from the Survey of Labour and Income Dynamics (SLID). In this paper, we concentrate in a non-parametric significance test to assess observed differences in empirical survival curves. We develop a test, based on the Gehan-Wilcoxon Statistic, that can be applied to data from p.p.s. design. The necessary results require the use of the structure of a product probability space which includes both the sampling space and the super-population.

Wednesday, June 13th, 2:10 – 2:30, AQ 4120

On the Treatment of Influential Observations in Household Surveys

Asma ALAVI (Statistics Canada)

Household expenditure or income surveys often deal with highly skewed distributions, which potentially lead to samples with some extreme observations. The problem is aggravated by the fact that there is a low amount of useful auxiliary information available at the design stage and that the sampling design is complex most of the time, leading to widely dispersed design weights. Therefore, it could happen that a large value be associated with a large design weight and that this combination have a great influence on the estimates produced by the survey. Design consistent estimators, such as Generalized REGression (GREG) estimators, are usually highly variable in the presence of influential observations but they have a low bias whereas model-based estimators are more stable but they are generally not consistent and more biased. In this paper, a compromise between these two types of estimators is proposed and a simulation study shows that it performs well with respect to the bias and mean squared error (MSE) criteria in comparison with some other robust estimators. Conditions under which the compromise should have a small design bias are also given.

Wednesday, June 13th, 2:30 – 2:50, AQ 4120

The Accuracy of a Prediction Model to Oversample Individuals with High Levels of Medical Expenditures in the Medical Expenditure Panel Survey

Steven COHEN (Agency for Healthcare Quality and Research)

Efforts to improve the precision of health care estimates associated with individuals that incur high levels of medical expenditures were considered in the Medical Expenditure Panel Survey. The concentration of health care expenditures in a relatively small percentage of the population served to motivate additional efforts to insure sufficient sample representation of this core analytical domain, to support subgroup analyses. Using data from the earlier National Medical Expenditure Survey, a probabilistic model was developed to identify individuals likely to incur high levels of medical expenditures, based on their prior health care experience. In this paper, the prediction model specification is described in detail. In addition, the accuracy of this model to predict individuals likely to incur high levels of medical expenditures is assessed, using data from the Medical Expenditure Panel Survey. The Medical Expenditure Panel Survey was designed to produce national and regional annual estimates of the health care utilization, expenditures, sources of payment and insurance coverage of the U.S. civilian non-institutionalized population. The MEPS includes a survey of medical providers, to supplement the data provided by household respondents. The design of the MEPS survey permits both person based and family level estimates. The scope and depth of this data collection effort reflects the needs of government agencies, legislative bodies, and health professionals for the comprehensive national estimates needed in the formulation and analysis of national health policies. The survey is sponsored by the Agency for Healthcare Research and Quality (AHRQ) with co-sponsorship from the National Center for Health Statistics (NCHS).

Statistics in Fisheries III

Wednesday, June 13th, 3:30 – 4:00, Images

Using Poststratification to Incorporate Habitat Information into Abundance Estimates from Multispecies Surveys

Sherri DRESSEL (University of Alaska, Fairbanks)

Numerous studies have linked the distribution, abundance and survival of groundfish species to habitat characteristics, but population assessments seldom incorporate habitat information or make full use of previously determined distributions of groundfish density. Because habitat parameters have been shown to affect the spatial distribution of groundfish density and overall groundfish abundance, their use in research and stock assessment should produce more precise and reliable population estimates. Often habitat information is not known prior to sampling, however, and as a result cannot be used for stratifying groundfish trawl surveys used for monitoring species' abundance. Even if habitat information is known prior to sampling, species monitored by groundfish trawl surveys may distribute differently in relationship to habitat parameters, making it difficult to stratify optimally for each species. This study used poststratification to incorporate habitat and fish density information from trawl surveys into assessments of total abundance for four species of juvenile flatfish: age-0 rock sole (*Pleuronectes bilineatus*), age-1 yellowfin sole (*Pleuronectes asper*), age-0 Pacific halibut (*Hippoglossus stenolepis*), and age-0 flathead sole (*Hippoglossoides elassodon*). I also describe a time-efficient method of determining statistically significant changes in annual abundance, which closely approximate annual differences in the poststratified estimate. Results indicate the importance of identifying species' habitat areas when conducting population assessments and depict the benefits and drawbacks of using the poststratification method. While this particular study focused on four juvenile flatfish species, the purpose of the study was to develop analytical methods for estimating total abundance and for detecting annual variations in abundance, which could be used for any species, particularly those from multispecies surveys.

Wednesday, June 13th, 4:00 – 4:30, Images

Sampling Problems for Which Textbooks Did Not Prepare You

Brenda NORCROSS (University of Alaska, Fairbanks)

It is a well-acknowledged fact that a good offense is the best defense. In conducting fish surveys that means that a good field sampling design will yield the most meaningful and defensible results. Unfortunately, in the real world of fisheries sampling, not only are there constraints to a good design, but also there are unknown, and thus unplanned, complications on the survey is underway. One example is found in bottom trawl surveys. Although a stratified random sample design may be the goal at the outset, in Alaska and British Columbia the bottom topography may constrain the ability to establish representative strata. This often is further confounded in the field by differences between actual and charted bottom structure and type, as well as weather-induced limits to sampling. Additional examples of problems with designing and conducting surveys will be presented. Statistical analyses employed will be described. Thought provoking discussion from the audience on statistical alternatives is encouraged.

Wednesday, June 13th, 4:30 – 5:00, Images

A Bootstrap Approach to Make Sample-Size Calculations for Trawl Studies

Steven SYRJALA (NMFS, Seattle)

A sometimes difficult aspect of designing a trawl study is making sample-size calculations. A common study design is a two-sample experiment looking for a proportional change in a continuous variable, for example, species density before versus after a disturbance. Sample-size calculations for two-sample experiments are frequently made using two-sample t-test theory. Trawl data are often drawn from non-negative, positively skewed distributions. When using such data to make sample-size calculations, the interest in a proportional change and the skewed distribution of the data suggest the idea of using a log transformation so that the situation more closely conforms to the structure and assumptions of a t-test. The transformation converts the proportional change to an additive change and usually makes the distribution of the data more symmetric, closer to a normal distribution. A problem, however, arises when the data to be used contain zeros. A common solution to this problem is to use the transformation $\ln(X_i + \delta)$ for the data X_i and some small value δ (most often 1). Unfortunately, the choice of a value for δ can dramatically affect the results of sample-size calculations. An alternative approach to making sample-size calculations is to use the bootstrap. Under the null hypothesis of no change, the distribution of a metric for the distance between two samples from the same population can be generated. Then the distribution of the same metric under a change in the population specified under the alternative hypothesis can be used to estimate power for a given sample size.

Brownian Sheets and Stochastic Partial Differential Equations

Wednesday, June 13th, 3:30 – 4:00, AQ 4140

Non-Independence Of Excursions Of The Brownian Sheet

Robert DALANG and T. MOUNTFORD (Ecole Polytechnique Federale, Switzerland)

A classical and important property of Brownian motion is that given its zero set, distinct excursions away from zero are independent. In this talk, we examine the analogous question for the Brownian sheet, and also for additive Brownian motion. The main result is that given the level set at level zero of the Brownian sheet, distinct excursions of the sheet away from zero are *not* independent. In fact, given the absolute value of the Brownian sheet in the entire non-negative quadrant, and the sign of all but a finite number of excursions away from zero, the signs of the remaining excursions are determined. For additive Brownian motion, we prove the following definitive result: given the absolute value of additive Brownian motion in the non-negative quadrant and the sign of a single excursion, the signs of all other excursions are determined.

Wednesday, June 13th, 4:00 – 4:30, AQ 4140

The Dimension Of The Boundary Of Brownian Sheet Via Additive Diffusions

Thomas MOUNTFORD (University of California, Los Angeles) and Robert DALANG (Ecole Polytechnique Federale)

The Dalang-Walsh algorithm for "evaluating" the bubble of additive diffusions is used to evaluate the Hausdorff dimension of the boundary of a bubble of additive Brownian motion. This method is then adapted to deal with the Brownian sheet.

Wednesday, June 13th, 4:30 – 5:00, AQ 4140

Some Recent Results Of Interest

John WALSH (University of British Columbia)

We investigate the two-point stochastic boundary-value problem on $[0,1]$:
$$U'' = f(U) \dot{W} + g(U, U') \dot{U}(0) + \xi \dot{U}(1) + \eta$$
 where \dot{W} is a white noise on $[0,1]$, ξ and η are random variables, and f and g are continuous real-valued functions. This is the stochastic analogue of the deterministic two point boundary-value problem, which is a classical example of bifurcation. We find that if f and g are affine, the problem does not exhibit bifurcation: for any r.v. ξ and η , (1) has a unique solution a.s. However, as soon as f is non-linear, bifurcation appears. We investigate when there is positive probability that there is either no solution whatsoever, a unique solution, or multiple solutions. While these all involve conditions on f and g , we conjecture that the only case in which there is no bifurcation is when f is affine.

Survey Methods Contributed Session: Estimation 2Wednesday, June 13th, 3:30 – 4:00, AQ 4150**Efficient Estimation of Quadratic Finite Population Functions in the Presence of Auxiliary Information**

Changbao WU and Randy R. SITTER (University of Waterloo)

By viewing quadratic and other second-order finite population functions as totals or means over a derived synthetic finite population, we show that the recently proposed model-calibration and pseudo empirical likelihood methods for effective use of auxiliary information from survey data can be readily extended to obtain efficient estimators of quadratic and other second-order finite population functions. In particular, estimation of a finite population variance, a covariance, or variance of a linear estimator can be greatly improved when auxiliary information is available. The proposed methods are model-assisted in that the resulting estimators are asymptotically design-unbiased irrespective of the correctness of a working model but very efficient if the working model is nearly correct. They have a number of attractive features which include: applicability to a general sampling design; incorporation of information on possibly multivariate auxiliary variables; the ability to entertain linear or nonlinear working models; and they result in nonnegative estimates for certain strictly positive quantities such as variances. Several existing estimators are shown to be special cases of the proposed general methodology under a linear working model.

Wednesday, June 13th, 4:00 – 4:30, AQ 4150**Existence and Consistency of GEE Estimators - Application to Design-Based Inference**

Ioana SCHIOPU-KRATINA(Statistics Canada)

Conditions for existence of roots of weighted and unweighted Generalized Estimating Equations are given. The consistency of the estimators thus obtained is studied. The results are applied to design-based inference when the design is multistage with replacement at the first stage.

Wednesday, June 13th, 4:30 – 5:00, AQ 4150**Benchmarking Parameter Estimates in Logit Models of Binary Choice**

Ian CAHILL(Statistics Canada)

An approach to exploiting the data from multiple surveys and epochs by benchmarking the parameter estimates of logit models of binary choice and semiparametric survival models is developed. The goal is to exploit the relatively rich source of socio-economic covariates offered by Statistics Canada's Survey of Labour and Income Dynamics (SLID), and also the historical time-span of the Labour Force Survey (LFS), enhanced by following individuals through each interview in their six-month rotation. A demonstration of how the method can be applied is given, using the maternity leave module of the LifePaths dynamic microsimulation project at Statistics Canada. The choice of maternity leave over job separation is specified as a binary logit model, while the duration of leave is specified as a semiparametric proportional hazards survival model with covariates together with a baseline hazard permitted to change each month. Both models are initially estimated by maximum likelihood from SLID data pooled from 1993-1997, then benchmarked to annual estimates from the LFS 1976-1992. In the case of the logit model, the linear predictor is adjusted by a log-odds estimate from the LFS. For the survival model, a Kaplan-Meier estimator of the hazard function from the LFS is used to adjust the predicted hazard in the semiparametric model.

Contributed Session: RegressionWednesday, June 13th, 3:30 – 3:45, AQ 4130**Bayesian Nonparametric Regression Using Wavelets**

Jean-Francois ANGERS (Universite de Montreal) and Mohan DELAMPADY, (ISI-Bangalore)

Multi-resolution analysis is used here to derive a wavelet smoother as an estimated regression function for a given set of noisy data. The hierarchical Bayesian approach is employed to model the regression function using a wavelet basis and to perform the subsequent estimations. The Bayesian model selection tool of Bayes factor is used to select the optimal resolution level of the multi-resolution analysis. Error bands are provided as an index of estimation error. The methodology is illustrated with two examples.

Wednesday, June 13th, 3:45 – 4:00, AQ 4130

A Markov Switching Negative Binomial Regression Model and Its Applications

Peiming WANG (Nanyang Technological University)

This talk presents a method for modeling repeated count data, in which the switching process is modeled by using a Markov chain, and conditional on the states of the Markov chain, the count response has a negative binomial distribution. It is allowed to incorporate covariates in either the negative binomial parameters or the transition probabilities of the Markov chain or both. As in the case of the switching-regression model, two regimes are distinguished with either known different data generating processes or the estimated posterior probabilities of the proposed model. The features of the proposed model, the maximum likelihood estimation and some finite-sample properties of the estimates are discussed. The method is illustrated with an application.

Wednesday, June 13th, 4:00 – 4:15, AQ 4130

Estimation of Regression Parameters with Missing Covariates

Cyntha STRUTHERS and D. L. McLEISH (St. Jeromes University and University of Waterloo)

It is common in applications of regression for one or more covariates to be unobserved for some of the experimental subjects, either by design (for example they are expensive or difficult to obtain) or by some random censoring mechanism. Specifically, suppose Y is a response variable, possibly multivariate, with a density function $f(y|x,v;\alpha)$ conditional on the covariates (x,v) where x and v are vectors and α is a vector of unknown parameters. We consider the problem of estimating the parameters α when data on the covariate vector v are available for all observations while data on the covariate x are missing for some of the observations. We assume $D=1$ or 0 as the covariate x is observed or not and $E(D|Y,X,V)=\pi(Y,V)>0$ where π is a known function depending only on the observable quantities (Y,V) . Thus, v is always observed but the covariate x are potentially missing and the probability that they are so depends only on the observed quantities Y and V . Variations on this problem have been considered by a number of authors, including Chatterjee and Breslow, Lawless and Kalbfleisch, Reilly and Pepe, Carrol and Wand, Pepe and Fleming, Robbins, Scott and Wild. We suggest and compare estimators and algorithms for this problem when the observations and covariates are discrete or continuous. Remarkably, even though x may be missing for a large fraction of the sample, this may result in little loss of efficiency for estimating the parameter α .

Wednesday, June 13th, 4:15 – 4:30, AQ 4130

Abrupt Change or Not? -- From the Broken Stick to the Bent Cable

Grace CHIU, Richard ROUTLEDGE and Richard LOCKHART (Simon Fraser University)

In the area of change-point analyses and segmented regression, the broken-stick model is widely used to describe continuous regression relationships which may differ radically on either side of an unknown change point. In my Ph.D. research, I develop estimation theory for a smooth transition model which I call the "bent cable." This model was originally suggested by Tishler and Zang (1981) as a numerical tool to smooth out the broken-stick function, permitting the use of a faster and more stable numerical algorithm to estimate the unknown break. The bent cable comprises two straight lines joined by a quadratic curve in the middle. In statistical modeling, it is an alternative to the broken stick which has been greatly criticized in many applications because of the abruptness assumption. Instead, the more flexible bent cable incorporates a possible region of smooth transition. In fact, the broken stick is merely a bent cable with an infinitesimal curved segment. Extensive estimation theory for the bent-cable model was unavailable until I successfully identified some interesting properties when the underlying relationship has (1) all three segments, or (2) no quadratic segment. In this presentation, I will mention some of these properties, and show examples of the application of the bent-cable model.

Wednesday, June 13th, 4:30 – 4:45, AQ 4130**A Paired Poisson Modeling Approach to Binomial Mixed Models**

Renjun MA (University of New Brunswick)

The generalization of the linear mixed model equations of Henderson (1975, Biometrics) has been the most widely adopted non-Bayesian approach to generalized linear mixed models (Breslow and Clayton 1993; Lee and Nelder 1996); however, these approaches are known to give inconsistent estimators for binomial mixed models. Jiang (1998, JASA) introduced a method of simulated moments for consistent parameter estimation in generalized linear mixed models, but this approach is inefficient. We propose a paired Poisson modeling approach to binomial mixed models based on our BLUP approach to Poisson mixed models. Our approach not only gives consistent estimators for both regression and dispersion parameters, but also leads to optimal regression parameter estimators in the sense of Godambe (1976, Biometrika).

Wednesday, June 13th, 4:45 – 5:00, AQ 4130**Event History Regression Models for Interval Censored Covariates**

Bingshu CHEN and Richard COOK (University of Waterloo)

In event history analysis it is often of interest to study how a marker process affects the distribution of another process, called the response process. The marker process may be reflected by a simple two-state model for the onset of some intermediate clinical event and the response may be a survival time (e.g. Guggins et al., 1999). Alternatively, the marker process may be represented by a point process and the response may be governed by a more general stochastic model. The motivating example for our work is a clinical trial of patients with cancer metastatic to bone. We are interested in assessing how the development of new metastatic bone lesions affects the incidence of skeletal complications. Here we investigate methods which allow one to express the association through regression models. This is achieved by considering the marker process (tumor development) as an internal time dependent covariate for a recurrent event defined by the occurrence of skeletal complications. New tumors are only detected upon period radiographic examination making the marker processes subject to interval censoring. An EM algorithm will be adopted to deal with this incomplete data problem.

Contributed Session: AssortedWednesday, June 13th, 3:30 – 3:45, AQ 4120**The Role of the Shape Parameter for the Shrinkage Estimators of the Mean**

Shahajahan KHAN (University of Southern Queensland)

The role of the shape parameter in determining the properties of the shrinkage and the positive-rule shrinkage estimators of the mean vector of multivariate populations is of main interest in this paper. The preliminary test approach to shrinkage estimation is used to define two Stein-type estimators based on the sample information and uncertain prior non-sample information. The impact of the change in the value of the shape parameter on the performances of the estimators with respect to the criteria of unbiasedness and quadratic risk is investigated. Graphical analysis of the effect of dimension of the population on the above properties is also provided.

Wednesday, June 13th, 3:45 – 4:00, AQ 4120**Interpretable Dimension Reduction**

Hong GU and Hugh CHIPMAN (University of Waterloo)

The analysis of high-dimensional data often begins with the identification of lower dimensional subspaces. Principal component analysis (PCA) is a dimension reduction technique that identifies linear combinations of variables along which most variation occurs or which "reconstruct" the original variables the best. For example, many temperature readings may be taken in a production process when in fact there are just a few underlying variables driving the process. A problem with principal components is that the linear combinations can seem quite arbitrary. To make them more interpretable, we introduce constraints on the coefficients of the linear combination. Two classes of constraints are considered - those in which coefficients are equal to one of a small number of values (homogeneous constraints), and those in which many coefficients are set to 0 (sparsity constraints). The resultant interpretable directions may be calculated to be "close" to the original principal components, or calculated in a stepwise manner that may make the components more orthogonal. A small dataset on characteristics of cars is used to introduce the techniques, which are then applied to a more substantial "data mining" application.

Wednesday, June 13th, 4:00 – 4:15, AQ 4120**Reduction Of The Bias For The Estimator Of A Population Size**

Tina LEVESQUE (Université Laval)

The estimators for a closed population size obtained from capture-recapture experiments that have been proposed have important biases when the capture probabilities are small. Therefore, it would be interesting to develop a method to reduce the bias of these estimators. In this talk, the Petersen model for two capture occasions will first be presented followed by the Mt model and the Removal model. Then a correction for the bias in the case of small samples will be introduced. Finally, simulation results will be presented. M. Louis-Paul Rivest, Professor of Statistics at l'Université Laval, supervised this work.

Wednesday, June 13th, 4:15 – 4:30, AQ 4120**A Noniterative Monte Carlo Method for Multivariate Sampling and Integration**

Liqun WANG and James C. FU (University of Manitoba)

We propose a non-iterative algorithm for random sample generation from a given multivariate distribution. This method requires the knowledge of the density function (up to a normalizing constant) only. It is based on random-discretization of the sample space and multivariate probability inversion. It is easy to implement, efficient to computation and applicable to almost all multivariate distributions arising in real applications. Due to its simplicity and efficiency, this algorithm enables flexible choice of prior distributions, which is important in non-conjugate Bayesian analysis of real world problems. This approach is applied to many benchmark examples in the recent literature.

Wednesday, June 13th, 4:30 – 4:45, AQ 4120**Normal Transformation-Based Percentile Limits for Lifetime Distributions**

Zhenlin YANG (National University of Singapore)

Common lifetime distributions are transformable to exact or near normality using a power transformation. Exact or approximate percentile limits are thus obtainable through a simple inverse transformation. When the transformation is unknown and is estimated from the data, a common practice is to replace the unknown transformation by its estimator and the usual inferences proceed as if the estimated transformation were the true one. In this paper, we show that the percentile limits following an estimated transformation do not have a correct coverage, even when sample size n is large. A simple analytical correction is given to ensure that the corrected percentile limits have a correct coverage when n is large. Extensive Monte Carlo simulation shows that the corrected percentile limits also perform well when n is not large. When compared with the existing frequentist percentile limits for a particular lifetime distribution such as Weibull, gamma, lognormal or inverse Gaussian, the new percentile limits are seen to be equivalent or better. The new percentile limits are practically attractive due to its simplicity and generality.

Wednesday, June 13th, 4:45 – 5:00, AQ 4120**On Linear and Nonlinear Mixed Model Approaches to Cosinor Analysis**

Susan MIKULICH, Gary O. ZERBE and Thomas J. CROWLEY (University of Colorado)

Research using cosinor analysis suggests that human activity patterns reflect aspects of the underlying biology of multiple psychiatric disorders. The cosinor model, applied to repeated-measures data governed by circadian and other biological rhythms, is a nonlinear model in the amplitude and acrophase parameters characterizing each harmonic. The cosinor model also has an intrinsically linear form, i.e. it can be transformed to a linear model where amplitude and acrophase for each harmonic can be computed as nonlinear functions of the estimated linear regression coefficients. Previous studies assessing activity have used antiquated methods to estimate and compare these nonlinear circadian parameters. Here a mixed model approach to cosinor analysis is considered, where the fixed effect parameters may enter nonlinearly as acrophase and amplitude for each harmonic or linearly after transformation to regression coefficients. In addition, the random effects may enter nonlinearly as subject-specific deviations from the acrophases and amplitudes or linearly as subject-specific deviations from the regression coefficients. We also entertain the “hybrid case”, where the fixed effects enter nonlinearly while the random effects enter linearly. The concept of linear random effects in combination with either linear or nonlinear fixed effects can be extended to include higher-order harmonic terms as random effects (i.e. Rao's covariance-adjustment in the mixed model) to improve precision and power for detecting group differences. Applying the delta method to linear mixed models produced results that were often identical to results from nonlinear mixed models. Consequently, by applying the delta method, traditional linear cosinor analysis can often be used to estimate and compare the nonlinear parameters of interest. One of the aims of this paper is to clarify when the two methods are identical and when they differ. Support: NIDA grants DA11015 & DA09842

Environmental Toxicology

Thursday, June 14th, 8:30 – 9:00, Halpern
Models and Methods for Time-Dependent Bioassays
 John HUBERT (University of Guelph)

In aquatic toxicity it is becoming more common to establish a dose-response relationship from a quantal bioassay which has been monitored over time. This response-dose-time relationship is essential for risk assessment issues. When there are associated covariables, there are at least two methods for finding the point and interval estimators for certain key parameters (such as the ED50). These methods will be explained through real data sets, and recommendations for their use will be clarified.

Thursday, June 14th, 9:00 – 10:00, Halpern
Statistical Issues in Canadian Pulp and Paper Environmental Effects
 Abdel EL-SHAARAWI and Jim SHERRY (National Water Research Institute)

The Fisheries Act of 1992 required the Canadian pulp and paper mills to conduct an Environmental Effects Monitoring (EEM) study by April 1, 1996. Following the submission of the first report the mills are required to complete a cycle of EEM studies every three years. The study cycle consists of field (adult fish survey), laboratory (sublethal toxicity tests), and fish consumption (tainting) components. The laboratory component consists of a series of standard toxicity tests on aquatic organisms. A test is a dose - response experiment in which the responses of individuals exposed to a series of the pulp mill effluent concentrations are measured. An overview of the scope and objectives of the program will be presented along with the statistical issues related to its design and data analysis. Data sets will be used for illustrations.

Contributed Session: Assorted

Thursday, June 14th, 8:30 – 8:45, AQ 4150
Robustness Of Depth-Based Procedures
 Ivan MIZERA (University of Alberta)

Robustness and stability of procedures based on the half space (Tukey) depth (depth contours and maximum depth estimators in location, regression, and other models) and their possible consequences for data-analytic practice will be addressed. The investigated criteria include pointwise continuity, qualitative robustness (weak continuity), breakdown point, and others.

Thursday, June 14th, 8:45 – 9:00, AQ 4150
Flexible Estimation Of Lag Distribution Of The Effect Of Instantaneous Exposures
 Bernard RACHE, M. Abrahamowicz. and A.J. Sasco

Epidemiologic studies often have to take into account a possible lag period, i.e. a time period between an exposure and the subsequent change in risk. Previous methods for modeling lags have been more or less empirical, relying more on a priori hypotheses than on the analyses of relevant data. In this study, we address the issue of estimating the distribution of individual lag periods from the classic right-censored time-to-event data. The challenge is that lags are latent, in the sense that change in risk is not directly observable but has to be inferred from the observed events, that may occur several years later. We propose here the following three-step method for estimating the distribution of lags: a) modeling changes over time in the hazard ratio (HR), b) estimating the limits of the lag distribution, and c) reconstructing the distribution of individual lags. The first step uses a non-parametric regression model based on a generalization of the proportional hazards (PH) model that allows for non-monotonic changes in hazard ratio over time, using quadratic spline functions¹. Statistical significance of changes over time in HR is tested by the likelihood ratio test (LRT) that compares the fit to data between the flexible model and the conventional PH model. Considering that a lag in the exposure effect cannot exist without a change in the HR over time, the following steps are carried out only if the LRT rejects the null PH hypothesis at the 0.05 significance level. The second step of the analyses focuses on estimating the lower and upper limits of the lag distribution, t_1 and t_2 , from the estimated HR function. Both t_1 and t_2 are first estimated based on the smoothed local slope of the estimated function in five separate mutually exclusive random subsets of the original data. Five-fold cross-validation procedure is then employed to identify the t_1 and t_2 values that maximize the cross-validated partial likelihood. In the third step, the distribution of lag values within the interval (t_1, t_2) , is reconstructed from the estimated HR function. The proposed approach is evaluated on simulated data and applied to the Framingham data to investigate the lagged effect of smoking cessation on risk of coronary heart disease.

Thursday, June 14th, 9:00 – 9:15, AQ 4150**Probabilistic Assessment of Study Quality in Meta-Analysis**

Shagufta SULTAN and Robert PLATT (McGill University)

Meta-analysis is a set of statistical procedures designed to accumulate experimental and correlational results across independent studies that address a related set of research questions. It uses the summary statistics from individual studies as the data points. The quality of studies in meta-analysis can vary widely. In order to take into account these variations, the variation is described by assigning quality scores to studies. A natural way to incorporate quality scores into an analysis is to weigh each study by its quality score. This method has been used by several authors and has been criticized due to its subjectivity. To improve on this approach, Tritchler (1999) introduced a probability model for the effect of quality on a summary effect measure. He derived a number of summarization methods and compared them for simulated data and also applied them on an actual set of studies for the fixed effect models. We performed two evaluations of Tritchler's approach. We compared the probability model to several other methods for quality adjustment in a simulation study. In this study we used a distribution of quality scores that more adequately reflects the quality scores in meta-analytic data. In general, the probability model gives little bias and good coverage probabilities for confidence intervals. In addition, we compared the probability model to several other methods for quality adjustment using an empirical study of 31 meta-analyses. We show that adjusting for study quality using Tritchler's approach gives important differences compared to other quality adjustment approaches.

Thursday, June 14th, 9:15 – 9:30, AQ 4150**Collaborative Activities Between Canadian And U.S. Governmental Health Statistics Organizations**

Jane GENTLEMAN and Gary CATLIN (National Center for Health Statistics)

Statistics Canada's Health Statistics Division and the U.S. government's National Center for Health Statistics have now had two of what has become annual Interchanges. The purpose of these meetings is to facilitate collaboration and cooperation between the two groups and to encourage continued interaction during the time between Interchanges. These contacts have resulted in synergy between the two groups, whose mandates and missions overlap greatly. Each year, staff members from one agency travel to the other agency, alternating between the Washington, D.C. area and Ottawa, for two days of formal and informal discussion. This talk will describe some of the joint activities that have arisen from the Interchanges. In particular, plans for a joint bi-national telephone survey will be described. The identical survey will be fielded once, at the same time in the United States and Canada. A 20-minute survey questionnaire is being developed jointly by the two groups, and cognitive testing will be done in both countries. Statistics Canada will collect data from both countries using its CATI data collection operations. The costs of the project will be shared by the two agencies. The objectives of the survey include: (1) Conducting methodological studies about the scientific and practical aspects of conducting joint multinational surveys; (2) Developing and testing identical questions to be part of the two countries' large, ongoing national health interview surveys, thus enhancing comparability of estimates and facilitating collaborative analysis; and (3) Selecting topic areas while remaining cognizant of the World Health Organization's Common Framework for Monitoring and Evaluating the Implementation of Health for All, which has a goal of developing a common health status survey to be used in multiple countries.

Thursday, June 14th, 9:30 – 9:45, AQ 4150**An Examination of the Bootstrap in the Cox Relative Risk Model**

Wenyu JIANG and John D. KALBFLEISCH (University of Waterloo)

In the context of Cox relative risk model, we propose bootstrap methods to give interval estimates of regression parameters based on the studentized partial likelihood score statistic. The distribution of this statistic is estimated by resampling its terms or by resampling subjects to fail over time while holding the risk sets fixed. We carry out simulation studies to compare these methods to Efron and Tibshirani's suggested procedure, which employs the bootstrap distribution of a Wald-type statistic based on the partial likelihood estimator. The methods based on the partial score function yield more accurate confidence intervals than those based on the partial likelihood estimator. The resampling methods applied to the score function also have substantial computational advantage over the more classical approaches.

Nonlinear TimeseriesThursday, June 14th, 10:30 – 10:50, Halpern**Nonlinear And Nonstationary Dependence Concepts In Time Series**

Dag Bjarne TJOSTHEIM (University of Bergen, Norway)

The difference between linear and nonlinear time series analysis is illustrated by two examples. The first example concerns the use of nonlinear dependence concepts in aligning time series. The dependence measures are based on the entropy and Hellinger distance functionals and applied to varve chronologies. The second example has to do with nonlinear dependence between nonstationary time series. It is indicated how Markov chain theory can be used to obtain an asymptotic theory for nonparametric estimates in such a context. The motivation comes from the problem of nonlinear cointegration in econometrics, but it can be applied to quite general input-output systems of nonstationary processes.

Thursday, June 14th, 10:50 – 11:10, Halpern**Nonlinear Time Series: Some Aspects Of Testing Model Adequacy And Detection**

Reg J. KULPERGER and Hao YU (University of Western Ontario)

ARCH and GARCH processes are amongst the most used nonlinear time series, with many applications in econometric time series. The main assumptions or features in these models are (i) an iid noise driving process, usually Gaussian and (ii) the structure of the conditional heteroscedastic variance term. Can these assumptions be statistically tested with a sufficient amount of data? Towards this end we examine the EDF process of the fitted residuals. The limiting Gaussian processes is obtained, which is a Brownian bridge plus some perturbation terms. Surprisingly this limit is easier to obtain than the corresponding residual EDF process for a linear process such as an AR process. A second but related question is the estimation under model misspecification. Some earlier work on a randomly weighted EDF process is reviewed. The process is useful in detecting nonlinearities in time series.

Thursday, June 14th, 11:10 – 11:30, Halpern**Varying Parameter Time Series Models**

Wai Keung LI (Hong Kong University)

The last two decades see a rapid growth on the variety of nonlinear time series models. This includes the threshold models, the functional coefficient models, the (G)ARCH type models, Markov switching models, single index models and the mixture type time series models. All these models can be grouped under the term varying parameter time series models. The modeling of such models can be tackled nicely using a semiparametric approach. In this talk we give a survey on some recent developments with emphasis on the modeling aspect. Some real examples will be given.

Thursday, June 14th, 11:30 – 11:50, Halpern**Nonparametric Tests of Conditional Heteroscedasticity in Time Series**

Min CHEN (Chinese Academy of Sciences)

A prominent feature of ARCH (autoregressive conditional heteroscedastic) models and GARCH (generalized autoregressive conditional heteroscedastic) models is the changing conditional variance. It is, however, not easy to determine whether a given time series has this feature. Two parametric tests have been proposed to detect this feature, but they heavily depend on the parametric structure specified by the user and the normality assumption for the error terms. In this talk, we outline a nonparametric approach to detect the existence of changing conditional variance.

Diagnostic TestingThursday, June 14th, 10:30 – 11:00, Images**Models, Methods and Inferences for Prevalence and Test-Accuracy Based on Imperfect-Dependent Screening Tests**

Wesley O. JOHNSON (University of California-Davis)

A common assumption made in estimating the sensitivity and specificity of two or more binary diagnostic tests in the absence of a gold-standard is one of conditional independence among tests given disease status. Although reasonable in some cases, this assumption is often untenable or untested and may lead to biased results. We propose a class of hierarchical models for the purpose of estimating the accuracies of two tests and the prevalence distribution in the absence of a gold-standard when several exchangeable populations with differing disease prevalence are available for sampling, relaxing the assumption of conditional independence between tests. We focus on the estimation of the prevalence distribution across populations. In addition, we discuss a simpler model when only a small number of non-exchangeable populations might be available for sampling. (This work is joint with Tim Hanson, Marios Georgiadis, Ramanpreet Singh and Ian Gardner)

Thursday, June 14th, 11:00 – 11:30, Images**Efficiency Of DNA Pooling To Estimate Joint Allele Frequencies And Measure Linkage Disequilibrium**

Ruth PFEIFFER (National Cancer Institute)

Pooling DNA samples for analysis can efficiently achieve two objectives: screening large numbers of individuals to estimate prevalence of genetic variants, and identifying those with the variants. We extend methods of analyzing pooled DNA samples to estimate the joint prevalence of variants at two or more loci. This joint probability is basic to understanding the risks of disease from joint exposures to the genetic variants, such as the risk modifying effects of a more common variant on a rare highly penetrant mutation. Joint prevalences of alleles at different loci are also needed to estimate the standard measure D of linkage disequilibrium, a parameter that is fundamental in population genetics and in determining the power of association studies. We show how to estimate allele frequencies and the allelic association coefficient from pooled samples, and we quantify the expected savings in number of assays used with pooling compared to the number of assays needed if every individual is tested. Our estimates are based on single-stage pooling designs, to keep laboratory protocols simple, and we allow for errors in the tests and provide simple estimators for the case of rare variants. We consider assays that can only determine whether an individual is carrying particular alleles (carrier status) as well as assays that can determine joint genotypes. The reduction in expected sample size attainable by pooling is demonstrated both for prevalence estimation and for identifying individuals with several rare mutations, and we study the operating characteristics of these procedures by simulations.

Thursday, June 14th, 11:30 – 12:00, Images**Bayesian Sample Size Estimation For Studies Involving Two Diagnostic Tests**

Nandini DENDUKURI (St. Mary's Hospital)

Sample size calculations for estimating disease prevalence or measures of diagnostic accuracy, such as sensitivity and specificity, have typically been based on frequentist methods for proportions. This approach involves making assumptions about the exact values of the parameters of interest, which are unknown or only partially known at the time of planning a study. Furthermore, the uncertainty surrounding all other parameter values are usually ignored when calculating sample sizes for a given parameter, leading to sample sizes that are much too small. We extend the Bayesian approach of Rahme, Joseph and Gyorkos (2000) to sample size estimation when two or more imperfect tests will be used. We use three different criteria based on the marginal posterior credible intervals: 1) to obtain a pre-determined average coverage for a fixed length, 2) to obtain a pre-determined average length for a fixed coverage, 3) to obtain a minimum coverage for a pre-determined length. The presence of a second test substantially decreases the sample size requirements from the situation when only one test is available. However, increasing prior uncertainty in any of the parameters sharply increases the sample size required. In some cases even an infinite sample size may not be sufficient to achieve the required precision. Our methods will be illustrated by application to a study involving two non-gold standard tests for Strongyloides Infection.

Contributed Session: AssortedThursday, June 14th, 10:30 – 10:45, AQ 4150**Confidence Interval for Dose Response Covers Subject to Certain Increase**

Xiaoli HOU (ASA)

Dose response experiments are widely used in scientific research. The dose of a compound associated with a certain increase in dose response is usually critical in dose response experiments and need to be estimated. A general method is introduced to establish a confidence interval for the dose subject to a certain increase in dose response.

Thursday, June 14th, 10:45 – 11:00, AQ 4150**Traffic Control Based on An Ergodicity Concept**

Otto MOESCHLIN (Fern Universität)

An optimal control of traffic lights means a maximal throughput per time unit as well as minimal and fair waiting times for the traffic participants. Prototypically for even more complicated forms of traffic organization as junctions and roundabouts is the control of the so-called asymmetric bottleneck, which means that the arrival processes from the both sides are different. The approach given here is decision theoretically oriented; it uses a fixed point theorem for operators on special sets of probability measures, which ensures ergodicity of the queueing process.

Thursday, June 14th, 11:00 – 11:15, AQ 4150**Structural Properties Of A Bandit Process**

Xikui WANG (University of Manitoba)

The talk is focused on exploring structural properties of the optimal strategy for a bandit process with delayed responses. The advantage of using these properties in computation and simulation is also discussed.

Thursday, June 14th, 11:15 – 11:30, AQ 4150**Estimation Of A Bernoulli Probability From Pooled Samples With Application To Virus Infection Rates In Mosquitoes**

Brad BIGGERSTAFF (US Centers for Disease Control and Prevention)

Likelihood methods are used to estimate a Bernoulli probability p where observations have been divided into pools and each pool's aggregate response recorded as success or failure. Current methodology assumes pool sizes are equal, often an unrealistic assumption in practice, so methods are developed allowing the pool sizes to differ. A convenient, iterative form of the maximum likelihood estimator (MLE) of p is given, and its asymptotic variance is derived. The three usual likelihood confidence intervals---Wald, likelihood ratio, and score---are computed. Simplified estimators based on either ignoring pooling or using the average pool size are also considered, and all estimators and confidence intervals are evaluated in a simulation study. The methods are demonstrated for a study of West Nile virus infection rate in mosquitoes sampled in Queens, New York, 1999.

Thursday, June 14th, 11:30 – 11:45, AQ 4150**Optimality Of Nonparametric Tests For Association Between Survival Time And Continuous Covariates**

Karen LEFFONDRE, J. LELLOUCH, C. COM-NOUGUÉ, and T. MOREAU (McGill University Health Center)

Liu et al. (JASA, 1993) have proposed a class of weighted logrank tests for ordered alternatives in survival analysis. We propose to investigate this class of nonparametric tests in the case of a continuous covariate. Nonparametric tests, that replace both values of dependent and independent variables by their ranks, have the advantages to do not depend on the choice of the measurement scale of the covariate, and to be less sensitive to outliers in the covariate space than the semi-parametric Cox model. A reformulation of the test statistics, using the ranks of the covariate, allows us to derive a new multivariable model and to introduce a new procedure to identify the alternative for which each test of the class is optimal. This differs from usual optimality studies in that we investigate the alternative for which a given test is optimal, instead of investigating which test is optimal for a given alternative. Using this new approach, we show that the ordered versions of the logrank and Gehan tests are optimal for particular proportional hazards and proportional odds alternatives, respectively. A simulation study compares the size and the power of these two tests with those of the logit-rank and Cox score tests, under the alternatives identified by our approach. Simulation results agree with the theoretical optimality results and suggest that the ordered logrank test may be a serious competitor for the logit-rank and Cox score tests even for a proportional hazards alternative.

Thursday, June 14th, 11:45 – 12:00, AQ 4150

Neo-Adjuvant Hormone Therapy + Radiation For T3, Margin Positive Prostate Cancer

Eric AGYEKUM, C. LUDGATE et al (University of Victoria)

Increasing patient selection has resulted in a decrease in post operative positive margins, however 40-80% of patients are likely to have residual cancer following radical surgery (Solway, CUOG). Following prostatectomy the prostate specific antigen (PSA) level should become undetectable. A rise in PSA signifies cancer recurrence. Radiation therapy to the prostate bed has become a standard treatment for definitely positive margin disease, however the results in terms of PSA disease free survival following radiation have varied from 53 to 98 % depending on risk factors. This appears better than surgery alone, with 75% of patient developing a rising PSA by 5 years (Partin). The survival benefit of post operative RT however remains unproven, and survival benefit may be limited to those treated with undetectable PSAs and/or a low Gleason score. This presumes that all residual disease is in the prostatic fossa, but if the cancer has already grown through the capsule, then it is behaving in an aggressive way, and the failure of local treatment could be due to cancer spread. The Messing study, combining immediate hormone therapy for node positive disease, shows a control and survival advantage for immediate hormone therapy. The EORTC study of locally advanced disease (Bolla), also showed an advantage for early HT in high risk patients. At the Vancouver Island Cancer Centre from 1992 onwards, there were delays of several months in starting radiation therapy due to resource limitations. Concerns over the possibility of disease progression during the radiotherapy wait times led to the use of hormone therapy to try and prevent disease progression. (Ludgate).

We present the results of patients managed in this way.

Résumés

Welcome and SSC Gold Medal Address

Lundi 11 Juin, 8h30 – 10h00, Images
La Vraisemblance Pondérée
 Jim ZIDEK (University of British Columbia)

Selon plusieurs, l'introduction de la vraisemblance a été une des plus grandes contributions à la statistique. Ses estimateurs associés, ses tests et leurs théories pour grands échantillons ont contribué à faire de la vraisemblance un des outils les plus puissants disponibles aux statisticiens. Cependant, comme l'échelle et l'envergure des études statistiques se sont accrues avec la technologie et les demandes grandissantes pour de l'information, une variété de nouvelles formes ont dû être créées. La vraisemblance pondérée est une de ces nouvelles formes dérivées des célèbres travaux de Stein et les lacunes du plus connu des estimateurs de vraisemblance: la moyenne échantillonnale. Prenant ce qui a pu être une des plus importantes contributions de la théorie de Wald, Stein a montré que l'information dérivée d'échantillons indépendants mais reliés pouvait être échangée pour obtenir des estimateurs pour les moyennes des populations séparées qui était globalement plus précis. L'idée d'échanger le biais pour la précision était née. Dans le contexte des études statistiques modernes, cette idée prend une importance énorme par exemple lors de cartographie de maladies ou l'estimation de petite surface. Dans cette présentation je décrirai comment l'estimateur de Stein peut devenir un estimateur de vraisemblance pondérée adaptatif (EVP). De plus, je passerai en revue des travaux récents avec quelques collègues, Feifang Hu, Weimin Sun, William Rosenberger, Constance van Eeden et Malay Ghosh, et nous montrerons ce qui peut être accompli avec l'EVP entre autres dans les contextes d'estimations simultanées et d'estimations de petite surface.

WNAR Presidential Invited Address

Lundi 11 Juin, 10h30 – 12h00, Images
Une Vue À Vol D'Oiseau De La Terre Et De Son Atmosphère
 Noel CRESSIE (Ohio State University)

Des satellites sur orbite polaire sondent à distance la Terre et son atmosphère, produisant des jeux de données donnant une couverture globale quotidienne. Le système « Earth Observing System » (EOS) de la NASA consistera en un ensemble de 10-15 satellites à plusieurs instruments et sera en place d'ici 10 ans. En 2010, il y aura environ 100 instruments EOS qui observeront la Terre, chacun produisant environ 80Gb de données par jour. Pour chacun des instruments, les données quotidiennes seront multiples et spatialement irrégulières. Notre but est modeste. Pour un produit de données particulier, nous souhaitons prédire des valeurs spatialement régulières à différentes résolutions; de telles valeurs sont souvent utilisées comme entrées entre autres dans des modèles généraux de circulation (MGC). Nous voulons non seulement prédire de manière optimale mais, puisque l'acquisition de données est sans arrêt, notre algorithme doit traiter les données très rapidement. Cette présentation présente une méthodologie de prédiction statistique qui préserve la « balance de la masse » dans toutes les résolutions spatiales et calcule les prédictions spatiales et les prédictions de (co)variances extrêmement rapidement. Des idées similaires peuvent être utilisées lorsque les données sont multivariées et spatio-temporelles. Des données de l'instrument Total Ozone Mapping Spectrometer (TOMS), sur le satellite Nimbus-7, seront utilisées comme exemple. Les recherches présentées sont conjointes avec Hsin-Cheng Huang, Academia Sinica, Taïwan.

Contrasting US 2000 and Canada 2001 Censuses

Lundi 11 Juin, 10h30 – 11h00, Halpern

Estimations De Calibrage Dans Le Recensement Canadien

Mike BANKIER (Statistics Canada)

Lors du recensement de la population canadienne en 1991, l'estimation de calibrage ou de régression ont été utilisés pour calculer un ensemble unique de pondérations au niveau des foyers pour être utilisé dans tous les estimés du recensement basés sur un échantillon national (1 sur 5) d'environ deux millions de foyers. Étant donné que plusieurs variables auxiliaires étaient disponibles, seul un sous-ensemble de celles-ci pouvait être utilisé. Sinon, quelques-unes des pondérations auraient pu être inférieures à un ou même négatives. Une procédure de sélection forward a été utilisée pour rejeter les variables auxiliaires pour lesquelles des pondérations inférieures à un étaient obtenues ou qui causaient un **'grand nombre de condition'** lors de l'inversion de la matrice des pondérations de calibrage. De plus, deux ajustements de calibrages ont été effectués pour obtenir un bon accord entre les décomptes auxiliaires de population et les estimés sur petites superficies. Ce dernier objectif a été atteint en 1991 mais les différences entre les décomptes et les estimés de population à l'échelle provinciale étaient parfois plus grandes que celles obtenues lors du recensement de 1986, qui utilisait l'estimation par apport de ratissage. Cela est peut-être dû à l'accumulation de biais présent dans l'échantillon ou introduit par l'estimation. L'impact de la réduction dans le nombre de variables auxiliaires disponibles dans le recensement de 1996 sera décrit. Finalement, les plans pour le recensement 2001 seront dévoilés.

Lundi, 11 Juin, 11h00 – 11h30, Halpern

Estimation De Petit Domaine Des Personnes Non Dénombrées Au Recensement 2001

Peter DICK (Statistics Canada)

Des estimés nets des personnes non dénombrées divisés selon l'âge et le sexe pour toutes les provinces et territoires sont requis pour la production d'estimés de la population en 2001. Même si des études de couverture peuvent donner des estimés fiables des personnes non dénombrées pour chaque province et territoire, la taille échantillonnale est trop restreinte pour fournir des estimés fiables à la précision voulue. Un mélange de procédures est utilisé pour produire ces estimés. Des estimés directs de sondage sont utilisés pour calculer les totaux marginaux des provinces et territoires. Une méthode de lissage à fonction spline est utilisée pour calculer les totaux nationaux d'âge-sexe pour l'autre marge. Un modèle de régression empirique bayésien donne des estimés pour des grands groupes d'âge, tandis qu'un modèle synthétique est utilisé pour des estimés plus détaillés. Ce papier discutera des méthodes utilisées dans le programme de 1996, et des améliorations prévues pour 2001.

Lundi 11 Juin, 11h30 – 12h00, Halpern

Recensement 2000 Aux É.U.: Un Survol Comparatif

Robert FAY (U.S. Census Bureau)

En juin 2001 le U.S. Census Bureau rendra public les données initiales les plus importantes du Recensement 2000 – un décompte de la population jusqu'au niveau des unités de recensement – même si d'autres données restent encore à travailler dont celles basées sur les données échantillonnales. Cet article résume le plan et l'état présent du Recensement 2000. L'intention première de ce survol est de comparer ce recensement à celui tenu au Canada en 2001. De plus, une partie de cette présentation reverra la planification des méthodes d'estimation des données échantillonnales du Recensement 2000. Statistiques Canada et le Census Bureau mènent des recherches depuis des décennies sur le sous-dénombrement de la population lié au recensement. Statistiques Canada a incorporé les estimés de sous-dénombrement net aux estimés de population post-recensement, mais la plupart des estimés officiels aux É.U. ne l'ont pas fait. Par contre, les plans pour le Recensement 2000 comprenaient des ajustements pour tous les types de données. Mais, comme ordonné par la Cour suprême des É.U., le Census Bureau a publié des décomptes par état non ajustés en décembre 2000 pour la répartition des représentants à la House of Representatives. Pendant l'hiver 2001, le Census Bureau a décidé s'il fallait statistiquement ajuster les données officielles subséquentes au Recensement 2000. Cette présentation discute des répercussions de cette décision et en revoit les fondements.

Unusual Stochastic Models

Lundi 11 Juin, 10h30 –11h00, AQ 4150

A Stochastic Adding Machine And Complex Dynamics

Tom TAYLOR (Arizona State University)

This paper considers properties of a Markov chain on the natural numbers which models a binary adding machine in which there a nonzero probability of failure each time a register attempts to increment the succeeding register and resets. This chain has a family of natural quotient Markov chains, and extends naturally to a chain on the 2-adic integers. The transition operators of these chains have a self-similar structure, and have a spectrum, which is, variously, the Julia set or filled Julia set of a quadratic map of the complex plane.

Lundi 11 Juin, 11h00 –11h30, AQ 4150

Convergence Des Échantillonneurs De Gibbs Pour Distributions Uniformes

Jeffrey ROSENTHAL (University of Toronto)

Nous considérons les échantillonneurs de Gibbs pour distributions uniformes sur des régions. Ces échantillonneurs généralisent les échantillonneurs markoviens « par section », qui échantillonnent uniformément la région sous la courbe d'une fonction de densité. Nous examinons les propriétés mathématiques de ces algorithmes. En particulier, nous prouvons les limites quantitatives de la rapidité de convergence des échantillonneurs par section, de même que les taux de convergence qualitatifs plus généraux. Nous montrons aussi que certaines propriétés de convergence dépendent grandement de la régularité des bornes de la région elle-même. Finalement, nous décrivons l'algorithme modifié « échantillonneur par section polaire ».

Ceci est un travail conjoint avec G.O. Roberts (Lancaster University).

Lundi 11 Juin, 11h30 –12h00, AQ 4150

Complétude Du Marché: Retour À L'Ordre

Georghe STOICA (St. Johns University of New Brunswick)

Si l'espace des sorties nettes, dans une économie à deux temps, est défini par un treillis linéaire de commandes remplies linéaire de fonctions à faibles unités, nous prouvons alors la complétude du marché par rapport à la topologie générée par relations d'ordre. Nous donnons aussi une définition alternative pour la complétude du marché, en termes d'étendue martingale approximative sur de tels treillis linéaires.

Session in Honour of V.P. Godambe's 75th Anniversary

Lundi 11 Juin, 1h30 –2h00, Halpern

Inférence Basée Sur La Vraisemblance Composée Lisse Et Simulée (VCLS) Pour Modèles Hiérarchiques

Subhash LELE (University of Alberta)

Les modèles hiérarchiques sont utiles pour modéliser entre autres des données spatiales non normales, des données avec erreurs de mesures et la reproduction animale. Un des problèmes majeurs de ces modèles est que l'inférence basé sur la fonction de vraisemblance est difficile à calculer. L'approche de la vraisemblance composée a été utilisée comme solution possible à ce problème. Dans cet article, nous proposons une extension de l'approche de la vraisemblance composée pour l'inférence de modèles espace-état non linéaires et non normaux. Les applications aux séries chronologiques non normales et aux séries écologiques spatiales et spatio-temporelles seront discutées.

Lundi 11 Juin, 2h00 – 2h30, AQ 4150

Abundance Estimation Based On Daily Pellet Heap Counts: An Estimating Function Approach
M.B. RAJARSHI, (University of Pune)

We analyze the data on daily counts of pellet heaps of deer in a sanctuary. Statistician is required to suggest an estimate of the deer population based on these data. In an earlier work (Gadgil et al, 1986, Biometrical Journal), we had assumed that pellet lifetime distribution is exponential or if it is not exponential, the model viewed as a queue, has an FIFO discipline. Likelihood for a non-exponential distribution is almost non-tractable for estimation of life-time distribution. We propose two estimating functions for estimation of parameters of the life-time distribution. The first one is fairly easy to work out, the other is somewhat complicated but still tractable for computer analysis. We report some real-life data analysis and a few simulation studies. Lastly, an algorithm for estimating functions in the context of incomplete data, is presented. This extends the Imputation-Maximization algorithm in the case of independent observations.

Lundi 11 Juin, 2h30 – 3h00, AQ 4150

Fonctions D'Estimation Et Inférence Pour Données De Sondages
Harold MANTEL and Milorad KOVACEVIC (Statistics Canada)

Dans cette présentation nous considérons l'usage de fonctions d'estimation pour l'inférence se basant sur des données de sondages. Les procédures d'estimation et d'estimation de variance seront décrites pour des paramètres descriptifs simples, pour des paramètres de régression linéaire multiple et pour des paramètres plus complexes, tels les coefficients de régression logistique. Les propriétés statistiques des estimés résultants seront investiguées. Les relations aux autres méthodes d'estimation seront explorées, et le rôle de l'optimalité sera discuté.

Spatial Statistics

Lundi 11 Juin, 1h30 – 2h00, Images

Le Bon, La Brute Et Le Truand: 30 Ans De Champs Aléatoires De Markov
Julian BESAG (University of Washington)

Les champs aléatoires markovien (CAM) sont des distributions définies via leurs caractéristiques locales (aussi connues comme conditionnelles complètes). Durant les 30 ans qui ont passé depuis que J. M. Hammersley et P. Clifford ont établi l'équivalence entre les CAM et les distributions de Gibbs en physique, leur théorème a eu une influence énorme sur le développement d'idées statistiques dans différents domaines, tels les modèles graphiques, l'analyse d'image bayésienne, les méthodes de Monte Carlo markoviennes, les tableaux de contingence multidimensionnels et les réseaux sociaux, sans compter les sujets qui tombent évidemment sous la bannière des statistiques spatiales. Cette présentation décrira quelques-uns des succès et des échecs survenus au fil des années.

Lundi Juin 11th, 2h00 – 2h30, Images

Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates
Charmaine DEAN (Simon Fraser University) and Y.C. MACNAB

Generalized additive mixed models are employed for the analysis of geographic and temporal variability of mortality rates. This class of models accommodates random spatial effects and fixed and random temporal components. Spatio-temporal models which use autoregressive local smoothing across the spatial dimension and B-spline smoothing over the temporal dimension are developed. The objective is the identification of temporal trends and the production of a series of smoothed maps from which spatial patterns of mortality risks can be monitored over time. Regions with consistently high rate estimates may be followed for further investigation. The methodology is illustrated by analysis of British Columbia infant mortality data.

Lundi 11 Juin, 2h30 – 3h00, Images
L'Inclusion De L'Autocorrélation Spatiale Dans Les Tests Statistiques
 Mark DALE (University of Alberta)

Lors de tests statistiques sur des caractéristiques écologiques, la présence d'autocorrélation spatiale dans les données peut avoir un effet considérable sur les résultats et leurs interprétations. Cette conférence présentera cette problématique et décrira quelques approches pouvant être utilisées pour la résoudre, en se concentrant sur l'analyse statistique plutôt que sur les plans d'expérience ou d'échantillonnage. Quelques-unes de ces approches sont la déflation de la statistique de test, l'ajustement pour la taille effective de l'échantillon ainsi que les techniques de permutation restreintes et de Monte-Carlo. En débutant avec une discussion sur des tests de tableaux de contingence, nous allons commenter ces approches et la validité des modèles sur lesquelles ils se basent pour l'analyse de données écologiques. Des exemples provenant de données de terrain et de simulations accompagneront la discussion. Nous concluons avec un résumé des tests pour lesquels nous savons comment inclure l'autocorrélation spatiale, et de ceux pour lesquels nous ne savons pas.

Robust Inference

Lundi 11 Juin, 1h30 – 2h00, AQ 4150
**Sur La Robustesse D'Intervalles De Confiance Du Rapport De Vraisemblance Empirique
 Pour Le Paramètre De Position**
 Min TSAO (University of Victoria)

Dans cette présentation, j'examinerai la robustesse des intervalles de confiance du rapport de vraisemblance empirique (RVE) pour la moyenne et l'estimateur M. Il a été démontré que l'intervalle RVE pour la moyenne a un point de rupture asymptotique de zéro. Une formule pour calculer le point de rupture de l'intervalle RVE pour l'estimateur M sera aussi présentée. Les avantages relatifs de l'intervalle RVE par rapport aux intervalles de confiance communément utilisés, basés sur la distribution asymptotique de l'estimateur M , seront examinés par le biais d'une étude empirique.

Lundi, 11 Juin, 2h00 – 2h30, AQ 4150
Plans Séquentiels Robustes Pour Modèles Mal-Spécifiés
 Doug WIENS (University of Alberta)

Nous allons discuter de récents travaux traitant de régression et de plans cliniques qui se veulent robustes contre une réponse incorrectement spécifiée. Par exemple, un expérimentateur veut appliquer un modèle Michaelis-Menten aux réponses, mais veut aussi un plan qui le protégera contre le biais des valeurs prédites dans le cas où la vraie fonction de réponse aurait une forme similaire mais différente. Un autre exemple, l'affectation de sujets à des traitements dans un essai clinique lorsque l'effet des facteurs pronostiques sur la variable de réponse est incorrectement modélisé. Le plan idéal protégerait contre l'augmentation de l'EQM résultant de ces erreurs dans les estimés des contrastes des traitements, tout en maintenant un niveau d'équilibre raisonnable.

Lundi 11 Juin, 2h30 – 3h00, AQ 4150
Sur Le Rôle D'Estimateurs M Redescendant En Vision Numérique (Et Autres Domaines)
 David TYLER (Rutgers University)

Il y a un long historique dans l'application de méthodes statistiques robustes en vision numérique. La plupart des premières méthodes utilisées par les chercheurs travaillant sur la vision numérique ont été proposées par ceux-ci et venaient du besoin de telles méthodes dans l'avancement de la recherche en vision numérique. Ces méthodes comprennent la transformation de Hough et RANSAC. Plus récemment, les chercheurs en vision numérique, voyant le développement de méthodes robustes en statistiques tels les estimateurs M et la médiane des erreurs au carré, ont incorporé ces nouvelles méthodes dans leur travail.

Contributed Session: Survival Analysis

Lundi 11 Juin, 1h45 – 2h00, AQ 4140

Sélection De Prédicteurs De Survie À L'Aide D'Une Méthode Non Paramétrique

Marylène TROUPÉ and J. VAILLANT (Université des Antilles-Guyane)

Le but de l'étude est de choisir des covariables avec une valeur prédictive élevée en utilisant l'information réciproque entre une variable de survie et des covariables quand les modèles paramétriques n'adaptent pas bien les phénomènes observés. La méthode, basée sur l'économie de l'information permet l'agrégation des courbes de survie correspondant à différentes modalités du même prédicteur choisi. Le taux d'information apporté par les covariables est décomposé similairement à l'analyse de la variance et alors peut être analysé. Une application au cancer du sein utilisant des données de 1 304 patientes qui ont été suivies pendant dix ans sera présentée.

Lundi 11 Juin, 1h45 – 2h00, AQ 4140

Vérification De L'Ajustement D'Un Modèle De Survie À Plusieurs Niveaux À Temps Discrets

Jennifer ORLICH and Susan MURPHY (University of Michigan)

L'analyse de survie à plusieurs niveaux à temps discrets est généralement utilisée pour analyser des données qui sont longitudinales, censurées, et en grappe. Le fait de mettre en application un modèle de survie à plusieurs niveaux à temps discrets est immédiat, le contrôle d'un tel modèle n'est pas bien développé. Deux méthodes graphiques pour évaluer l'ajustement de ce modèle seront fournies. La première méthode vérifie que la probabilité prédite de l'événement, basée sur le modèle, est approximativement l'estimation de la probabilité de ce même événement basé sur les données. La deuxième méthode contrôle si les probabilités prédites dérivées du modèle et cumulées dans le temps reflètent le nombre d'événements cumulés correspondant qui sont observés dans les données. Nous démontrons et donnons des arguments sur la façon dont le graphique devrait apparaître pour un modèle convenablement adéquat et nous fournirons des illustrations de la façon dont les méthodes détectent le manque d'ajustement. La situation où la plupart des probabilités prédites sont près de zéro est d'un intérêt particulier.

Lundi 11 Juin, 2h00 – 2h15, AQ 4140

Estimation Efficace Pour Le Modèle De Risque Proportionnel Tronqué À Gauche Et Avec Données Censurées Par Intervalle De Type 1

Jong Sung KIM (Portland State University)

L'estimateur du maximum de vraisemblance (MLE) pour les modèles de risque proportionnel tronqué à gauche et avec données censurées par intervalle de type 1 est étudié. Sous certaines conditions de régularité, le MLE du paramètre de régression s'avère asymptotiquement normal avec un taux de convergence de l'ordre de racine de n et atteint la borne d'information, quoique la différence entre le temps de troncature et le temps de censure du MLE de la fonction de répartition de risque de base converge seulement à un taux de $n^{1/3}$. Deux méthodes pour estimer la matrice de variance-covariance du paramètre de régression MLE sont à considérer. Un est basé sur un principe d'information manquante généralisé et l'autre est basé sur l'information du profil. Les études de simulations prouvent que les deux méthodes proposées fonctionnent bien en termes de biais et variance pour des échantillons de tailles modérées. Un exemple est fourni pour illustrer les méthodes.

Lundi 11 Juin, 2h15 – 2h30, AQ 4140

Estimation Des Taux De Survie Fauniques À L'Aide De Séries Chronologiques À Partir De Données Par Structure D'Age

Mark UDEVITZ (Alaska Biological Science Center)

Les données par structure d'âge ont une longue histoire d'utilisation pour estimer des taux de survie dans des populations fauniques. Des taux de survie peuvent être estimés à partir des proportions de classes d'âge soit dans la population actuelle (données selon la structure d'âge actuelle) soit dans les décès naturels (données de l'âge au temps de la mort). La plupart des estimateurs ont exigé l'acceptation souvent peu raisonnable d'une répartition par âge stable. Les répartitions par âge deviendront stables seulement si les taux de recrutement et de survie demeurent constantes sur une période de temps assez longue. Récemment, on a développé un nouvel ensemble d'estimateurs de taux de survie qui utilisent la structure d'âge actuelle et les données de l'âge au temps de mort d'une seule année, éliminant de ce fait la condition sur la stabilité de la structure d'âge. Maintenant, je généralise les fonctions de vraisemblance employés pour dériver ces estimateurs pour considérer plusieurs années de données par structure d'âge. L'utilisation de plusieurs années de l'un ou l'autre des types de données de répartition par âge permet également la relâchement de l'hypothèse de stabilité. Les procédures standards de sélection modèles basées sur la vraisemblance peuvent être employées pour tester des hypothèses au sujet de stabilité de la répartition par âge et des patrons relatifs à l'âge dans des taux de survie et pour identifier le modèle le plus approprié pour estimer ces taux.

Lundi 11 Juin, 2h30 – 2h45, AQ 4140

Un Modèle De Risque Proportionnel Partiellement Linéaire

Xuewen LU (Agriculture and Agri-Food Canada)

Le modèle de risque proportionnel de régression suppose habituellement que la covariable a un effet log-linéaire sur la fonction de risque. Dans cet article, nous considérons un modèle de survie semi paramétrique avec des effets de covariables flexibles. Nous supposons que la fonction de base de risque est paramétrée, alors que la fonction de risque associée aux covariables est modélisée d'une manière semi paramétrique. Un « algorithme à une étape » est employé pour estimer les composantes non paramétriques et les composantes paramétriques en maximisant la vraisemblance locale et globale, respectivement. En utilisant la méthode linéaire locale, des estimateurs des paramètres inconnus et de la fonction de covariables inconnue sont proposées, et leurs distributions asymptotiques sont obtenues. La performance de « l'algorithme à une étape » est comparée à celle de l'algorithme entièrement réitéré et à la méthode du profile de vraisemblance. Une étude de simulations illustre les algorithmes et les modèles.

Lundi 11 Juin, 2h45 – 3h00, AQ 4140

Utilisation D'Informations Auxiliaires Pour Des Intervalles De Survie Censurés Et Pour L'Analyse Des Données Des Événements Historiques

Jaroslaw HAREZLAK (Indiana University School of Medicine)

Dans l'analyse de survie ou dans l'analyse des événements historiques, lorsqu'un événement a lieu, nous observons la période de l'événement avec une bonne exactitude. Cependant, dans certaines situations, nous savons seulement que l'événement s'est produit entre deux points d'observation. Dans cette situation, une analyse standard traite l'événement comme un intervalle censuré. La méthodologie disponible sur les intervalles censurés a tendance à être moins efficace lorsque les intervalles sont grands. Dans une étude des maladies transmises sexuellement (MTS), nous avons suivi un groupe de jeunes femmes, qui étaient infectées par une MTS ou dont le partenaire était infecté (c.-à-d. contact), afin de détecter une récurrence de l'infection. À l'aide des visites régulières (à un, trois, cinq et sept mois), nous savons seulement si la ré-infection se produisait entre ces visites. En plus des données de suivi, nous avons également enregistré dans des carnets, les données quotidiennes sur les activités sexuelles des sujets expérimentaux. Avec l'information auxiliaire fournie par les carnets, nous proposons d'utiliser le temps d'apparition d'une infection possible (c.-à-d. relation sexuelle) pour améliorer l'estimation du temps de récurrence de l'infection. Un échantillon des relations sexuelles est tiré au hasard de l'intervalle de ré-infection. Chaque échantillon est analysé en utilisant une des méthodes disponibles pour des données censurées à droite, et les estimateurs sont combinés d'une façon appropriée. Notre méthode est comparée à d'autres méthodes employées pour des données avec des intervalles censurés. Une généralisation est aussi proposée pour des événements multiples par sujets. Dans nos simulations, nous utilisons les modèles non paramétriques, semi-paramétriques et paramétriques. Nos résultats indiquent que la méthode proposée a une bonne performance dans toutes les situations considérées. L'analyse de l'exemple de motivation indique un plus haut risque d'infection pour les sujets qui ont été infectés au tout début que pour ceux qui sont des contacts.

WNAR Student Paper Presentations

Lundi 11 Juin, 1h30 – 2h00, AQ 4130

Modèle Markovien Vectoriel Pour L'Analyse Des Données Historique D'Une Maladie Avec Des Résultats Multidimensionnels Et Des Observations Incomplètes

Shenghai ZHANG (University of Waterloo)

Un certain nombre de processus à plusieurs états dépendants surgissent lorsque plusieurs variables de réponse sont exigées pour mesurer les résultats d'intérêt. Chaque sujet est périodiquement visité et des résultats ont été enregistrés simultanément. L'information de l'état dans lequel un sujet se trouve est incomplète. Bien que quelques méthodes analysant de telles données aient été fournies précédemment, elles sont seulement valides pour des processus irréversibles bidimensionnels et le même intervalle de temps entre chacune des visites doit être assumé. Le but de cet article est de développer une méthodologie pour surmonter ces difficultés. Les méthodes fournies ici servent à estimer non seulement l'effet d'un traitement mais également les probabilités de transition entre les états. D'abord, l'intensité commune pour le processus markovien vectoriel à temps continu est discutée et des formules qui décrivent le rapport entre les probabilités de transition et les intensités sont établies. Basé sur une famille de distributions exponentielles multidimensionnelles qui est donnée dans cet article, les intensités communes peuvent être spécifiées par les intensités marginales qui peuvent habituellement être facilement obtenu dans des applications pratiques. D'ailleurs, un procédé pour estimer des paramètres d'intérêt est décrit.

Lundi 11 Juin, 2h00 – 2h30, AQ 4130

Comparaison Des Méthodes Statistiques Pour Des Données Traitées Par Filtration À Haut

Yuan Yuan WANG, Hugh CHIPMAN, and William WELCH (University of Waterloo)

Le traitement par filtration à haut rendement (HTS) est souvent employé dans la découverte pharmaceutique pour mesurer l'activité biologique des composés par rapport à diverses cibles biologiques. Les données produites par HTS ont des dizaines de milliers, voire des centaines de milliers de composés. De plus, les chimistes et les biologistes sont intéressés à prévoir les composés les plus actifs, qui sont relativement rares. Je comparerai les performances de divers modèles statistiques sur plusieurs jeux de données de HTS en termes de capacité à prévoir des cas extrêmes. J'identifierai également des caractéristiques particulières aux données de HTS qui mènent à certains types de bons modèles.

Lundi 11 Juin, 2h30 – 3h00, AQ 4130

Méthodes De Listes Multiples Et De Capture-Recapture Avec Stratification Partielle Ou Incomplète Appliquées À L'Estimation De L'Incidence Des Feux De ForêtJason SUTHERLAND, C.J. SCHWARZ, J.G. FALL, B. DORNER, K.P. LERTZMAN
(Simon Fraser University)

Les méthodes de listes multiples et de capture-recapture sont des méthodes généralement utilisées pour estimer la taille des populations évasives. Si possible, la stratification est une voie utile de réduire les biais provoqués par l'hétérogénéité des probabilités d'adhésion de la liste parmi des membres de la population. Dans cet article, nous développons une méthode pour traiter des cas quand toutes les listes ne sont pas en activité dans toutes les strates. La motivation de cet article est d'estimer le nombre de feux de forêt qui ont eu lieu récemment, où les différents arbres fonctionnent comme listes enregistrant les feux. Cet article présente une approche généralisée pour plus de deux listes quand toutes les listes ne couvrent pas toutes les strates et présente un estimateur de l'incidence des événements. Nous tenons compte des dépendances de liste et des probabilités différentielles de l'évaluation dans chaque liste en utilisant le cadre des modèles log-linéaires.

IMS Special Invited Session I

Lundi 11 Juin, 3h30 – 5h00, Images

Chain Graph Models and Their Causal Interpretations

Steffen LAURITZEN and Thomas S. RICHARDSON (Aalborg University, Denmark)

Chain graphs are a natural generalization of directed acyclic graphs (DAGs) and undirected graphs. However, the apparent simplicity of chain graphs belies the subtlety of the conditional independence hypotheses that they represent. There are a number of simple and apparently plausible, but ultimately fallacious interpretations of chain graphs that are often invoked, implicitly or explicitly. These interpretations also lead to flawed methods for applying background knowledge to model selection. We present a valid interpretation by showing how the distribution corresponding to a chain graph may be generated as the equilibrium distribution of dynamic models with feedback. These dynamic interpretations lead to a simple theory of intervention, extending the theory developed for DAGs. Finally, we contrast chain graph models under this interpretation with simultaneous equation models which have traditionally been used to model feedback in econometrics.

Business and Industry Section Special Invited Session

Lundi 11 Juin, 3h30 – 5h00, Images

Accelerated Testing: A Method for Obtaining Reliability Information Quickly

Bill MEEKER and Luis A. ESCOBAR (Iowa State University)

Accelerated tests are used to obtain timely information on products reliability. Changes in technology, the calls for rapid product development, and the need to continuously improve product reliability have combined to increase the need for developing improved methods for accelerated testing. Laboratory tests with increased use rates or higher than usual levels of accelerating variables like temperature or voltage are used to accelerate failure mechanism. Then the results are used to make predictions about product life or performance over time at use or design conditions. The predictions involve extrapolation in several dimensions. Interesting statistical problems arise in modeling physical phenomena, use of engineering/physical information, planning accelerated tests, and quantifying uncertainty. This talk reviews the basic physical and statistical models and methods used in accelerated testing. Current research in this area will be outlined and areas for future research will be described.

Gene Ancestry, Identity and Allelic Associations

Lundi 11 Juin, 3h30 – 4h00, AQ 4130

Un Test De Score Pour L'Analyse De Liaison De Traits Qualitatifs Et Quantitatifs Basé Sur Des Données D'Identité Par Descendance

Sandrine DUDOIT (MSRI, Berkeley)

Des facteurs génétiques contribuent à plusieurs maladies humaines en conférant susceptibilité ou résistance, et en affectant la sévérité et la progression de la maladie. La cartographie de liaison est un outil puissant pour localiser les gènes de centaines de maladies mendéliennes rares (par exemple, la maladie d'Huntington, la fibrose kystique). À l'opposé, l'identification de gènes influençant des maladies plus communes telles que le diabète ou l'hypertension a progressé lentement. Ces maladies sont complexes pour plusieurs raisons incluant des incertitudes de diagnostique (apparition tardive de la maladie), des modes de transfert inconnus et non mendéliens, et l'implication de plusieurs gènes. L'importance pour la santé publique de la compréhension des bases génétiques des maladies complexes a provoqué un intérêt accru pour le développement de méthodes statistiques pour la cartographie génétique. Cette présentation présentera une approche unifiée basée sur la vraisemblance pour la cartographie de traits humains complexes, qualitatifs ou quantitatifs, en utilisant des données d'identité par descendance (IPD) tirées de petites généalogies. L'approche unifiée considère la vraisemblance des données IPD conditionnelles aux phénotypes, et teste pour la liaison entre un locus marqueur et un gène influençant le trait en utilisant un test de score. Pour un type de généalogie donné, la forme de la statistique du score de liaison est déterminée par la deuxième plus grande valeur propre et par les vecteurs propres correspondants à la matrice de transition pour les configurations IPD sur la longueur d'un chromosome. Lorsque comparé aux approches déjà existantes, le test de score de liaison fournit une méthode robuste et puissante pour la cartographie génétique de traits complexes. Les applications des microtableaux dans le contexte de la cartographie génétique seront aussi discutées.

Lundi 11 Juin, 4h00 – 4h30, AQ 4130
Distinction Entre Loci Génétiques Causaux Et Liés
 Laura LAZZERONI (Stanford University)

Les plans de paires de germains affectés et les tests de déséquilibres de transmission sont les méthodes usuelles utilisées pour identifier et cartographier les facteurs de risques génétiques. Lorsque de tels tests donnent des résultats positifs pour un ou plusieurs polymorphismes, il est intéressant de déterminer lesquels, s'il y en a, des loci testés ont un lien causal au phénotype de la maladie et lesquels ne sont liés qu'à un locus causal. Cette présentation décrira les informations fournies par les données servant à faire cette distinction.

Lundi 11 Juin, 4h30 – 5h00, AQ 4130
Une Nouvelle Méthode Statistique Pour La Reconstruction De L'Haplotype
 Matthew STEPHENS (University of Washington)

Les méthodes expérimentales courantes pour la détermination des types génétiques des individus dans une population (génotypage) ne fournissent habituellement aucune information relative à l'haplotype – c'est-à-dire quel allèle de chaque locus est hérité de la mère ou du père. L'information sur l'haplotype est importante pour plusieurs analyses de données génétiques moléculaires à petite échelle, comprenant la cartographie de maladies et la détermination de l'histoire des populations. Moyennant de grands frais, les haplotypes peuvent être obtenus de façon complète ou partielle, expérimentalement ou par le génotypage d'autres membres de la famille. Une méthode statistique peut aussi être utilisée. Nous décrirons deux approches statistiques existantes pour cette problématique, et proposerons une nouvelle méthode statistique comportant des améliorations considérables par rapport aux algorithmes courants, et ce par l'incorporation d'idées de la théorie de la génétique des populations. (Travaux conjoints avec Nicholas Smith et Peter Donnelly)

Survey Methods Contributed Session: Imputation and Variance Estimation

Lundi 11 Juin, 3h30 – 3h50, AQ 4150
Inférence Sur Les Paramètres De Régression Avec Imputation Des Données Manquantes
 David HAZIZA (Statistics Canada)

La moyenne d'une population peut être estimée de façon non biaisée sous l'imputation par la moyenne pour des données manquantes et des réponses uniformes, mais l'estimateur obtenu d'un paramètre de régression est généralement biaisé. D'après l'approche de Skinner et de Rao (1999), nous obtenons un estimateur ajusté pour le biais d'un paramètre de régression sous un plan d'expérience arbitraire. Nous dérivons des estimateurs convergents de la variance, en utilisant une méthode présentée par Fay (1991) dans laquelle la voie habituelle de la réponse de l'échantillon est renversée. Nous allons également considérer le cadre avec modèle assisté, dans ce cas l'estimateur imputé est non biaisé pour le modèle du plan d'échantillonnage et on trouve des estimateurs de la variance convergents.

Lundi 11 Juin, 3h50 – 4h10, AQ 4150
Estimation De La Variance Pour Les Équations D'Estimation En Présence De Valeurs Manquantes
 Wesley YUNG, Mike HIDIROGLOU and J.N.K. RAO (Statistics Canada)

Des paramètres d'intérêt d'une population peuvent être exprimés comme solutions aux équations d'estimations appropriées pour la population. Des estimations échantillonales de ces paramètres peuvent être obtenues par la solution d'équations d'estimation échantillonales appropriées. Hidiroglou, Rao et Yung (1999) ont montré comment des estimateurs de la variance peuvent être obtenues comme solution des équations d'estimation qui incorpore des ajustements pondérés dus à l'information auxiliaire, telle que l'estimation pondérée généralisée de la régression (GREG). Dans cet article, nous étendons les résultats de Hidiroglou et al. (1999) au cas où des ajustements sont également faits pour la non-réponse quand des méthodes d'imputation sont appliquées pour expliquer les valeurs manquantes.

Lundi 11 Juin, 4h10 – 4h30, AQ 4150

Estimation De La Variance Pour L'Enquête Annuelle Sur Les Manufactures

Wisner JOCELYN and Garriguet DIDIER (Statistics Canada)

L'étude annuelle sur les manufactures rassemble de l'information financière et sur les produits manufacturés. Elle est constamment replanifiée. Une des conditions est de fournir des détails sur des produits à partir d'un recensement. Un plan stratifié à deux phases simultanées est utilisé. Des détails limités sur des produits sont recueillis dans le premier échantillon de la phase initiale où davantage de détails des produits sont rassemblés pour le sous-échantillon. La deuxième phase a détaillé l'information des produits rassemblée et est combinée avec les données restreintes des produits obtenues dans la première phase et l'information financière auxiliaire est disponibles pour la population entière pour dériver des modèles de régression. Les modèles sont employés pour produire des évaluations des produits pour les unités non échantillonnées afin de répondre à l'exigence d'un recensement. Dans cet article nous discutons du procédé utilisé pour l'estimation de la variance pour ce plan particulier à deux phases où nous employons les données échantillonnées pour assigner aléatoirement des produits à chacune des unités non échantillonnées. Nous comparons l'efficacité d'un tel arrangement à l'imputation de donneurs pour chaque unité non échantillonnée, par l'intermédiaire d'une simulation de Monte Carlo. Certains des résultats, avec des recommandations pour la mise en place de la stratégie, seront discutés.

Lundi 11 Juin, 4h30 – 4h50, AQ 4150

Estimation De La Variance Lors D'Imputation

Charlie ARCARO and Wesley YUNG (Statistics Canada)

Le problème de données manquantes ou incomplètes lors d'études est un problème commun rencontré par les agences statistiques. Dans beaucoup d'études conduites par Statistique Canada, les données manquantes sont remplacées (ou imputées) par des valeurs dérivées des données actuelles ou passées. Tandis que les méthodes d'imputation peuvent produire des estimateurs sans biais, les estimateurs conventionnels pour la variance peuvent mener à la sous-estimation de la vraie variance si les valeurs imputées sont traitées en tant que valeurs vraies. Plusieurs méthodes ont été proposées pour calculer la variance due à l'imputation. Nous étudions deux d'entre elles; l'une basée sur une version modifiée du "jackknife" et l'autre basée sur un approche basée sur un modèle. Nous comparons les propriétés de ces deux méthodes en utilisant des arguments théoriques et empiriques.

Contributed Session: Optimal Designs

Lundi 11 Juin, 3h30 – 3h45, AQ 4140

Un Plan De Phase II À Deux Niveaux Avec Point D'Arrêt Multinomial

Alex SMITH, Katherine S. PANAGEAS, Mithat GONEN, Paul B. CHAPMAN (Memorial Sloan-Kettering Cancer Center)

Des essais cliniques de phase II en oncologie sont exécutées pour évaluer l'efficacité thérapeutique d'un nouveau régime de traitement. Une mesure commune d'efficacité pour ces épreuves est la proportion de patients qui obtiennent une réponse mesurée par rétrécissement des tumeurs. C'est une technique courante pour classifier cette réponse dans les catégories suivantes: (1) réponse complète (CR); (2) réponse partielle (PR); (3) la maladie est stable (SD); et (4) progression de la maladie (PD). La réponse de la tumeur est alors traitée comme variable binaire par laquelle les patients qui réalisent un CR ou un PR sont considérés comme des répondants et tous les autres des non-répondants. Un plan à deux phases qui tient compte de l'arrêt hâtif de l'épreuve si les expositions au traitement ont peu d'efficacité tels que le plan de Gehan ou de Simon qui donnent un poids égal à un CR et à un PR. Cependant, un CR est défini en tant que disparition complète de la tumeur et est plus propice à un résultat avantageux de survie. Nous discutons qu'un CR devrait être plus pondéré et par conséquent, nous proposons une conception à deux phases avec un point final à plusieurs niveaux, c'est-à-dire CR, PR, ou un non répondant. Ce plan d'expérience est une extension du plan optimal à deux phases de Simon et est basé sur un modèle polynomial. Des comparaisons sont faites avec le plan optimal à deux phases de Simon et de nouveaux plans avec des groupements de paramètres communs sont tabulés.

Lundi 11 Juin, 3h45 – 4h00, AQ 4140

Construction De Plans Optimaux Avec Contraintes

Saumendranath MANDAL, K.C. CARRIERE (University of Alberta)

Il y a diverses classes de problèmes d'optimisation dans lesquels nous souhaitons trouver des distributions de probabilités optimales. Les plans de régression optimaux en sont un exemple. Dans ce travail, nous nous concentrons sur le problème de trouver un plan contraint approximatif maximisant un critère sous contraintes. Les contraintes sont les covariances (ou les corrélations) des estimateurs des fonctions paramétriques d'intérêt nulles. Dans beaucoup de plans de régression, on désire estimer des paramètres indépendamment des autres, en faisant des covariances ou des corrélations entre les estimateurs appropriés des paramètres aussi petites que possible. Nous abordons ce problème en formulant au départ le lagrangien avec deux contraintes, mais en retirant les paramètres de Lagrange par substitution, en transformant le problème d'optimisation sous contraintes à celui de maximiser trois fonctions de plans pondérés simultanément. Ils ont un maximum commun de zéro qui est simultanément atteint aux plans maximaux pondérés sous contraintes. Dans le cas d'une contrainte, l'approche se réduit à un problème de maximisation de deux fonctions simultanément. Pour trouver le plan optimal, nous considérons une classe d'algorithmes multiplicatifs, indexée par une fonction $f(\cdot)$, satisfaisant certaines conditions (positive et strictement croissante) et pouvant dépendre d'un ou de plusieurs paramètres libres. Ces itérations satisfont d'une manière ordonnée les contraintes du problème. Nous considérons également des techniques pour améliorer les taux de convergence avec quelques choix appropriés de la fonction $f(\cdot)$. L'approche peut être étendue au cas de plus de deux contraintes.

Lundi 11 Juin, 4h00 – 4h15, AQ 4140

Plan D'Expérience Pour Réponses Suivant Une Binomiale Négative Basé Sur L'ANCOVA

Wei WEI and R.J.COOK (University of Waterloo)

Des plans randomisées sont plus fréquemment conçues en se basant sur des données de comptage. La variabilité entre les patients dans les taux d'occurrence signifie que les modèles mixtes, tels que le modèle binomiale négative, conviennent à l'analyse. Quand les comptages de base sont disponibles, les analyses devraient être fonction de ces comptes menant à l'analogue des données de comptage de l'ancova. Nous élaborons des critères de plan d'expérience pour de telles configurations, aussi bien que pour des configurations où des patients sont choisis pour être inclus basé sur les comptes de base. La sensibilité des critères du plan d'expérience pour modéliser la spécification erronée est également étudiée. Nous contrastons nos méthodes de plan d'expérience avec celles de McMahon et al (1994).

Lundi 11 Juin, 4h15 – 4h30, AQ 4140

Un Plan À Deux Niveaux Pour Le Scan Du Génome

David VERBEL, Jaya M. SATAGOPAN, E.S. VENKATRAMAN, Colin B. BEGG (Memorial Sloan-Kettering Cancer Center)

Le but de cet article est de décrire un plan à deux niveaux qui maximise la puissance à détecter des associations gène-maladie en utilisant la lecture de génome avec comme contrainte principale le coût total du plan, représentée par le nombre total d'évaluations de gène plutôt que le nombre total de patients. Dans le premier niveau, tous les gènes d'intérêt sont évalués sur un sous-ensemble de patients. Les gènes les plus prometteurs sont alors évalués sur les sujets supplémentaires dans le deuxième niveau. Ceci éliminera le gaspillage des ressources sur des gènes peu susceptibles d'être associé à la maladie basée sur les résultats du premier niveau. Nous considérons le cas où les gènes sont corrélés et le cas où les gènes sont indépendants. En utilisant des résultats de simulations il est montré que, en général, lorsque les gènes sont indépendants ou quand la corrélation est petite, l'examen de tous les gènes sur approximativement 25% des patients et ensuite l'évaluation de 10% des gènes les plus prometteurs sur les patients restants fournit la puissance presque-optimale pour un large intervalle de configurations paramétriques.

Lundi 11 Juin, 4h30 – 4h45, AQ 4140

Sur La Modélisation Des Excédents De L'Écoulement Des Rivières Au-Dessus Des Seuils Par La Distribution De Pareto Généralisée (GPD)

Vartan CHOULAKIAN (Université de Moncton), M. A. Stephens (Simon Fraser University)

La distribution de Pareto généralisée est employée en hydrologie pour modéliser les excédents dans le contrôle des inondations; les excédents sont les niveaux des rivières au-dessus d'un seuil T , et l'adaptation du GPD a été suggéré comme méthode pour décider de T . À cette fin, on doit pouvoir examiner le GPD. Des tests seront donnés et la méthode sera étudiée pour les excédents de 238 rivières au Canada.

Lundi 11 Juin, 4h45 – 5h00, AQ 4140

Analyse D'Événements Historiques Et Échantillonnage Ignorable
Christian BOUDREAU (University of Waterloo)

Une quantité considérable de données d'analyse d'événements historiques est obtenue par des études de panel longitudinales, ou des études de cohorte. Le but est de comprendre les différents événements que les individus subissent dans le temps. Les exemples de tels événements incluent le mariage, le divorce, la fertilité, le chômage, etc. L'analyse d'événements historiques traite de la forme et l'analyse de tels processus. Notez que l'analyse de survie est un cas spécial d'analyse d'événements historiques. Cependant, les études longitudinales comportent généralement l'utilisation de designs complexes d'enquêtes telles que l'échantillonnage par grappe et par strates. Ces dispositifs d'études complexes compliquent l'analyse statistique et seront le sujet de cette présentation. En particulier nous explorons le cas des plans d'échantillonnage ignorable. Avec de tels plans il est possible de simplifier l'inférence statistique en portant des analyses non pondérées. Cependant, la dépendance entre les observations obtenue par la technique d'échantillonnage par grappes ou autres doit être prise en considération lors du calcul de la variance des quantités impliquant des événements historiques. Ceci est réalisé en utilisant la théorie des équations d'estimation.

Measurement Error

Mardi 12 Juin, 8h30 – 9h00, Halpern

Extended Partial Likelihood Methods For Failure Time Models, With Measurement Error In A Time-Varying Covariate

Donna SPIEGELMAN and R. LOGAN (Harvard School of Public Health)

Occupational and environmental epidemiologists often assess prospective associations of cumulative exposure variables with chronic disease outcomes. These variables are usually measured with moderate to considerable error. An empirically verified model for the error in the point exposures at single instants of time is identified from validation data, and the model for the measurement error in the cumulative exposure variable is derived from this model for the point exposures. An extended partial likelihood is then constructed and maximized, and estimates are obtained which fully account for the measurement error process, including uncertainty about it, assuming proportional hazards for the perfectly measured data. Identifiability, consistency and asymptotic normality of the estimator of the parameter of interest are considered for several common study designs. This method is compared to other methods currently proposed for addressing related problems, models and data structures. A detailed example illustrates these methods, examining mortality from lung cancer and cumulative exposure to radon progeny among underground miners in New Mexico, USA. A gamma measurement error model fit the annual individual exposure data well. For a 500 WLM increase in cumulative radon exposure, the estimated lung cancer mortality rate ratios, uncorrected and corrected for exposure measurement error, were 5.8 (95% confidence interval (CI) 3.0-11.1) and 16.7 (95% CI 5.2, 53.9), respectively.

Mardi 12 Juin, 9h00 – 9h30, Halpern

Régression D'Erreurs De Mesures Et Maximum De Vraisemblance Non Paramétrique

Dan SCHAFFER (Oregon State University)

En m'adressant aux statisticiens et biostatisticiens familiers à ce problème, je passerai en revue les problèmes statistiques associés à l'imprécision des variables explicatives, et ma perception de l'état présent des méthodes statistiques utiles. De plus, je discuterai de la méthodologie liée au maximum de vraisemblance non paramétrique pour divers types de régressions ayant une erreur de mesure prédite. La partie critique de l'analyse pour plusieurs applications épidémiologiques est la modélisation des incertitudes reliées à la dose ou l'exposition. Nous discuterons de ce dernier point, et donnerons aussi une liste de contrôle pour les modélisations classiques et de Berkson.

Mardi 12 Juin, 9h30 – 10h00, Halpern

Analyse De Cas-Témoins Avec Connaissances Partielles Des Probabilités De Mauvais Classement De L'Exposition

Paul GUSTAFSON (University of British Columbia)

Considérons une analyse cas-témoin avec une variable d'exposition dichotomique sujette à un mauvais classement. Si les probabilités de classification sont connues, des méthodes sont disponibles pour ajuster les estimés des rapports de cotes par rapport au mauvais classement. Nous étudions le scénario réaliste où un jugement raisonnable, mais pas nécessairement une valeur exacte, est fait par rapport aux probabilités de classification. Si l'analyse procède en traitant simplement les jugements comme étant des valeurs exactes, de petites différences entre les jugements et les vraies probabilités peuvent affecter sérieusement les estimés des rapports de cotes. Nous démontrons que ce problème est atténué par une analyse bayésienne qui incorpore les incertitudes par rapport aux probabilités de classification comme informations a priori.

Survey Methods Contributed Session: Survey Applications

Mardi 12 Juin, 8h30 – 8h50, AQ 4150

La Nouvelle Conception De L'Enquête Mensuelle Du Commerce En Gros Et Au Détail Des Statistique Canada

Hélène BÉRARD (Statistics Canada)

L'Enquête mensuelle du commerce de gros et de détail (EMCGD) est une enquête très importante menée par Statistique Canada. L'EMCGD produit des estimations sur les ventes mensuelles et les stocks pour divers domaines d'intérêt tels que les provinces et les secteurs industriels. Les estimations mensuelles sont utilisées dans le calcul du Produit Intérieur Brut (PIB). La tendance des ventes représente un indicateur économique majeur. Les estimations mensuelles du commerce de détail sont utilisées pour le calcul des paiements de péréquation (transfert de fonds du fédéral vers les provinces). L'EMCGD constitue de plus la base de sondage de l'Enquête trimestrielle sur les marchandises vendues au détail (ETMVD). L'enquête courante recourt à un plan de sondage stratifié élaboré en 1988. Plusieurs améliorations ont été apportées à l'enquête au fil des ans afin d'améliorer et de maintenir la qualité des données, toutefois un remaniement s'imposait. Un remaniement est nécessaire afin de produire les estimations selon le Système de Classification des Industries de l'Amérique du Nord (SCIAN), et de bénéficier de la disponibilité des données administratives du programme de la taxe sur les produits et services. Le remaniement vise aussi à réduire les coûts d'enquêtes et le fardeau de réponse, à mettre à jour les systèmes informatiques et à harmoniser les concepts utilisés avec les enquêtes annuelles nouvellement remaniées. Cette communication présente les défis associés au remaniement d'une enquête mensuelle auprès des détaillants et des grossistes. Je discuterai du plan de sondage, de la stratification de la population à l'aide de données provenant de fichiers administratifs, de la sélection de l'échantillon ainsi que des mécanismes envisagés pour la mise à jour de l'échantillon. Je discuterai aussi des contraintes additionnelles reliées à l'utilisation de la base de sondage et des données de l'EMCGD par l'ETMVD.

Mardi 12 Juin, 8h50 – 9h10, AQ 4150

Défis Dans La Quantification Du Commerce Électronique

Patricia WHITRIDGE (Statistics Canada)

L'intérêt pour le commerce électronique avait augmenté au cours des dernières années. Statistique Canada a entrepris une étude pour mesurer la quantité de ventes ayant lieu sur internet pendant l'année 1999. On a décidé de tirer profit du plan d'échantillonnage d'une grosse enquête sur l'économie déjà existante. Un questionnaire a été développé avec des questions au sujet de l'utilisation de la technologie dans les entreprises. L'information a été rassemblée sur la disponibilité des ordinateurs, sur l'utilisation du courriel et sur l'accès à internet. On a également demandé aux entreprises si elles avaient une page W3, si elle pouvait être utilisée pour des ventes, et la quantité de ces ventes. Les données ont été disponibles à l'été 2000. Maintenant, plusieurs possibilités sont discutées au sujet des méthodes les plus pertinentes à utiliser pour estimer toutes les ventes réalisées par internet. Par exemple, devrions-nous poser des questions sur les pages W3 aux établissements, ou plutôt aux sièges sociaux? Une étude générale des entreprises est-elle la meilleure méthode? Pourrions-nous tirer un échantillon d'entreprises et rechercher alors nous-même les pages W3? La présentation commencera par de l'information de base au sujet de l'étude de 1999, incluant ses défis. Puis, les plans actuels pour l'étude de 2000, qui est en cours, seront discutés. Quelques possibilités qui se sont déjà présentées seront identifiées, et différents scénarios d'avenir à plus long terme seront examinés. La présentation se conclura par quelques observations générales sur l'expérience de Statistique Canada pour mesurer le commerce électronique.

Mardi 12 Juin, 9h10 – 9h30, AQ 4150

Plan Échantillonnal De L'Enquête Sur La Santé Mentale Canadienne
Yves BÉLAND (Statistics Canada)

En tant qu'élément de la stratégie bisannuelle de l'enquête sur la santé de la Communauté Canadienne (CCHS), la composante de l'étude provinciale du premier cycle du CCHS se concentrera sur différents aspects de la santé mentale et du bien-être des Canadiens vivant dans des logements privés. D'ailleurs, cette nouvelle étude rassemblera des données sur des facteurs positifs et négatifs affectant la santé mentale, l'utilisation des services de santé mentale aussi bien que des données sur des impacts sociaux et les coûts associés à la santé mentale. Tout ceci sera ajouté à la collecte d'une série de caractéristiques sociales et démographiques. L'échantillon prévu pour cette étude se composera de 30 000 répondants choisis parmi un cadre de probabilité de zone sous un plan d'échantillonnage en grappes à plusieurs niveaux où le logement est l'unité échantillonnale finale. Une personne âgée de 15 ans ou plus sera choisie au hasard dans les ménages échantillonnés. Les plans sont de suréchantillonner les personnes dans la catégorie d'âge 15-24 en augmentant leur probabilité de sélection dans le but d'améliorer leur représentation dans l'échantillon. La collecte de données commencera en janvier 2002 et s'étendra sur plus de 12 mois pour éliminer des effets saisonniers possibles sur certaines caractéristiques principales. Toutes les entrevues seront conduites en tête-à-tête en utilisant une approche assistée par ordinateur. (cependant certaines questions sensibles tel que ceux au sujet des pensées suicidaires seront posées à l'aide d'un questionnaire écrit auto-administré.) On s'attend à ce que la longueur moyenne d'une entrevue soit de 90 minutes. Cet article décrira plusieurs aspects clés du plan d'échantillonnage de cette nouvelle et passionnante étude de même que quelques défis méthodologiques pour le test et la validation du questionnaire. En parallèle, une brève vue d'ensemble du plan d'échantillonnage d'une étude spéciale sur la santé mentale et le bien-être des membres des forces armées canadiennes sera donnée.

Mardi 12 Juin, 9h30 – 9h50, AQ 4150

Quoi De Neuf Pour Les Tableaux De Mortalité 1995-1997 Pour Le Canada

Brad W. THOMAS, Doreen DUCHESNE (STC), Marek WYSOCKI (STC), Robert BOURBEAU (University of Montréal)
(Statistics Canada)

Une table de mortalité présente l'espérance de l'événement mortalité pour une cohorte fictive de 100 000 personnes nées au même moment, et sujettes aux taux de mortalité connus par âge et par sexe pour la population courante. Les dernières tables de mortalité détaillées produites pour le Canada et les provinces couvraient la période 1990-1992 (Statistique Canada numéro 84-537 au cat.). Ces tables étaient basées sur le Recensement de la population (REPOP) de 1991 et sur les données de mortalité de 1990 à 1992. La disponibilité des données du REPOP de 1996 et de mortalité de 1997 nous permet de produire les tables de mortalité détaillées pour la période 1995-1997. La publication de Statistique Canada citée précédemment décrit en détails la méthodologie utilisée pour créer les tables de 1990-1992, auxquelles on a fait des améliorations pour obtenir celles de 1995-1997. Le présentation portera principalement sur les améliorations les plus importantes, notamment l'utilisation d'un modèle pour estimer la probabilité de décès pour les 85 ans et plus (Coale et Kisker, 1990), ainsi qu'une nouvelle méthodologie pour estimer ces probabilités pour les âges 0 à 4 ans. On utilise les estimations de population de Statistique Canada au 1er janvier pour les années 1995-1998. On présentera également une nouvelle méthode pour assurer une tendance décroissante monotone dans les probabilités à l'échelle provinciale.

WNAR Student Presentations

Mardi 12 Juin, 8h30 – 8h50, AQ 4130

Détection De L'Hétérogénéité Génétique Au Sein De Familles Avec Une Parenté Infectée

Lucia MIREA and Shelley B. BULL (Samuel Lunenfeld Research Institute, University of Toronto)

Les études actuelles de liaison visent à identifier les sites chromosomiques qui peuvent héberger des gènes susceptibles pour des maladies complexes caractérisées par la variabilité dans les symptômes, la prévalence, et/ou le mode de la transmission. Les méthodes statistiques évaluent si la distribution des allèles partagées observées, mesurant la similitude génétique parmi les membres affectés d'une famille, est différente de celle prévue par le hasard. Ces méthodes souffrent d'une perte de puissance statistique lorsque l'hétérogénéité systématique est présente parmi la distribution des allèles partagées des familles échantillonnées. Pour détecter l'hétérogénéité, à une position chromosomique donnée, nous proposons deux approches qui utilisent des covariables binaires au niveau familial. Le sous-diagnostic clinique, l'appartenance ethnique et l'âge au début de la maladie sont des exemples de covariables qui peuvent indiquer des différences génétiques à l'intérieur de familles. La première approche est un test du rapport de vraisemblance (LRT), basé sur une extension de la vraisemblance de liaison de Kong et de Cox (1997, *Am J Hum Genet* 61:1179-1188) qui incorpore les effets d'une covariable binaire. La méthode du LRT compare le modèle aux covariables du modèle sous l'hypothèse nulle n'assumant aucun effet de covariable. Dans la deuxième approche (PERM), des méthodes de permutations sont appliquées pour tester les différences dans les score de liaison moyens entre les sous-groupes définis par les covariables. Le LRT implique le calcul des estimateurs du maximum de vraisemblance sous certaines contraintes et peut ne pas avoir la distribution asymptotique prévue. En revanche, le test PERM ne fait aucune hypothèse sur la distribution et ne comporte pas d'évaluations explicites des paramètres. L'effort de calcul des deux tests est similaire. Pour comparer les propriétés de ces méthodes, nous avons conçu et conduit une étude de simulations pour estimer l'erreur du type I et la puissance dans des études de familles avec des enfants de mêmes parents et de cousin affectés. Plusieurs scénarios ont été évalués comprenant les modèles dominants et récessifs de la transmission avec les divers niveaux (20%, 40% et 50%) de l'hétérogénéité du locus.

Mardi 12 Juin, 9h00 – 9h30, AQ 4130

Incertitude Des Résultats Pour La Modélisation Des Risques Proportionnels À Temps Discret

Amalia MEIER and Barbra RICHARDSON (University of Washington)

L'estimation du temps d'échec a été typiquement fait quand on observe l'échec avec certitude. Cependant, dans le contexte où les résultats du test sont des faux négatifs et des faux positifs, le temps de l'analyse d'événement est également d'intérêt. On désire une procédure qui s'occupe de la perte d'information due à l'imperfection des résultats au niveau de la sensibilité et de la spécificité. Richardson et Hughes (2000) fournissent un algorithme pour l'estimation du produit des limites pour des données à temps discrets avec des résultats incertains. Grâce à leur travail, on développe une méthode qui permet l'estimation de la courbe de survie et l'estimation des effets de covariables. Les différences du temps de survie entre les groupes de traitements, sous l'hypothèse des risques proportionnels, sont d'un intérêt particulier. Il est montré que le plus grand biais pour le risque et pour l'estimation d'effet de covariable provient du fait qu'on ne tient pas compte des tests faussement positifs. Cependant, les simulations prouvent que la spécification correcte de ces tests à faux taux permettent (presque) l'estimation non biaisée du risque considéré au tout début et des effets de covariables. Une comparaison graphique aux modèles de Cox est faite, évaluant la robustesse des estimateurs en ce qui concerne la spécification erronée de ses faux taux de test. La méthode est appliquée surtout pour la recherche dans les pays en voie de développement, où l'exactitude des tests peut être limitée par le coût et les difficultés technologiques.

Mardi 12 Juin, 9h30 – 10h30, AQ 4130

Models For Zero-Inflated Count Data Using The Neyman Type A Distribution

Melissa DOBBIE and A.H. WELSH (Australian National University)

In animal abundance studies, where, for example, the estimated abundance of a rare animal may be of primary importance, often a large proportion of the data is zero, leading to an excess of zeros in relation to a Poisson or negative binomial distribution. We refer to this phenomenon as zero-inflation. Ignoring zero-inflation implies that the underlying distributional assumptions will not be met. This can affect the results of an analysis, and hence lead to incorrect conclusions concerning the data. Ridout et al (Proc. XIXth IBC Cape Town, 1998, 179-190) review and compare several approaches to modeling count data with an excess of zeros in relation to the Poisson distribution. One approach models such data by a mixture of Poisson distributions; the most well-known and commonly used mixture is the negative binomial distribution. However, zero-inflated count data typically exhibit either a half-mode (at zero) or one-and-a-half modes (at zero and at a count greater than one) implying that the unimodal negative binomial distribution may be unsuitable. The Neyman type A distribution (Neyman, Ann. Math. Stat., 1939, 10, 35-57) is also generated by a mixture of Poisson distributions but is flexible enough to model multimodal data (Barton, Trab. Estad., 1957, 8, 13-22). We explore the possibility of modeling zero-inflated count data using the Neyman type A distribution. Three parameterizations of the distribution are extended to allow their parameters to depend on covariates. To illustrate the methodology, we develop models for counts of Leadbeater's possum (*Gymnobelideus leadbeateri*) of which 63% are zeros. Using the method of Cox (J. Royal Stat. Soc. B, 1962, 24, 406-424) to test non-nested hypotheses, we then formally compare the Neyman type A models and also compare them to a model from a competing family.

Small Area Estimation

Mardi 12 Juin, 10h30 – 11h00, AQ 4150

Second Order Asymptotics in Small Area Estimation: Accurate

David SMITH, G.S. DATTA and J.N.K. RAO (University of Georgia)

A simple area level linear mixed model, popularly known as the Fay-Herriot model, is widely used to produce model-based small area estimators; in particular, empirical best linear unbiased prediction (EBLUP) estimators of small area means. The EBLUP estimators depend on the estimator of the between area variance component. Fay and Herriot proposed a simple iterative method to estimate this variance component, not requiring normality. In this paper we obtain a second order approximation to the MSE of EBLUP estimators for the Fay-Herriot iterative method and an estimator of MSE also correct to second order, and compare our results to previous second order approximations based on the method of moments, maximum likelihood (ML) and residual maximum likelihood (REML). We also obtain a correction to normal theory confidence intervals, based on EBLUP estimator and its estimator of MSE, to achieve the target coverage probability asymptotically as the number of small areas increases.

Mardi 12 Juin, 11h00 – 11h30, AQ 4150

Small Area Estimation using Multiple Surveys

Narasimha PRASAD and Sharon LOHR (University of Alberta)

In this paper, we provide a general method for small area estimation when information is derived from multiple surveys. Fuller and Harter (1987) and Battese, Harter and Fuller (1988) proposed nested-error regression models for small area estimation. We adopt a multivariate approach to extend these models for combining information from several surveys, with possibly different sampling designs, in order to improve the accuracy of small area estimates. Computational algorithms are discussed, and a simulation study indicates that if quantities in the different surveys are sufficiently correlated, gains in efficiency can be achieved.

Mardi 12 Juin, 11h30 – 12h00, AQ 4150

Benchmarking of Small Area Estimators in a Bayesian Framework

Avi SINGH and Ralph FOLSOM (Research Triangle Institute)

It is often desirable in practice to benchmark the small area estimates in that they sum to the direct large area estimates. This provides robust estimates that protect against possible model mis-specification as well as possible overshrinkage. Moreover, benchmarking provides internal consistency of small area estimates with direct aggregate level estimates which are often published before the small area estimates. For random benchmarks, it is not possible to incorporate these constraints in the (Hierarchical) Bayes theory, because the benchmarks are based on the data used in the likelihood, and are not part of the prior. However, the class of estimators can be constrained to satisfy the benchmarks. To this end, we propose a generalization of the constrained Bayes theory of Ghosh (1992) in that benchmarks need not be posterior means of moments of the histogram of small area parameters. For nonrandom benchmarks, the parameter space becomes constrained, and for this purpose we propose, for the Metropolized MCMC for HB estimation, a method of construction of the proposal distribution such that benchmarks are exactly satisfied by the candidate sample realizations. Computation of MSE of optimal estimates under both random and nonrandom benchmarks is discussed. Some illustrative examples are also presented.

Likelihood

Mardi 12 Juin, 10h30 – 11h00, Halpern

Inférence Basée Sur La Vraisemblance, Information À Haute Dimension Et Manquante De Façon Aléatoire

Susan MURPHY (University of Michigan)

Nous considérons deux problèmes où l'information manquante de façon aléatoire est plausible uniquement lorsque nous conditionnons sur un grand nombre de covariables auxiliaires. Mais dans les deux problèmes l'estimateur désiré est marginal par rapport aux covariables auxiliaires. Puisque l'estimateur est marginal, il est indésirable de demander qu'un modèle conditionnel (par rapport à des variables auxiliaires) soit correct. Nous proposons une méthode qui dépend sur deux jugements, un pour chaque distribution conditionnelle. Si un des jugements est correct, nous montrons que les estimateurs sont compatibles. De plus, si les deux jugements de modélisation sont corrects, l'estimateur est efficace localement. Ce type d'estimateur fut nommé doublement robuste par J. Robins. Le premier problème concerne l'estimation d'une fonction de survie marginale dans le cas où les temps de rupture et de censure sont indépendants conditionnellement aux covariables auxiliaires. Le second problème est l'estimation du coefficient de régression dans un modèle de risques proportionnels pour la distribution du temps de rupture pour une covariable donnée (p.ex. un indicateur de traitement). Cependant, comme dans le problème précédent, les temps de rupture et de censure sont indépendants seulement lorsque conditionnels à plusieurs covariables auxiliaires additionnelles.

Ce travail à été fait conjointement avec Donglin Zeng.

Mardi 12 Juin, 11h00 – 11h30, Halpern

Modèles De Vraisemblance Pour L'Évaluation De L'Agrégation Familiale

Nan LAIRD (Harvard School of Public Health)

La première étape dans l'étude de la base génétique de la maladie consiste à démontrer que la maladie est groupée ou agrégée dans certaines familles. Les plans cas-témoins sont souvent utilisés pour amasser des données sur l'apparition de la maladie dans la famille proche des individus atteints et de ceux contrôles. Nous discuterons de certains modèles de vraisemblance pour l'analyse de données familiales tirées d'études cas-témoins et comparerons des méthodes basées sur la vraisemblance avec celles basées sur la régression logistique. Nous discuterons aussi de l'extension des approches pour une seule maladie afin d'étudier la co-agrégation de plus d'un troubles.

Mardi 12 Juin, 11h30 – 12h00, Halpern

Ajustements Au Profil De Vraisemblance

Nancy REID (University of Toronto)

Dans plusieurs modèles paramétriques des paramètres de nuisance sont introduits afin de rendre le modèle plus réaliste, mais l'intérêt principal de l'étude est l'inférence sur un petit nombre de paramètres d'intérêt. Souvent, un ajustement simple au profil de vraisemblance est suffisant pour tenir compte de ces paramètres de nuisance. Cet ajustement est généralement dérivé par un argument utilisant la théorie asymptotique d'ordre supérieur. Des travaux récents sur ce sujet seront présentés, en mettant l'accent sur les similarités entre les approches bayésienne et non bayésienne.

Statistics in Industry and Technology

Mardi 12 Juin, 10h30 – 11h00, AQ 4130

Some Simple Research Problems from Industrial Consulting

Jock MACKAY and Stefan STEINER (University of Waterloo)

Routine consulting in an industrial setting can generate many interesting research questions for a perceptive observer. These questions tend to be directly related to application or at most one step removed, and hence are usually simple to state. It is our belief that these problems are best discovered by "being there" since they often arise separately from the purpose of the immediate consultation. In this talk, we present several examples of such problems in an attempt to convince academic researchers of the value of seeing industrial processes and their control systems in action. Reality is often stranger than the best imagination.

Mardi 12 Juin, 11h00 – 11h30, AQ 4130

Statistical Investigation of Resistance Spot Welding Quality Using a Two-Stage, Sliding Level Experiment

Wei LI, Shaowei CHENG, Jack HU and Justin SHRIVER (University of Washington)

Quality variation in resistance spot welding is a major concern in industry. However, the relationship between weld quality and various process conditions, including abnormal process conditions, has not yet been systematically studied. This talk discusses this relationship using a newly developed two-stage, sliding-level experiment. In the experiment, welding current is treated as a "slid factor" and is determined based on the settings of other process variables. Engineering knowledge is applied in statistical model selection. From the analysis, it is found that abnormal process conditions, such as axial misalignment, angular misalignment, poor fitup, edge weld, and electrode wear, significantly affect weld size and thus cause large variations in weld quality. Although they may help increase the weld size, abnormal process conditions generally lead to less robustness of the process. In order to minimize the effects of the abnormal process conditions, a robust parameter design is formulated using the statistical models developed from the experimental data. The analysis suggests that high current and large electrodes should be used in the process. Developed for this study, the new experimental design and analysis procedures are generic for manufacturing processes where process variables are interdependent.

Mardi 12 Juin, 11h30 – 12h00, AQ 4130

Non-Regular Robust Parameter Design

Derek BINGHAM and William LI (University of Michigan)

Robust parameter designs performed using a combined experimental plan for the noise and control factors often provides cost-savings and increased estimation capacity for the effects of interest when compared to Taguchi's crossed arrays. Selecting the "best" fractional factorial design in this setting has received much consideration in recent years. In this work, we introduce methodology for finding optimal non-regular robust parameter designs and demonstrate that these designs often have better estimation capacity than fractional factorial robust parameter designs.

Contributed Session: Timeseries

Mardi 12 Juin, 10h30 – 10h45, AQ 4140

Autorégression Généralisée Des Rangs Signés

Jeff TERPSTRA (North Dakota State University)

Les estimateurs par les rangs signés sont connus pour réaliser un bon compromis entre la robustesse et l'efficacité lorsque les erreurs sont normales. Cependant, les estimateurs par les rangs signés sont seulement robustes pour la réponse, et sont assez sensibles aux valeurs aberrantes dans le plan d'expérience. Puisque les observations d'une série chronologique auto-régressive jouent un double rôle en tant que réponse et variable explicative, une analyse basée sur les rangs signés des modèles auto-régressifs peut ne pas être totalement robuste. Dans cet article, un estimateur pondéré des rangs signés qui se réduit à l'estimateur des rangs signés de Wilcoxon quand les poids sont constants est considéré. La généralisation est motivée à partir d'une identité algébrique sur la norme des rangs signés de Wilcoxon et elle est semblable au rapport entre l'estimateur GR et l'estimateur R de Wilcoxon de Hettmansperger et de McKean (1998). L'utilisation des poids induit une grande classe des estimateurs dont le comportement varie de manière très robuste à fortement efficace. Sous l'hypothèse d'un deuxième moment fini, des propriétés asymptotiques de linéarités sont dérivées pour ces prétendus estimateurs GSR. En outre, assumant une distribution symétrique des erreurs, les estimateurs GSR s'avèrent asymptotiquement normaux avec un taux égal à la racine carrée de n . La théorie des statistiques U avec une caractérisation de la dépendance faible qui est inhérente aux modèles stationnaires $AR(p)$ sont les outils primaires utilisés pour obtenir les résultats. Des tests d'hypothèses aussi bien que des erreurs types pour les intervalles de confiance peuvent être basés sur la théorie asymptotique. Les GSR et les estimateurs GR sont comparés par l'intermédiaire d'un exemple et une étude d'efficacité. Les résultats indiquent que les GSR et estimateurs GR sont robustes pour l'estimation d'une série chronologique auto-régressive. Cependant, pour un ensemble de poids bien choisi, l'estimateur GSR est plus efficace que l'estimateur GR. Pour finir, l'estimateur GSR a un avantage de plus par le fait que l'ordonnée à l'origine peut être estimée simultanément; à la différence de l'estimateur GR où l'ordonnée à l'origine est basée sur un estimateur d'emplacement pris de l'ensemble initial des résidus.

Mardi 12 Juin, 10h45 – 11h00, AQ 4140

L'Estimateur Du Maximum De Vraisemblance Exact Pour Les Processus Stationnaires Normaux AR(1) Et Applications Au Test Pour La Racine Unité

Ying ZHANG and A.I. McLEOD

La forme analytique de la solution pour l'estimateur exact du maximum de vraisemblance pour des processus stationnaires normaux $AR(1)$ est donnée. Ses moments asymptotiques sont dérivés par la notion de puissance. En conséquence, nous développons un algorithme pour les moments asymptotiques d'ordre supérieur de l'estimateur exact du maximum de vraisemblance. En utilisant l'estimateur exact du maximum de vraisemblance, nous proposons des tests unilatéraux pour la racine unité par la technique du modèle mal spécifié. Nous établissons leur convergence. Nous présentons également une étude de puissance de Monte Carlo comparant ces tests aux méthodes alternatives existantes. La puissance de nos tests pour la racine unité surpasse légèrement cela des propositions précédentes.

Mardi 12 Juin, 11h00 – 11h15, AQ 4140

Application De Méthodes D'Ajustement Saisonnier À Des Séries Chronologiques Difficiles: Le Cas Croate
Ante ROZGA (University of Split)

Ces dernières années nous avons été témoin d'une forte élaboration des méthodes d'ajustement saisonnier . Ceci est dû particulièrement à la demande des analystes économiques pour une meilleure qualité des données, et également pour trouver des méthodes plus justes au niveau théorique. Des méthodes basées sur des modèles telles que TRAMO/SEATS sont confrontées aux méthodes traditionnelles empiriques utilisées comme la famille X-11. Nous avons testé le comportement de TRAMO/SEATS et de X-12-ARIMA sur 470 séries chronologiques économiques croates et avons trouvé une différence substantielle. Ici nous avons concentré notre analyse sur des séries chronologiques difficiles. Les séries chronologiques économiques croates sont davantage "difficiles" comparées à celles de EU et US pour des raisons évidentes de guerre et de processus de transition. Nous avons utilisé le procédé automatique avec l'interface DEMETRA développé par Eurostat, Luxembourg. Il y avait 77,45% de toutes les séries qui ont été trouvés difficile par X-12-ARIMA comparé à 46,17% par TRAMO/SEATS. La raison du pourcentage tellement élevé de séries chronologiques difficiles est dans le niveau élevé de leur désagrégation. Si nous utilisions plus de séries chronologiques agrégées, ce pourcentage serait de beaucoup inférieur. Ceci est dû au comportement erratique des séries chronologiques non significative d'un point de vue économique. Nous avons également étudié les raisons pour que la série soit considérée comme difficile. Avec l'intervention d'un utilisateur, le nombre de séries chronologiques difficiles pourrait être réduit, en particulier en utilisant X-12-ARIMA. De façon plus générale, TRAMO/SEATS a besoin de moins d'interventions de l'utilisateur et est plus pratique pour modéliser plusieurs séries chronologiques. Pour éviter des difficultés, les séries pourraient être divisées en deux parties modélisées séparément. Dans le cas croate, c'est la période qui inclut la guerre et l'autre période après la guerre. Cette approche a eu comme conséquence une meilleure performance des deux approches. En résumé, nous avons trouvé TRAMO/SEATS plus flexible à la majorité des séries chronologiques économiques croates.

Mardi 12 Juin, 11h15 – 11h30, AQ 4140

Les Fluctuations Financières : Quelques Propriétés Des Processus Stochastiques Extrêmes
Don MCLEISH (University of Waterloo)

Les observations sur les prix des valeurs mobilières, des taux de change, des taux d'intérêts, et d'autres séries chronologiques financières, incluent habituellement non seulement un prix d'ouverture et de fermeture, mais aussi un maximum et un minimum pour la période. Supposez que nous observons des processus stochastiques $X(t)$ pendant l'intervalle $[0, T]$. Nous enregistrons le $X(0)$ prix d'ouverture, le $X(T)$ prix de fermeture, le grand H .

Mardi 12 Juin, 11h30 – 11h45, AQ 4140

Détection Des Grappes De Durée Et Adéquation Des Modèles Conditionnels Auto-Régressifs De Durée En Utilisant Les Ondelettes

Pierre DUCHESNE and Yongmiao HONG (Ecole des Hautes Etudes Commerciales)

Il y a eu récemment un intérêt croissant dans l'analyse des données financières arrivant à des intervalles irréguliers. Une classe importante de modèles est le modèle conditionnel auto-régressif de la durée (ACD) présenté par Engle et Russell (1998), et ses diverses généralisations. Ces modèles ont été employés pour modéliser les grappes de durée pour des données financières telles que les temps d'arrivées des transactions et les changements des prix. Cependant, très peu de procédures d'évaluation pour l'adéquation des modèles ACD sont disponibles dans la littérature. Un test diagnostic généralement utilisé est la statistique de Box-Pierce/Ljung-Box adaptée à l'estimation des résidus de durée normalisés, mais sa distribution asymptotique est inconnue et peut-être qu'elle doit être ajustée. Dans cette présentation nous proposons un nouveau test pour le groupement selon la durée et un test de diagnostic pour des modèles ACD à l'aide d'un estimateur de densité spectral de la durée basée sur les ondelettes . Le premier test exploite la nature unilatérale du groupement selon la durée. Un processus de ACD est positivement autocorrélé à tous les décalages, ayant pour résultat un mode spectral à fréquence zéro. En particulier, il y a un sommet spectral à la fréquence zéro lorsque le groupement selon la durée est persistant ou lorsque le groupement selon la durée est petit à chaque décalage individuel mais se reporte sur un long décalage distributionnel. Comme méthode commune de décomposition de temps-fréquence, les ondelettes peuvent efficacement capturer les sommets spectraux et ils devraient être puissants. Notre deuxième test contrôle l'adéquation d'un modèle de ACD en utilisant une densité spectrale en ondelettes pour l'estimation des résidus de durée normalisés sur la fréquence entière. Le test de diagnostic proposé a une propriété asymptotique pratique, car elle ne dépend pas du paramètre de nuisance-- l'incertitude d'estimation de paramètre n'a aucun impact sur la distribution asymptotique de la statistique de test. D'ailleurs, il peut contrôler un éventail de solutions alternatives et il est particulièrement puissant lorsque le spectre pour l'estimation des résidus de durée est non lisse, ce qui peut arriver en raison du groupement selon la durée persistant négligeable, de l'effet saisonnier, de l'effet calendrier et des cycles économiques. Nous proposons et justifions des méthodes appropriées dépendantes des données pour choisir de meilleures échelles -- le paramètre de lissage dans l'estimation par ondelettes. Ceci rend les méthodes proposées totalement opérationnelles en pratique. Nous présentons une étude de simulation, illustrant les mérites des procédures proposées.

Session in Honour of J.N.K. Rao's Retirement

Mardi 12 Juin, 1h30 – 2h00, Images

Jon Rao: Survol De Ses Contributions En Recherche

David BELLHOUSE (University of Western Ontario)

Jon Rao a publié son premier article de recherche en échantillonnage à l'âge de vingt ans, quatre ans avant d'obtenir son doctorat en 1961. Rapidement Jon est devenu un des chercheurs de pointe en échantillonnage, et il reste encore aujourd'hui un des meneurs dans ce domaine. Pendant ces quarante années, la recherche en échantillonnage a considérablement changé et Jon en a souvent été l'initiateur. Les publications de recherche de Jon seront examinées en relation à l'évolution de la théorie et des méthodes d'échantillonnage modernes.

Mardi 12 Juin, 2h00 – 2h30, Images

Méthodes De Rééchantillonnage Pour Sondages Complexes : Un Survol

Randy SITTER and Jeff WU (Simon Fraser University)

Cette présentation essaiera de passer en revue quelques-uns des développements en méthodes de rééchantillonnage dans le contexte de sondages. Ces méthodes comprennent la méthode du jackknife, les répliques répétées balancées et le bootstrap. L'accent sera mis sur l'estimation de la variance. Des structures d'échantillonnage complexes et des problématiques telles que la stratification, l'échantillonnage en grappe à plusieurs niveaux, l'échantillonnage sans remise, les données manquantes et l'imputation, etc., seront considérées et l'impact qu'elles ont sur la structure et la complexité des méthodes de rééchantillonnage sera discuté. Nous finirons par une discussion sur les plus récents développements, les limites fondamentales et les défis futur.

Mardi 12 Juin, 2h30 – 3h00, Images

Contributions Faites Par J.N.K. Rao À La Méthodologie De Sondage À Statistiques Canada

Mike HIDIROGLOU (Statistics Canada)

J.N.K. Rao est consultant en échantillonnage à Statistiques Canada depuis 1973. Ce papier met l'accent sur les contributions directes et indirectes apportées par Jon en recherche et sur le programme de méthodologie appliquée à Statistiques Canada. Ses contributions les plus directes comprennent plusieurs articles qu'il a écrits en collaboration avec d'autres méthodologistes à Statistiques Canada, sa longue et active participation sur le Comité conseil sur les méthodes statistiques à Statistiques Canada, son travail comme éditeur associé de la revue Méthodes d'enquête, et l'application de ses développements théoriques en échantillonnage de sondage dans plusieurs sondages menés par Statistiques Canada. Ses contributions indirectes comprennent le simple fait que Statistiques Canada ait eu la chance d'avoir un consultant de l'envergure de Jon. Il a donné de précieux conseils statistiques à plusieurs méthodologistes de Statistiques Canada, et aussi un grand éventail de problèmes de sondage. Son encouragement continu à la recherche, et ses interactions avec les méthodologistes plus orientés vers la recherche ont clairement été des facteurs qui ont contribué au leadership de Statistiques Canada en méthodologie d'enquête. En dernier, mais non les moindres, sont l'attention que Jon porte aux gens et son amour pour les méthodes d'enquêtes.

Bioinformatics Tutorial I

Mardi 12 Juin, 1h30 – 3h00, Halpern

Inférence En Biologie Moléculaire À Haut Débit

Larry HUNTER (University of Colorado)

La biomédecine est entrée dans une nouvelle ère de production de données à haut débit. La taille des bases de données de séquences macromoléculaires double environ tous les 18 mois, et elles contiennent maintenant plus de 7 millions de séquences représentant plus de 9 milliards de nucléotides. Les puces géniques permettent l'analyse simultanée du niveau d'expression de milliers de gènes à la fois. Des technologies connexes permettent l'identification de millions de mutations ponctuelles chez certains individus, le criblage simultané de milliers de composés pour une affinité d'adhésion similaire à celle d'un médicament, et aussi la détermination relativement rapide de la structure tri-dimensionnelle des macromolécules. Les défis naissants de l'analyse de cet assaut de données extrêmement intéressantes ont donné jour à la bioinformatique. Inférer des faits ayant un sens biologique pertinent de ces jeux de données massifs est un aspect clé de la bioinformatique, et prend source dans des techniques statistiques, de reconnaissance des formes, et d'apprentissage machine. Dans ce survol, plusieurs problèmes représentatifs et techniques de solution seront présentés pour un public ayant de bonnes connaissances en statistiques. Les techniques présentées comprennent des modèles markoviens cachés, des machines à vecteur-support et des réseaux bayésiens.

Discrete Probability and Combinatorics

Mardi 12 Juin, 1h30 – 2h15, AQ 4150

Décroissance Exponentielle Et Propriétés De Mélange Pour ???Lattice System in Finite Volumes

Ken ALEXANDER (University of Southern California)

Pour les systèmes « spin » (tels les modèles Ising ou Potts) et les modèles dépendants de percolation (tel le modèle FK), nous examinons quelques questions reliées au degré auquel les influences d'un événement survenant dans une région de volume fini peuvent se propager et altérer les probabilités d'événements survenant dans des régions distantes du même volume. Ces questions comprennent les suivantes : si le modèle démontre une décroissance exponentielle des corrélations ou des connectivités dans un volume infini, y a-t-il une décroissance exponentielle analogue dans les volumes finis qui est uniforme dans le choix de volume et de conditions aux frontières? Comment une telle décroissance exponentielle uniforme de la fonction à deux points (connectivité ou corrélation) est-elle reliée aux versions uniformes des propriétés de mélange? Les propriétés de mélange pour des modèles en treillis (mélange faible et fort) sont habituellement formulées en termes de l'influence maximale qu'un événement survenant dans une région Γ peut avoir sur un autre événement survenant dans une région distante Δ , ou Γ et Δ sont déterministes. Jusqu'à quel point ces propriétés de mélange sont-elles valides lorsque Γ et Δ sont aléatoires?

Mardi 12 Juin, 2h15 – 3h00, AQ 4150

Fonctions Critiques À Deux Points Pour Marche S'Auto-Évitant, Percolation, Arbres En Treillis Et Animaux En Treillis

Gordon SLADE (University of British Columbia)

Il est bien connu que la fonction de marche aléatoire de Green décroît comme $|x|^{-2-d}$ en d dimensions ($d > 2$). Cette présentation discutera d'une preuve récente de cette même décroissance de loi de puissance pour les fonctions critiques à deux points pour marche s'auto-évitant, percolation, arbres en treillis et animaux en treillis, pour les modèles assez étendus en $d > 4$ (marche s'auto-évitant), $d > 6$ (percolation), et $d > 8$ dimensions (arbres et animaux.) Cette preuve utilise une nouvelle approche unifiée du développement en lacet, et elle est un travail conjoint avec Takashi Hara et Remco van der Hofstad.

IMS Special Invited Session II

Mardi 12 Juin, 3h30 – 5h00, Images

Interpreting Genomic Polymorphism Data: What History Has To Tell Simon

TAVARE (University of Southern California)

Technological advances in molecular biology have made it possible to survey genome-wide DNA sequence variation in natural populations. These data include restriction fragment length polymorphisms, micro satellite repeats, single nucleotide polymorphisms (SNPs) and complete DNA sequences of particular loci. The analysis and interpretation of the patterns of variation seen in such data is complicated by the fact that the sampled chromosomes share a common ancestry, thus making the data highly dependent. To make matters worse, the nature of this common ancestry is not known precisely and therefore needs to be modeled. Since the early 80s, population geneticists have used the coalescent as a stochastic description of the ancestry of a sample of chromosomes, and there is now an extensive literature on inference and estimation for such processes. In this talk I will give an overview of coalescent methods, touching on a number of applications including inference about demographic history, selective sweeps, the age of mutations, and the hunt for disease genes using linkage disequilibrium mapping. A number of open problems will be discussed.

Bioinformatics Tutorial II

Mardi 12 Juin, 3h30 – 5h00, Halpern

Prévision De Réseaux Métaboliques

Imran SHAH (University of Colorado)

La biologie moléculaire à haut débit fournit une quantité sans précédent de données ayant une influence indirecte sur notre compréhension des entités et processus biologiques. Ces données peuvent nous aider à décomposer les réseaux entrelacés d'interactions biochimiques connus sous le nom de métabolisme. La compréhension du métabolisme d'organismes vivants aura un gros impact sur la recherche en sciences de la vie. Les approches bioinformatiques visant à approcher ce problème font face à beaucoup de défis informatiques. Cette présentation décrira une stratégie bioinformatique pour la prévision de réseaux métaboliques mettant en évidence les défis statistiques et informatiques de cette approche.

Canadian Initiative on Social Statistics

Mardi 12 Juin, 3h30 – 5h00, AQ 4150

Perspectives on the Canadian Initiative on Social Statistics

Doug NORRIS (Statistics Canada), Paul BERNARD (Université de Montréal), Mary THOMPSON (University of Waterloo)

The Canadian Initiative on Social Statistics (CISS) is a cooperative project of SSHRC and Statistics Canada. Its aim is to promote research and training in the application of social statistics. Once CISS has implemented the recommendations of a joint National Task Force of leading Canadian researchers and statisticians, it is anticipated that there will be a great increase in Canada's capacity to undertake major, innovative, social research based on quantitative evidence. The purpose of this SSC session is to give various perspectives on the kinds of research the CISS will help to bring about. The speakers will tackle some provocative topics such as (i) what social statistics are and are not; and (ii) what the cooperation should be between statisticians and social scientists involved in social statistics research.

SSC Presidential Invited Address

Mercredi 13 Juin, 8h30 – 10h00, Images

The Geometry Of Random Images In Astrophysics And Brain Mapping

Keith WORLSEY (McGill University).

The geometry in the title is not the geometry of lines and angles but the geometry of topology, shape and knots. For example, galaxies are not distributed randomly in the universe, but they tend to form clusters, or sometimes strings, or even sheets of high galaxy density. How can this be handled statistically? The Euler characteristic (EC) of the set of high density regions has been used to measure the topology of such shapes; it counts the number of connected components of the set, minus the number of 'holes,' plus the number of 'hollows'. Despite its complex definition, the exact expectation of the EC can be found for some simple models, so that observed EC can be compared with expected EC to check the model. A similar problem arises in human brain mapping, where the EC is used to detect local increases in blood flow (positron emission tomography) or oxygenation (functional magnetic resonance imaging) due to an external stimulus. This allows researchers to find the regions of the brain that are 'activated' by the stimulus. Recent work has extended these ideas to manifolds so that we can detect changes in brain shape via structure masking, surface extraction, and 3D deformation fields. Finally, in organic physical chemistry, vortices in excitable media are low density sets that form random closed loops. These are stable if they are linked or knotted, so the question is whether this can actually happen.

Statistics in Fisheries I

Mercredi 13 Juin, 10h30 – 11h00, Images

Définition D'Unités De Conservation Basées Sur La Capacité D'Adaptation Locale: Vérification Des Prédications D'Un Modèle Heuristique

Chris WOOD (Fisheries and Oceans Canada)

Les biologistes en conservation ont mis l'accent sur la protection des espèces rares et en voie d'extinction pour maintenir la biodiversité, la complexité écologique et la capacité à évoluer. Ce discours est raisonnable puisque ces espèces ont évolué à travers des milliers ou des millions d'années et maintenant existent comme bassins de diversité génétique fermés et s'autoperpétuants qui, une fois éliminés, ne peuvent être reproduits sur une échelle de temps humaine. Cependant, les espèces sont des entités résultant des processus d'évolution de mutations, de changements génétiques aléatoires, de migration et de sélection naturelle s'opérant à l'intérieur des populations. La diversité génétique doit aussi être conservée à l'intérieur des espèces en identifiant et privilégiant des unités de population qui requièrent une protection spéciale ou une gestion séparée. Je suggère que notre but devrait être de conserver la capacité d'adaptation locale, et je vais passer en revue les conditions théoriques sous lesquelles cette adaptation locale devrait se produire. Je vais aussi présenter un modèle heuristique pour définir la conservation des unités basées sur des sondages et recensements génétiques. Quoique ce modèle d'équilibre comprenne plusieurs hypothèses irréalistes, ses prédictions se montrent qualitativement raisonnables lorsque testés sur des populations bien connues de saumons 'sockeye'.

Mercredi 13 Juin, 11h00 – 11h30, Images

Un Modèle Statistique Hiérarchique Pour Lier Des Données D'Abondance Du Saumon Spatialement Explicites Aux Caractéristiques De L'Habitat Des Bassins Des Rivières Salmon, Snohomish Et Willamette

Ashley STEEL Blake E. FEIST, George R. PESS, and Robert E. BILBY (U.S. National Marine Fisheries Service)

Parmi les bassins des rivières Salmon, Snohomish et Willamette, une petite proportion des affluents reçoit de manière constante la majorité des saumons en fraie malgré la fluctuation annuelle de la population totale. Nous avons utilisé la constance annuelle de la distribution des poissons pour identifier les relations entre les caractéristiques générales de l'habitat et l'abondance des saumons en fraie. Nous avons relié des comptes « redd » à plusieurs tranches d'information disponibles sur une large échelle spatiale (géologie, température annuelle moyenne de l'air, densité des routes, couverture forestière); ces couches d'habitat ont été sélectionnées pour l'analyse en se basant sur les relations attendues, spécifiques aux espèces, entre l'habitat dans la rivière et l'abondance du poissons à plusieurs stades de vie. Pour tester l'existence de différence statistiquement significative parmi les patrons observés nous avons utilisé un modèle linéaire hiérarchique (MLH) combiné à un test de permutation. Notre méthode a identifié les caractéristiques de l'habitat associées aux effluents les plus productifs et elle a prédit la position de sites dans les bassins ayant les caractéristiques d'habitat propice à supporter de grands nombres de poissons.

Mercredi 13 Juin, 11h30 – 12h00, Images

Méthodes Bayésiennes Pour L'Inférence Relative À Des Populations De Poissons S'Hybridant

Eric ANDERSON (University of Washington)

Un souci grandissant pour la conservation des salmonidés est l'hybridation entre différentes populations et espèces. Par exemple, de petites populations sauvages pourraient subir un flot génétique important venant de poissons s'échappant de piscicultures. De plus, l'utilisation de données génétiques pour définir les 'unités évolutives significatives' peut être compliquée par l'hybridation naturelle survenant entre deux espèces proches. Cependant, l'hybridation entre deux espèces génétiquement non similaires conduit à des changements prévisibles dans la fréquence des génotypes à plusieurs loci et peuvent être étudiés par l'utilisation de marqueurs génétiques. Je vais présenter de récents progrès dans les méthodes bayésiennes pour l'analyse de données génétiques venant de population contenant des hybrides. Je vais illustrer ces méthodes polyvalentes en estimant des fréquences d'allèles et en identifiant la truite fardée et l'omble de mer et leurs hybrides provenant d'échantillons de poissons juvéniles en provenance de l'Oregon et de Washington.

Nonparametric Econometrics

Mercredi 13 Juin, 10h30 – 11h15, AQ 4130

Nonparametric Regression Estimation Using Weak Separability

Joris PINKSE (University of British Columbia)

In this paper I propose a nonparametric kernel estimator of weakly separable regression functions for independent and identically distributed (i.i.d.) data. Weak separability nests a number of other separability conditions, including the additively separable model used by among others Linton and Nielsen (1995) and the generalized additive model of Linton and Hardle (1996) and Horowitz (1999). Like other papers my method uses marginal integration. A desirable and unusual feature of my estimation method is that the integration region can depend on the value of the sub-vector that is not being integrated and can be infinite. Additionally, support of the joint density of regressors can be any open, possibly infinite, convex set. Finally, in most cases the estimator can be computed without numerical integration and computation is hence straightforward.

Mercredi 13 Juin, 11h15 – 12h00, AQ 4130

Développement Économique Et Les Retours Au Capital Humain

Thanasis STENGOS (University of Guelph)

Ce papier examine l'impact de l'accumulation du capital humain sur l'expansion des sorties tout en laissant les contributions des entrées traditionnelles (capital et travail) ainsi que celles du capital humain varier par rapport aux pays et le temps. Le premier point est accompli en construisant un index de croissance du facteur de productivité totale (FPT) pour les entrées traditionnelles, tandis que les effets du capital humain sont estimés par des méthodes semi-paramétriques de régression locale polynomiale utilisant des noyaux. Nous avons trouvé que l'élasticité des sorties du capital humain est en moyenne de 0.25 pour les pays à haut revenus et plus basse et non significative pour les économies à bas et moyen revenus. Nous allons aussi présenter des taux sociaux de retour au capital humain basés sur des données agrégées. Pour les pays à haut revenus, ceux-ci sont fidèles aux estimés de micro études.

Contributed Session: Longitudinal and Correlated Data Analysis

Mercredi 13 Juin, 10h30 – 10h45, AQ 4140

Autoregressive Spatial Smoothing and Temporal Spline Smoothing for Mapping Rates

Ying MACNAB (University of British Columbia)

Within the context of disease mapping and surveillance, there have been a few recent literature on spatio-temporal modeling of disease rate; those publications concern mainly with methodological developments to accommodate temporal effects of various forms. In this presentation I develop generalized additive mixed spline (GAMS) models for an additive modeling of spatial effects, temporal effects and age effects, presenting a spline-based methodology for flexible fitting of nonlinear fixed and random temporal effects while permitting spatial dependence. I propose the use of B-splines, in particular, the use of random B-splines for temporal smoothing and show that the GAMS inference can be implemented using existing methods that are within the generalized linear mixed model (GLMM) framework. I describe an analytic strategy that addresses a variety of issues, including random splines and model invariance, knot selection, computation, interpretation of temporal trends of moderate nonlinearity and production of smoothed maps. The methodology is illustrated by two real-life examples, one within the context of disease mapping and the other of spatio-temporal disease surveillance.

Mercredi 13 Juin, 10h45 – 11h00, AQ 4140

GEE De Second Ordre Pour Des Données En Grappes Avec Des Observations Manquantes Intermittentes

Grace Yun YI and Richard J. COOK (University of Waterloo)

Les équations d'estimation généralisées (GEE) ont été largement répandus pour l'analyse des données en grappes dans des études longitudinales pour estimer des paramètres dans des modèles de régression pour les moyennes marginales. Les équations d'estimation avec probabilités inverses pondérées généralisées (Robins, Rotnitzky, et Zhao 1995) ont été développées pour traiter les données incomplètes qui résultent d'un mécanisme de données « manquantes aléatoirement » (MAR). Cependant, dans plusieurs cas, les mesures répétées peuvent surgir dans en grappes, ce qui mène à une structure de corrélation en coupe et longitudinale. Dans quelques applications, la structure de corrélation peut elle-même devenir d'intérêt. Nous décrivons les procédures d'inférence qui expliquent des données « manquantes aléatoirement » intermittente basées sur des équations d'estimation pondérées généralisées de second ordre pour les coefficients du modèle de régression pour les moyennes marginales. Les méthodes sont appliquées aux données du projet de prévention du tabagisme à Waterloo dans lequel des réponses binaires indiquent l'état du tabagisme.

Mercredi 13 Juin, 11h00 – 11h15, AQ 4140

Maximisation De La Vraisemblance Spatiale Et Ses Applications

Hongtu ZHU, Minggao GU, Julie ZHOU, Fanliang HE (University of Victoria)

We propose a two-stage algorithm for computing maximum likelihood estimates for a class of spatial models. The algorithm combines Markov chain Monte-Carlo methods such as the Metropolis-Hastings-Green algorithm and the Gibbs sampler, and stochastic approximation methods such as the off-line average (Polyak and Juditsky, 1992) and adaptive search direction (Gu and Kong, 1998). A new criterion is built into the algorithm so stopping is automatic once the desired precision is set. Simulation studies have been conducted for two spatial models. Applications are provided to model the distribution of two vegetation types in British Columbia, Canada. This one is different than the one I have..which is this....

Dans cet exposé, je présente un nouvel algorithme stochastique d'approximation à l'aide de la méthode MCMC. On considérera des données d'une espèce de plante

Mercredi 13 Juin, 11h15 – 11h30, AQ 4140

Analyses De Données À Mesures Répétées Avec Groupement À Zéro

Janet TOOZE, Gary K. GRUNWALD, Richard H. JONES (National Cancer Institute)

Les données longitudinales ou à mesures répétées avec groupement à zéro se produisent dans beaucoup d'applications en biométrie, y compris la recherche sur la politique de santé, l'épidémiologie, la nutrition, et la météorologie. Ces données montrent une corrélation parce qu'elles sont mesurées sur le même sujet à travers le temps ou parce que des sujets peuvent être considérés comme des mesures répétées dans une plus grande unité telle qu'une famille. Elles présentent des défis spéciaux en raison de la non-normalité extrême des distributions impliquées. Un modèle pour des données à mesures répétées avec groupement à zéro, en utilisant un modèle de mélanges à effets mixtes avec des effets aléatoires corrélés, est présenté. Le modèle contient des composantes pour modéliser la probabilité d'une valeur différente de zéro et de la moyenne des valeurs différentes de zéro, tenant compte des mesures répétées en utilisant des effets aléatoires et permettant la corrélation entre deux composantes. Des méthodes pour décrire l'effet des variables de prédiction sur la probabilité des valeurs différentes de zéro, sur la moyenne des valeurs différentes de zéro, et sur la moyenne globale sont données. Cette interprétation s'applique également au modèle mixte pour des données en coupe. Les méthodes proposées sont illustrées par des analyses des effets de plusieurs covariables sur des dépenses médicales en 1996 pour des sujets groupés dans des ménages en utilisant des données de l'enquête sur les dépenses médicales.

Mercredi 13 Juin, 11h30 – 11h45, AQ 4140

Modèles De Simplexe Mixte Pour Des Proportions Expérimentales Longitudinales

Zhenguo QIU, Peter X.-K. SONG and Ming TAN (York University)

Nous présenterons un modèle linéaire mixte généralisé (GLMM) basé sur la distribution simplexe de Barndorff-Nielsen et de Jorgensen (1991), qui est appropriée pour modéliser des proportions expérimentales continues longitudinales. L'estimation des paramètres pour les coefficients de régression est effectuée avec la méthode de quasi-vraisemblance pénalisée (PQL) proposée par Breslow et Clayton (1993). La méthode de REML (maximum de vraisemblance restreint) est alors employée pour obtenir les évaluations des paramètres de variance pour des effets aléatoires. Les résultats d'une étude de simulation sont présentés pour étudier les évaluations des paramètres de variance choisis à différents niveaux des paramètres de dispersion. La méthode proposée est illustrée en analysant des données de chirurgie oculaire.

Mercredi 13 Juin, 11h45 – 12h00, AQ 4140

Modèles Mixtes De L'Espace D'État Pour Des Observations Longitudinales Binomiales
 Claudia CZADO (Munich University of Technology) and Peter X.-K. Song (York University)

Nous proposons une nouvelle classe de modèles d'états spatiales pour des données discrètes longitudinales où l'équation d'observation est indiquée par une forme additive impliquant les composantes déterministes et dynamiques. Ces modèles nous permettent d'expliquer l'effet de la tendance saisonnière ou autres covariables variant dans le temps, tout en conservant la puissance des modèles d'espace d'état en modélisant la forme dynamique des données. Nous développons des différents algorithmes de MCMC et comparons leur propriétés de convergence et leurs mélanges dans une simulation expérimentale. Finalement, l'application des modèles proposés est illustrée pour des données de précipitations de Tokyo (Kitagawa, 1987).

Contributed Session: Inference

Mercredi 13 Juin, 10h30 – 10h45, AQ 4120

Test Séquentiel Pour Des Hypothèses Composées
 Edit GOMBAY (University of Alberta)

Nous revisons brièvement les tests séquentiels qui sont utilisés en statistiques appliquées. Nous analysons un type qui est basé sur la vraisemblance, en présence des paramètres de nuisance. On suggère quelques nouvelles procédures qui ont des propriétés optimales de puissance.

Mercredi 13 Juin, 10h45 – 11h00, AQ 4120

Estimation Convergente Des Distributions Avec Un Biais De Type II
 Hari MUKERJEE and H. El BARMI (Wichita State University)

On dit qu'une variable aléatoire X a un biais de type II si $P(X \text{ dans } [0, x])$ est supérieure ou égale à $P(X \text{ dans } [-x, 0])$ pour tout x supérieur ou égal à 0. On sait que l'estimateur non paramétrique du maximum de vraisemblance de la fonction de distribution de X est non convergent lorsque la distribution est continue. Dans ce cas, nous fournissons un estimateur de type projection qui est fortement uniformément convergent. Nous dérivons ses propriétés de convergence faibles et fournissons des procédures asymptotiques d'inférence. On montre que la même technique mathématique peut être employée pour estimer deux fonctions cumulatives d'incidence dans un problème de concurrence de risques lorsque leur taux de risque spécifique sont ordonnées. Tous les résultats sont applicables au cas des données censurées.

Mercredi 13 Juin, 11h00 – 11h15, AQ 4120

Sur L'Estimation Semi-Paramétrique De La Densité D'Une Population Dans Un Échantillonnage En Lignes Transversales

R. KARUNAMUNI and S. ZHANG (University of Alberta)

Cet article étudie quelques méthodes semi-paramétrique potentiellement utiles et probablement appropriées dans l'échantillonnage en lignes transversales. Nous obtenons des expressions des estimateurs semi-paramétriques bayésiens et de non bayésiens pour la densité des populations biologiques. Les estimateurs proposées sont comparées à la méthode du noyau traditionnelle et du maximum de vraisemblance. Quelques exemples d'échantillonnage en lignes transversales seront également analysés.

Mercredi 13 Juin, 11h15 – 11h30, AQ 4120

Estimation De La Densité A Priori Via La Déconvolution De Haar
 Alexandre LEBLANC and Jean-François Angers (Université de Montréal)

Le problème du choix de la densité a priori se présente très souvent en inférence bayésienne. Les solutions les plus populaires à ce problème sont la modélisation hiérarchique, les lois a priori non informatives, diverses techniques bayésiennes empiriques, etc. Une alternative à ces approches qui nous semble fort intéressante consiste à estimer directement, à l'aide des observations, la densité a priori inconnue. Nous considérons cette approche pour l'estimation de la densité a priori d'un paramètre de position. Plus précisément, nous développons une technique permettant d'obtenir une estimation non paramétrique de la densité inconnue. L'approximation résultante est une fonction constante par intervalles qui est facile d'utilisation pour faire des simulations et intégrer numériquement.

Mercredi 13 Juin, 11h30 – 11h45, AQ 4120

Un Test Statistique Pour Comparer Des Populations Binomiale, Poisson Et Binomiale Négative

Humberto VAQUERA, Jose VILLASEÑOR and B.C. ARNOLD (ITESM-CP)

Dans quelques zones de recherches comme la biologie, les assurance, etc., l'utilisation de variables discrètes est fréquente. Les modèles de probabilités discrets les plus utilisés sont binomiale, de Poisson et binomiale négative. Dans cette étude, on propose un procédé statistique pour tester l'hypothèse nulle H_0 : Poisson contre H_a : binomiale ou binomiale négative. Le test est basé sur l'index de dispersion de Fisher. La distribution du test statistique proposée est obtenue asymptotiquement. Pour de petits échantillons, nous utilisons la simulation par ordinateur pour étudier la distribution du test statistique sous l'hypothèse nulle. Une étude de la puissance est également réalisée pour comparer le test proposé à d'autres méthodes. Deux jeux de données sont employés pour illustrer la méthode.

Mercredi 13 Juin, 11h45 – 12h00, AQ 4120

Test Pour Le Nombre De Composantes De Modèles Mixtes Finis Pour Des Modèles Mixtes Finis Avec Vraisemblance Modifiée

Jiahua CHEN, Hanfeng CHEN, Jack KALBFLEISCH (University of Waterloo)

Les modèles mixte finis appartiennent à une classe de modèles non réguliers de probabilité. Les procédures statistiques standards sont souvent inappropriées. Il est particulièrement difficile de tester le nombre de composantes des modèles mixtes finis. Dans cette présentation, nous discutons la pertinence de l'utilisation de la vraisemblance modifiée en construisant un test d'hypothèses avec 2 composantes contre la possibilité de trois composantes ou plus. Nous prouvons que le test statistique proposé a une distribution limite relativement simple. À la différence de nos résultats précédents pour tester l'homogénéité, la distribution limite sous l'hypothèse nulle dépend de la vraie distribution. Nous constatons que la distribution limite sous l'hypothèse nulle peut être déterminée en substituant des paramètres inconnus par leurs estimateurs. Par conséquent, le test peut être fait à un niveau asymptotiquement correct. Nos résultats préliminaires de simulation indiquent que les résultats asymptotiques sont très utiles lorsque la dimension de l'échantillon est à un niveau de 500.

WNAR Student Presentations

Mercredi 13 Juin, 10h30 – 11h00, AQ 4150

Diagnostic Mammaire Assisté Par Ordinateur À L'Aide D'Agrégation De Rééchantillonnage

Mireya DIAZ and J. Sunil RAO (Case Western Reserve University)

La mammographie fournit un outil pertinent pour la détection hâtive des masses irrégulières du sein. Le diagnostic assisté par ordinateur (CAD) utilisant des mammogrammes digitaux peut être un complément très pertinent à un diagnostic visuel. Une approche classifie chaque pixel indirectement dans la masse ou la non-masse par l'intermédiaire d'une classification dichotomique dérivé du profil d'intensité de pixel d'un point d'origine. Une méthode de rééchantillonnage (bagging) combinant ces approximations circulaires simples peut être une méthode pertinente pour détecter plus de forme complémentaires. Deux arrangements d'échantillonnage en ligne brisée des données sont mis en Application : rééchantillonnage des résidus et le rééchantillonnage du pixel du candidat entré. Les deux approches ont été testées sur des « masses » simulées de formes circulaire et elliptique sous différentes conditions de bruit. Les résultats suivent la théorie du "bagging" avec rééchantillonnage des résidus ayant un effet bénéfique pour les masses généralement circulaires, et rééchantillonnant les centres fournissent de grandes diminutions de mauvaise classification pour une forme plus générale. Les techniques seront détaillées et illustrées. En plus, extensions au suivi des profils temporels de forme seront discutées.

Mercredi 13 Juin, 11h00 – 11h30, AQ 4150

Simulation Parfaite D'Équations Différentielles Stochastiques

Xiaoqiang LI and Duncan MURDOCH (University of Western Ontario)

Propp et Wilson (1996, Random Structures & Algorithms) ont proposé un algorithme appelé le couplage du passé (CFTP). Leur approche permet d'utiliser une simulation finie pour produire un échantillon exact à partir de la distribution limite d'une chaîne de Markov. Nous décrivons notre travail en appliquant les idées fondamentales du CFTP à la simulation d'équations différentielles stochastiques (SDEs), par exemple Sdx .

Mercredi 13 Juin, 11h30 – 12h00, AQ 4150

Responder Cell Frequency Estimation and Binomial Three-level Nonlinear Mixed Effects Model in Limiting Dilution Assays

Misoo ELLISON (University of Colorado Health Sciences Center/National Jewish Medical and Research Center)

Responder cell frequencies (RCF), which describe boosted immune responses in Herpes Zoster (HZ) prevention studies, are estimated by using Limiting Dilution Assays (LDA). The theoretical linearity assumption, between the logarithm of the proportion of nonresponding wells (S) and the cell concentration (N) (or dilution level) in LDA, based on the single-hit Poisson model is often violated with observed data resulting in biased estimates of RCF. This Poisson assumption will be modified by applying a mixture of Poisson and Gamma distributions, resulting in a Negative Binomial assumption, which presents a better fit between S and N. In LDA for HZ prevention studies, the responses (nonresponding wells) are repeatedly measured at different cell concentrations and over time. I propose a model, where the response variable is binomial and the probability (S) has a nonlinear functional relationship with covariates, cell concentrations and times. The nonlinear functional relation, between S, N, and time, is based on the negative binomial assumption. This binomial three-level nonlinear mixed effects model accounts for correlation between responses in dilution levels from the same individual and the correlation between repeated assays of an individual over time, simultaneously.

Statistics in Forestry

Mercredi 13 Juin, 1h30 – 2h00, Halpern

Statistiques Bayésiennes Pour Des Décisions Complexes Relatives Aux Ressources Naturelles

Gay BRADSHAW and Jeffrey BORCHERS (USDA Forest Service)

La reconnaissance de la complexité sociale et écologique impliquée dans les plans de conservation a encouragé le développement d'efforts conjoints entre la communauté scientifique, le public et les institutions d'aménagement du territoire. Cependant, étant donné que ces trois secteurs sont soumis à des contraintes et cultures différentes, la communication et la collaboration sont souvent difficiles. Par exemple, les perceptions des risques et des niveaux de confiance face à l'incertitude scientifique sont généralement différentes pour les divers partis. Pour faciliter le processus de collaboration, un modèle bayésien couplé avec une hiérarchie objectif a été développé pour représenter les valeurs, objectifs et données scientifiques utilisés dans le processus décisionnel pour le développement d'un plan de conservation. Le modèle et le processus ont été développés pour le bassin hydrographique de Little Applegate, Oregon, É.U., situé dans une zone de gestion adaptative. Puisque l'intégration de la modélisation et des processus collaboratifs rendent le processus décisionnel transparent, le niveau de confiance entre les trois groupes de collaborateurs a augmenté. Le modèle a aussi été utilisé pour développer une stratégie de suivi et pour évaluer la faisabilité d'activités projetées de gestion du territoire soumises à des contraintes d'infrastructure.

Mercredi 13 Juin, 2h00 – 2h30, Halpern

The Size Distribution Of Wildfires

William REED and Kevin S. MCKELVEY (University of Victoria)

Examination of extensive records on areas burned by wildfires in six regions in North America reveal that there are some commonalities in the size distributions across the regions. However a simple power-law distribution for size, as has been suggested by several authors, is too simple to describe the size-distribution over its full range. A simple stochastic model for the spread and extinguishment of fires is used to examine conditions for power-law behavior and departures from it. Some alternative parametric models for the size distribution are considered with a 'competing hazards' model providing the best fit.

Mercredi 13 Juin, 2h30 – 3h00, Halpern

Système De Sondage En Sylviculture : Propriétés Statistiques Et Production Anticipée

Wendy BERGERUD (B.C. Ministry of Forests)

Le ministère des forêts de la Colombie Britannique possède un système de sondage en sylviculture permettant d'établir si les populations forestières récemment établies ont atteint le stade de « croissance libre ». Les compagnies forestières sont tenues d'établir des populations forestières à croissance libre dans les zones qui ont été exploitées. Lorsque c'est fait, la responsabilité de la gestion des populations forestières incombe alors à la Couronne. Même si les sondages actuels en sylviculture peuvent être vus comme un échantillonnage aléatoire simple, la méthodologie comprends certains paramètres inhabituels ne permettant pas l'utilisation des formules statistiques traditionnelles. Le cadre d'échantillonnage d'acceptation et des méthodes de simulation ont été utilisés pour examiner comment la modification des paramètres de sondage affecte les chances d'une mauvaise identification des populations forestières qui ne sont pas en croissance libre. De plus, puisque le but à long terme du système de sondage est l'identification des populations forestières qui ne fourniront pas une production suffisante lors de la rotation, le modèle ministériel de croissance des arbres TASS à été utilisé pour projeter la production anticipée. Les résultats préliminaires de cette étude seront présentés.

Statistics in Fisheries II

Mercredi 13 Juin, 1h30 – 2h00, Images

Plans De Sondages Idéaux Pour Diverses Distributions Spatiales D'Invertébrés De Fonds Marins

Zane Z. ZHANG (Department Fisheries & Oceans, Nanaimo)

Les populations d'invertébrés sédentaires des fonds marins, tels les oursins de mer, sont souvent sondées en suivant des coupes transversales perpendiculaires à la rive. Ces coupes transversales peuvent être faites de manière systématique, aléatoire ou aléatoirement stratifiée. Des cadres (« quadrats ») sont placés de façon consécutive, systématique, aléatoire ou aléatoirement stratifiée suivant chaque coupe transversale. Les animaux contenus dans chaque cadre sont comptés et/ou échantillonnés pour estimer les paramètres de population. Une simulation informatique à été développée pour examiner la précision relative de ces différents plans de sondage. Cette simulation évalue les différentes stratégies d'échantillonnage par rapport à différents niveaux d'agrégation d'animaux. Elle nous permet de rechercher les plans idéaux pour diverses distributions spatiales d'animaux.

Mercredi 13 Juin, 2h00 – 2h30, Images

Échantillonnage En Grappe Adaptatif Du Sébaste Du Golfe De L'Alaska

Terry QUINN (University of Alaska, Juneau)

L'échantillonnage en grappe adaptatif (ÉGA) a soulevé beaucoup d'intérêt pour l'échantillonnage de populations groupées depuis la publication de Thompson en 1990. Même si plusieurs articles ont été publiés sur l'ÉGA, les expériences pratiques utilisant des populations de poissons sont rares. En 1998 et 1999, deux voyages pour échantillonner le sébaste, en particulier la perche de l'océan Pacifique (POP) et le sébaste boréal et à oeil épineux (SB/OE), ont testé l'échantillonnage en grappe adaptatif. On pense que les POP sont plus regroupées que les SB/OE. La variable d'intérêt particulier était le nombre de prises de sébastes dans une tire au chalut de 15 à 30 minutes. Un échantillonnage adaptatif était effectué sur chaque tire qui excédait le critère fixé. Une règle d'arrêt empêchait l'échantillonnage de s'étendre indéfiniment. Nos hypothèses de travail étaient (1) que l'ÉGA performerait mieux que l'échantillonnage aléatoire simple (ÉAS) pour la POP, et (2) que l'ÉGA serait plus approprié à la POP qu'aux SB/OE puisqu'en 1998 la stratégie d'échantillonnage était l'ÉGA utilisant les statistiques d'ordre pour la valeur du critère. Cette approche a nécessité la prise d'échantillons aléatoires simples sur 12-15 stations à travers 6 strates. En 1999, nous avons utilisé une zone d'étude plus grande et avons considéré quatre alternatives pour le choix de la valeur de critère avant que le sondage ait lieu. L'alternative qui donnait la valeur de critère la plus basse a été utilisée sur le terrain, et les valeurs de critère plus élevées ont été simulées après le sondage. Par rapport à la première hypothèse, la performance de l'ÉGA était parfois meilleure, parfois pire que l'ÉAS considérant un effort d'échantillonnage similaire. Lorsque la population étudiée était modérément agrégée (telle que jugé par l'analyse de variogrammes), l'ÉGA avait une meilleure précision absolue et était plus rapide, mais la précision relative était parfois équivalente ou même pire. Curieusement, l'estimateur ÉAS était souvent considérablement plus élevé que les estimateurs ÉGA, même si ces estimateurs sont non biaisés. Ceci était le résultat d'une grosse prise par tire dans l'échantillon initial, par l'entremise d'un échantillonnage adaptatif. Par rapport à la seconde hypothèse, l'ÉGA était moins efficace pour les SB/OE dû à « l'itshosen », les valeurs de critère élevées résultant en un échantillonnage considérablement moins adaptatif sans grande perte de précision. Un problème important avec l'ÉGA sur le terrain est que les unités en bordure contribuent peu à l'efficacité de l'estimateur mais peuvent prendre un temps considérable à échantillonner. Globalement, nos résultats d'ÉGA n'étaient pas aussi optimistes que ceux publiés dans la littérature mais quand même assez optimistes pour encourager des applications ultérieures.

Mercredi 13 Juin, 2h30 – 3h00, Images

La Détermination De Population De Poissons Utilisant La Capture-Recapture: Un Guide Pratique

Carl SCHWARZ (Simon Fraser University)

Les méthodes de capture-recapture sont des outils bien connus en gestion des pêcheries. Ces méthodes vont du simple modèle de Petersen aux modèles plus complexes à captures multiples et strates multiples. Dans cette présentation, nous ferons un survol de plusieurs types de modèles de capture-recapture utilisés en pratique et passerons en revue leurs avantages et inconvénients.

Innovative Applications of Statistics in Business

Mercredi 13 Juin, 1h30 – 2h00, AQ 4150

Early Detection of High Risk Claims at the Workers Compensation Board of British Columbia

Ernest URBANOVICH, Ella E. YOUNG, Martin L. PUTERMAN, Sidney O. FATTEDAD (Workman's Compensation Board)

The Workers Compensation Board of British Columbia (WCB) annually accepts more than 74,000 short term disability claims for work place injuries and pays out \$900 million dollars in benefits. A small fraction (4%) of these claims, which we refer to as high-risk, require vocational rehabilitation and/or long-term disability payments in addition to initial short-term disability payments. These high-risk claims incur average benefit payments over 40 times greater than the average for claims that do not require these benefits. For the 321,000 claims received between 1989 and 1992, high-risk claims incurred \$1.1 billion (64%) of the total payment of \$1.8 billion for all claims. The purpose of this study was to develop a computer implementable approach for early detection of these costly claims so that more intensive case management may be employed early in the claim process. Our approach combined decision analysis and logistic regression and was applied to data on the 321,000 claims received between 1989 and 1992. Data was stratified by injury type and the risk factors age of worker and number of days of initial payments were found to be highly predictive of high risk status. A decision theoretic approach was used to develop a classification rule that was shown to have high out-of-sample predictive power. The model is now being incorporated in a claims risk scorecard that will be used to determine which claims need early intervention. We estimate annual savings of approximately \$5 million dollars as a result of this targeted early intervention.

Mercredi 13 Juin, 2h00 – 2h30, AQ 4150

Analyzing Telecommunications Network Call Traffic Flow at Telus

Jason GOTO, Isabelle SMITH, Kelly CHUNG and Hong CHEN (Center for Operations Excellence)

Call traffic data detailing the start time, duration and route of individual calls across a telecommunication switching network was analyzed to identify clusters of network sectors with similar time of day demand patterns. The study revealed opportunities to increase network utilization through alternate call routing in adjacent network sectors with complementary time of day demand patterns.

Mercredi 13 Juin, 2h30 – 3h00, AQ 4150

Dynamic Linear Models for Motion Pictures Box-Office Forecasting

Remi DESMEULES (University of British Columbia)

I use a Dynamic Linear Models to predict the gross box-office of movies from week to week, particularly trying to replicate the Top 10 box-office results. I use multiple linear regression to obtain a prior for the parameters of the exponential decay model and then use Dynamic Linear Models to dynamically update the forecast as data is gained on the stream of revenue of the movie. I also use an "attraction model" to fine tune my Top 10 predictions and account for seasonality and competition.

Recent Results on Normal and Related Approximations

Mercredi 13 Juin, 1h30 – 2h00, AQ 4140

An Inequality For Tail Probabilities Of Martingales With Bounded Differences

V. BENTKUS (Vilnius Institute of Mathematics)

We provide precise bounds for tail probabilities $P\{M_n \geq x\}$ of martingales with bounded differences. The bounds are expressed through tail probabilities of sums of i.i.d. Bernoulli random variables. In other words, we show that the tails are sub-Bernoullian. Sub-Bernoullian tails are strictly dominated by Gaussian tails, since their ratio Gaussian/sub-Bernoullian tends to infinity exponentially fast as $x \rightarrow \infty$. Possible extensions of the methods are discussed.

Mercredi 13 Juin, 2h00 – 2h30, AQ 4140

Stein's Method For Computing Error Bounds

Susan HOLMES and Gesine REINERT (Stanford University)

We present new proofs for known results about the convergence in law of the bootstrap distribution to the true distribution of smooth statistics, whether the samples studied come from independent realizations of a random variable or dependent realizations with weak dependence. The techniques employed are based on Stein's method for empirical processes as developed by Gesine Reinert in her Ph D thesis, with whom this is joint work.

Mercredi 13 Juin, 2h30 – 3h00, AQ 4140

Asymptotic Expansions For Studentized Statistics Under Weak Dependence

Soumendra N. LAHIRI (Iowa State University)

Let X_1, X_2, \dots be a sequence of stationary strongly mixing random vectors with mean vector μ . Let $\hat{\theta}_n = H(\bar{X}_n)$ be an estimator of a parameter of interest $\theta = H(\mu)$, where H is a smooth function. In this talk, we consider asymptotic expansions for Studentized statistics of the form $T_n = n^{1/2}(\hat{\theta}_n - \theta) / s_n$ for a class of Studentizing factors s_n that estimate an (possibly unbounded) number of lagged autocovariances of the sequence $\{X_n\}$. Some applications of the expansions to block bootstrap methods are also indicated.

Block Designs: Directions and Inspirations

Mercredi 13 Juin, 1h30 – 1h50, AQ 4130

A- And MV-Efficient Designs For Comparing A Set Of Test Treatments To A Set Of Controls

M. JACROUX (Washington State University)

The problem of comparing t test treatments to s controls using experimental units arranged in b blocks of size k is considered. Methods are given for deriving highly A- and MV-efficient designs for such situations. The methods given combine optimality results obtained for one-way classification models with the optimal augmentation of blocks with controls of block designs which are known to be efficient for making comparisons among the test treatments. Examples are given to show how the results obtained can be applied.

Mercredi 13 Juin, 1h50 – 2h10, AQ 4130

Design Of Computer Experiments To Optimize The Mean Of A Response Subject To Constraints

W. NOTZ (Ohio State University)

For purposes of this talk, a computer experiment is assumed to consist of computer code that produces deterministic responses for a given set of inputs. We consider the situation in which there are two types of inputs; manufacturing variables and noise variables. Manufacturing variables are those that can be controlled by the product designer. Noise variables cannot be controlled but have values that follow some probability distribution. We are interested in the means of the responses over the distribution of the noise variables, and we seek the values of the manufacturing variables that optimize the mean of one of the responses, subject to constraints on the mean of another response. The approach is Bayesian; in the case of two responses, the prior information is that the responses are draws from a covariance stationary bivariate Gaussian stochastic process with correlation functions belonging to a parametric family with unknown parameters. A latin hypercube design is used to provide an initial fit of this model. Subsequently, a sequential strategy is used to select values of the inputs at which to observe the response. This strategy involves computing, for each unobserved set of values of the inputs, the posterior expected "improvement" over the current best guess at the optimum and the next set of inputs to observe are those that maximize this expected improvement. The strategy is illustrated with examples.

Mercredi 13 Juin, 2h10 – 2h30, AQ 4130

Optimalité Des Designs Par Bloc : Résultat Pour Le Cadre Combinatoire

Brian RECK (Old Dominion University) and J. P. Morgan (Virginia Tech)

Étant donné v traitements à comparer et ayant b blocs de k unités chacun de disponible, un statisticien songeur peut se demander « qu'elle est la répartition optimale des traitements sur les unités? » Ceci est le problème initial des designs par bloc. Si n_{ij} est le nombre de fois que le traitement i est utilisé dans le bloc j et si N est une matrice $v \times b$,

Mercredi 13 Juin, 2h30 – 2h50, AQ 4130

Plans Polygonaux

John STUFKEN (Iowa State University)

Il est bien connu qu'il existe des relations intéressantes entre certains concepts en plan d'expérience et en échantillonnage. Il y a en particulier une relation intime entre les plans en bloc et les plans d'échantillonnage. Motivés par le désir d'éviter la sélection d'unités adjacentes dans un échantillon, A. Hedayat, C.R. Rao et J. Stufken (Journal of Statistical Planning and Inference 19, 159-170, 1988) ont initié l'étude d'une classe de plans en bloc maintenant connue sous le nom de plans polygonaux. Nous discuterons des résultats récents sur les plans polygonaux et les plans en bloc reliés, et donnerons des directions pour la recherche future, cette dernière motivée par l'application à l'échantillonnage.

Survey Methods Contributed Session: Estimation 1

Mercredi 13 Juin, 1h30 – 1h50, AQ 4120

Un Test Pour Des Fonctions De Survie Utilisant Des Données D'Échantillons Complexes

Susanna Rubin BLEUER (Statistics Canada)

La distribution des durées de périodes et d'autres distributions de temps de vie peuvent être estimées en appliquant des méthodes d'analyse de survie. Des méthodes d'analyse de survie ont été développées pour des données indépendantes identiquement distribuées (i.i.d.) d'une super-population. Quand des données sont obtenues à partir d'une méthodologie complexe de sondage, l'hypothèse des données i.i.d. n'est pas validée et les résultats standards ne peuvent pas s'appliquer. Les exemples des données de temps de vie de Statistique Canada dans des études complexes sont les longueurs des périodes du bas revenu, lesquelles peuvent être obtenues à partir de la base de données en fonction du chômage peuvent être obtenues à partir de l'étude sur le travail et la dynamique du revenu (SLID). Dans cet article, nous nous concentrons sur un test d'hypothèses non paramétrique pour évaluer des différences observées dans les courbes de survie empiriques. Nous développons un test, basé sur la statistique de Gehan-Wilcoxon, qui peut être appliqué à des données provenant de plans p.p.s. Les résultats nécessaires exigent l'utilisation d'un nouvel espace de probabilité qui inclut l'espace échantillonnal et l'espace de la super-population.

Mercredi 13 Juin, 1h50 – 2h10, AQ 4120

Complexités Dans L'Analyse De Données De Durée Dans Des Enquêtes Longitudinales - Avec Application À L'Étude Sur La Dynamique Du Travail Et Du Revenu

Georgia R. ROBERTS (Statistics Canada)

Plusieurs études ont été présentées par Statistique Canada dans les années 90 pour répondre à un besoin d'information longitudinale sur les individus et les familles. L'étude sur la dynamique du travail et du revenu (SLID) est une de ces études. Le but de SLID est de supporter la recherche sur des sujets tels que la dynamique du revenu (par exemple des mouvements vers et hors de la pauvreté) et sur la dynamique du travail (par exemple des mouvements vers et hors du chômage). Avec des données maintenant disponibles pour chacune des 6 années (1993 - 1998) du premier panneau de SLID, les analystes ont une source de données riche pour effectuer leur recherche longitudinale. Un type de recherche pour lequel SLID a été développé est l'examen des facteurs affectant la longueur de période puisque des données rétrospectives suffisamment détaillées ont été rassemblées à l'aide d'entrevues annuelles pour que la mesure précise des périodes soit possible. Quelques approches typiques trouvées dans la littérature pour ce type de recherche sont (i) régression logistique; (ii) régression logistique de personnes-temps; et (iii) modélisation du risque proportionnel de Cox. Dans cet article, chacune des trois approches est considéré pour identifier des facteurs affectant les longueurs des périodes de chômage. Nous examinerons si la complexité du plan d'enquête et des idiosyncrasies des données de l'étude affectent ces méthodes. Les problèmes dont : quel individu échantillonné inclure dans l'étude; quelle pondération de l'étude doit-on utiliser; quelle est la façon de traiter plus d'une période fourni par le même individu échantillonné et de la censure seront discutés. Un exemple basé sur des données du premier panneau de SLID sera employé pour illustrer les issues discutées dans la recherche.

Mercredi 13 Juin, 2h10 – 2h30, AQ 4120

Sur Le Traitement Des Observations Influentes Pour L'Enquête Sur Les Ménages

Asma ALAVI (Statistics Canada)

Les enquêtes sur les dépenses ou sur le revenu des ménages traitent souvent les distributions fortement asymétriques, qui mènent potentiellement à des échantillons avec quelques observations extrêmes. Le problème est aggravé par le fait qu'il y a une faible quantité d'information auxiliaire utile disponible à l'étape de la conception du plan et que le plan d'échantillonnage est complexe la plupart du temps, menant à des poids d'échantillonnage largement dispersés. Par conséquent, il pourrait se produire qu'une grande valeur soit associée à une grande pondération et que cette combinaison ait une grande influence sur les estimateurs produits par l'étude. Des estimateurs convergents, tels que les estimateurs généralisés pour la régression (GREG), sont habituellement fortement variables en présence des observations influentes, mais ils ont un faible biais, tandis que les estimateurs basés sur un modèle sont plus stables mais ils ne sont généralement pas convergent et ils sont davantage biaisés. Dans cet article, on propose un compromis entre ces deux types d'estimateurs et une étude de simulations prouve qu'il performe bien en ce qui concerne les critères du biais et de l'écart quadratique moyen (MSE) en comparaison avec quelques autres estimateurs robustes; des conditions pour lesquelles le compromis devrait avoir un petit biais dû au plan d'expérience sont également données.

Mercredi 13 Juin, 2h30 – 2h50, AQ 4120

L'Exactitude D'Un Modèle De Prévion Sur Des Individus Suréchantillonnés Ayant Des Niveaux Élevés De Dépenses Médicales Dans L'Enquête Sur Les Dépenses Médicales

Steven COHEN (Agency for Healthcare Quality and Research)

Les efforts pour améliorer la précision des estimateurs des soins de santé associés aux individus qui encourent des niveaux élevés des dépenses médicales ont été considérés dans l'enquête sur les dépenses médicales. La concentration des dépenses de santé dans un pourcentage relativement petit de la population a servi à motiver des efforts supplémentaires pour assurer la représentation suffisante de ce noyau du domaine analytique, pour supporter des analyses de sous-groupe. En utilisant des données de l'enquête nationale sur les dépenses médicales antérieures, un modèle probabiliste a été développé pour identifier des individus à risque d'encourir un niveau élevé des dépenses médicales, basé sur leur expérience antérieure de santé. Dans cet article, la spécification du modèle de prévision est décrite en détail. De plus, l'exactitude de ce modèle pour prévoir des individus à risque d'encourir un niveau élevé de dépenses médicales est évaluée, en utilisant des données de l'enquête sur les dépenses médicales. L'enquête sur les dépenses médicales (EDM) a été conçue pour produire des estimations nationales et régionales annuelles de l'utilisation de soins de santé, des dépenses, des sources de paiement et de la protection d'assurance des individus civils non institutionnalisés des États-Unis. L'EDM inclut une étude des fournisseurs médicaux, pour compléter les données fournies par des répondants des ménages. La conception de l'EDM permet des estimations basées autant sur une personne que sur un ménage. La portée et la profondeur des efforts pour la collecte de données reflètent les besoins des organismes gouvernementaux, des corps législatifs, et des professionnels de santé, pour des estimations nationales complètes, nécessaires dans la formulation et l'analyse des politiques sanitaires nationales. L'étude est commanditée par Agency for Healthcare Research and Quality (AHRQ) avec la co-commandite du National center for Health Statistics (NCHS).

Statistics in Fisheries III

Mercredi 13 Juin, 3h30 – 4h00, Images

Utilisation De La Post-Stratification Pour L'Incorporation D'Information Sur L'Habitat Pour Des Estimations D'Abondance Lors De Sondages Comportant Plusieurs Espèces

Sherri DRESSEL (University of Alaska, Fairbanks)

Plusieurs études ont fait le lien entre la distribution, l'abondance et la survie d'espèces de poissons démersaux et les caractéristiques de l'habitat, mais les estimés de population utilisent rarement l'information sur l'habitat ou les distributions de densité de poissons démersaux déterminées auparavant. Puisqu'il a été démontré que les paramètres de l'habitat ont un effet sur la distribution spatiale de la densité et de l'abondance des poissons démersaux, leur utilisation en recherche et estimations de stocks devrait mener à des estimés de population plus précis et fiables. Puisque les informations sur l'habitat ne sont pas souvent connues avant l'échantillonnage, elles ne peuvent pas être utilisées pour la stratification des sondages au chalut des poissons démersaux, ces sondages étant utilisés pour évaluer l'abondance des espèces. Même si les informations sur l'habitat sont connues avant l'échantillonnage, les espèces examinées par ces sondages au chalut peuvent être distribuées différemment, relativement aux paramètres de l'habitat, rendant difficile la stratification optimale pour chaque espèce. Cette étude utilisait la post-stratification pour inclure les informations sur l'habitat et la densité de poissons, obtenue lors de sondages au chalut, aux estimés d'abondance totale pour quatre espèces de poissons plats juvéniles: la fausse limande (*Pleuronectes bilineatus*) d'âge 0, la limande à nageoires jaunes (*Pleuronectes asper*) d'âge 1, le flétan du Pacifique (*Hippoglossus stenolepis*) d'âge 0, et la plie à tête plate (*Hippoglossoides elassodon*) d'âge 0. Je vais aussi décrire une méthode plus rapide pour déterminer les changements statistiquement différents des abondances annuelles, qui se rapprochent beaucoup des différences annuelles des estimés post-stratifiés. Nos résultats montrent l'importance de l'identification de l'habitat des espèces lors de l'estimation des populations, et illustrent les avantages et inconvénients de la méthode de post-stratification. Même si cette étude mettait l'accent sur quatre espèces de poissons plats juvéniles, le but ultime était de développer des méthodes analytiques pour l'estimation d'abondance totale et la détection de variations annuelles dans l'abondance. Ces méthodes pourraient être utilisées pour toutes espèces, en particulier celles présentes lors de sondages à plusieurs espèces.

Mercredi 13 Juin, 4h00 – 4h30, Images

Problèmes D'Échantillonnages Auxquels Les Livres Ne Nous Ont Pas Préparés

Brenda NORCROSS (University of Alaska, Fairbanks)

Il est reconnu qu'une bonne attaque constitue la meilleure défense. Appliqué aux sondages de poissons cela signifie qu'un bon plan d'échantillonnage sur le terrain donnera les résultats les plus significatifs et justifiables. Malheureusement, dans le monde de l'échantillonnage en pêcheries, non seulement y a-t-il de nombreuses contraintes à une bonne planification, mais aussi des complications inconnues, donc non prévues par le plan, qui peuvent survenir lors de l'étude même. Un exemple de ceci se retrouve dans les sondages de fond fait au chalut. Même si le plan original est un plan d'échantillonnage aléatoire stratifié, la topographie des fonds marins en Alaska et en Colombie Britannique peut être un obstacle à la définition des strates. De plus, les différences entre les structures et les types de fonds marins actuels et répertoriés ainsi que les limites imposées à l'échantillonnage par le climat confondent davantage le plan original. D'autres exemples de problèmes de planification et d'exécution de sondages seront présentés. Les analyses statistiques utilisées seront décrites. Les commentaires de l'assistance sur les alternatives statistiques sont les bienvenus.

Mercredi 13 Juin, 4h30 – 5h00, Images

Une Approche Bootstrap Pour Études Au Chalut Servant À Effectuer Des Calculs De Taille D'Échantillon

Steven SYRJALA (NMFS, Seattle)

Un aspect souvent difficile dans la planification d'études au chalut est de faire des calculs de taille d'échantillon. Un plan d'étude usuel est une expérience à deux échantillons examinant un changement proportionnel par rapport à une variable continue, par exemple la densité d'une espèce avant et après une perturbation. Des calculs de taille d'échantillon lors d'expériences à deux échantillons sont souvent effectués en utilisant la théorie du test-t à deux échantillons. Les données de chalut sont souvent obtenues à partir de distributions non négatives avec une asymétrie positive. Lorsque l'on utilise de telles données pour effectuer des calculs de tailles d'échantillon, l'intérêt porté aux changements proportionnels et la distribution asymétrique des données suggèrent l'utilisation d'une transformation logarithmique dans le but de se rapprocher de la structure et des hypothèses du test-t. Cette transformation convertit un changement proportionnel en changement additif et rend habituellement la distribution des données plus symétrique, plus près d'une distribution normale. Un problème survient lorsque les données contiennent des zéros. Une solution commune à ce problème est d'utiliser la transformation $\ln(X_i + \delta)$ où les données sont représentées par X_i , et où δ est une petite valeur (la plupart du temps 1). Malheureusement, le choix de la valeur δ peut affecter considérablement les résultats des calculs de tailles d'échantillon. Une autre approche pour effectuer des calculs de taille d'échantillon est d'utiliser le bootstrap. Sous l'hypothèse nulle d'aucun changement, la distribution d'une métrique pour la distance entre deux échantillons provenant de la même population peut être générée. La distribution de la même métrique lors d'un changement de la population spécifié sous l'hypothèse alternative peut alors être utilisée pour estimer la puissance pour une taille échantillonnale donnée.

Brownian Sheets and Stochastic Partial Differential Equations

Mercredi 13 Juin, 3h30 – 4h00, AQ 4140

Non-independence of excursions of the Brownian sheet and of additive Brownian motion

Robert DALANG and T. MOUNTFORD (Ecole Polytechnique Federale, Switzerland)

A classical and important property of Brownian motion is that given its zero set, distinct excursions away from zero are independent. In this talk, we examine the analogous question for the Brownian sheet, and also for additive Brownian motion. The main result is that given the level set at level zero of the Brownian sheet, distinct excursions of the sheet away from zero are $\{ \em not \}$ independent. In fact, given the absolute value of the Brownian sheet in the entire non-negative quadrant, and the sign of all but a finite number of excursions away from zero, the signs of the remaining excursions are determined. For additive Brownian motion, we prove the following definitive result: given the absolute value of additive Brownian motion in the non-negative quadrant and the sign of a single excursion, the signs of all other excursions are determined.

Mercredi 13 Juin, 4h00 – 4h30, AQ 4140

The Dimension Of The Boundary Of Brownian Sheet Via Additive Diffusions

Thomas MOUNTFORD (University of California, Los Angeles) and Robert DALANG (Ecole Polytechnique Federale)

The Dalang-Walsh algorithm for "evaluating" the bubble of additive diffusions is used to evaluate the Hausdorff dimension of the boundary of a bubble of additive Brownian motion. This method is then adapted to deal with the Brownian sheet.

Mercredi 13 Juin, 4h30 – 5h00, AQ 4140

Some Recent Results Of Interest

John WALSH (University of British Columbia)

We investigate the two-point stochastic boundary-value problem on $[0,1]$:
$$U'' = f(U) \dot{W} + g(U, U') U(0) + \xi U(1) + \eta.$$
 where \dot{W} is a white noise on $[0,1]$, ξ and η are random variables, and f and g are continuous real-valued functions. This is the stochastic analogue of the deterministic two point boundary-value problem, which is a classical example of bifurcation. We find that if f and g are affine, the problem does not exhibit bifurcation: for any r.v. ξ and η , (1) has a unique solution a.s. However, as soon as f is non-linear, bifurcation appears. We investigate when there is positive probability that there is either no solution whatsoever, a unique solution, or multiple solutions. While these all involve conditions on f and g , we conjecture that the only case in which there is no bifurcation is when f is affine.

Survey Methods Contributed Session: Estimation 2

Mercredi 13 Juin, 3h30 – 4h00, AQ 4150

Évaluation Efficace Des Fonctions Quadratiques De Population Finies En Présence D'Information Auxiliaire

Changbao WU and Randy R. SITTER (University of Waterloo)

En regardant les fonctions quadratiques de population finie ou de second ordre comme totaux ou moyennes au-dessus d'une population finie synthétique dérivée, on montre que le modèle calibré récemment proposé et les pseudo-fonctions de vraisemblance empiriques, pour un usage efficace de l'information auxiliaire des données d'étude, peuvent être aisément étendus pour obtenir des estimateurs efficaces des fonctions quadratiques de population finie ou de second ordre. En particulier l'estimation, pour une population finie, de la variance, de la covariance, ou de la variance d'un estimateur linéaire peut être considérablement améliorée lorsque l'information auxiliaire est disponible. Les méthodes proposées sont basées sur des modèles au sens où les estimateurs résultants sont asymptotiquement non biaisés par leur plan d'expérience indépendamment de l'exactitude du modèle choisi mais très efficaces si le modèle choisi est presque correct. Elles ont un certain nombre de particularités attrayantes qui incluent: applicabilité à un plan d'échantillonnage général; incorporation de l'information sur des variables auxiliaires possiblement multivariées; la capacité de considérer des modèles linéaires ou non linéaires; et il en résulte des estimateurs non négatifs pour certaines quantités strictement positives telles que les variances. Plusieurs estimateurs existants s'avèrent des cas spéciaux de la méthodologie générale proposée en considérant un modèle linéaire.

Mercredi 13 Juin, 4h00 – 4h30, AQ 4150

Existence Et Uniformité Des Estimateurs EEG. Application À L'Inférence Basée Sur Un Plan D'Expérience

Ioana SCHIOPU-KRATINA (Statistics Canada)

Les conditions d'existence des racines des équations d'estimation généralisées sont données pour les modèles pondérés ou non. L'uniformité des estimateurs obtenus est étudiée. Les résultats sont appliqués à l'inférence basé sur un plan d'expérience quand le plan est à plusieurs étapes avec le remplacement à la première étape.

Mercredi 13 Juin, 4h30 – 5h00, AQ 4150

Évaluations De Paramètres Par Calibrage Dans Des Modèles Logit À Choix Binaire

Ian CAHILL (Statistics Canada)

Une approche pour exploiter des données d'enquêtes multiples et des époques par calibrage les estimateurs des paramètres des modèles de logit à choix binaire et des modèles de survie semi-paramétrique est développée. Le but est d'exploiter la source relativement riche des covariables socio-économiques offerte par l'enquête de Statistique Canada sur la dynamique du travail et de revenu (SLID), et également l'envergure historique de l'enquête sur les forces de travail (LFS), mise en valeur par le suivi des individus au moyen d'une entrevue basée sur une rotation de six mois. Une démonstration de la façon dont la méthode peut être appliquée est effectuée, à l'aide du module de congé de maternité du projet de microsimulations de la dynamique d'un cheminement de vie de Statistique Canada. Le choix du congé de maternité plutôt que la séparation du travail est indiqué comme modèle binaire de logit, tandis que la durée du congé de maternité est indiquée comme modèle de risque de survie proportionnel semi-paramétrique avec covariables ainsi qu'un risque de base pouvant changer à chaque mois. Les deux modèles sont initialement estimés par la méthode du maximum de vraisemblance des données de SLID prises de 1993-1997, puis calibrés sous la forme d'estimateurs annuels pour LFS 1976-1992. Dans le cas du modèle de logit, le prédicteur linéaire est ajusté par estimateur log-ratio pour le LFS. Pour le modèle de survie, un estimateur de Kaplan-Meier de la fonction de risque du LFS est utilisé pour ajuster le risque prévu dans le modèle semi-paramétrique.

Contributed Session: Regression

Mercredi 13 Juin, 3h30 – 3h45, AQ 4130

Régression Non Paramétrique Bayésienne À L'Aide D'Ondelettes

Jean-Francois ANGERS (Université de Montréal) and Mohan DELAMPADY, (ISI-Bangalore)

Dans cette conférence, nous voulons construire une fonction de régression non paramétrique basée sur des observations bruitées. Cette fonction de régression est construite à l'aide d'ondelettes et de l'analyse multi-résolution et en utilisant une approche bayésienne hiérarchique. En utilisant les cotes de Bayes, nous faisons aussi de la sélection de modèle afin de choisir le niveau de résolution optimale. Des bandes de confiance sont aussi données comme mesure de l'erreur d'estimation. Deux exemples sont discutés afin d'illustrer la méthodologie présentée dans cet article.

Mercredi 13 Juin, 3h45 – 4h00, AQ 4130

Un Modèle De Régression Binomiale Négative Transitif De Markov Et Ses Applications

Peiming WANG (Nanyang Technological University)

Cet exposé présente une méthode pour modéliser les données de compte répétées, dans laquelle le processus de transition est modélisé en utilisant une chaîne de Markov, et conditionnel aux états de la chaîne de Markov, la réponse de compte a une distribution binomiale négative. Il est permis d'incorporer des covariables soit dans les paramètres de la binomiale négative ou dans les probabilités de transition de la chaîne de Markov ou dans les deux. Comme dans le cas du modèle de régression transitif, deux régimes sont différenciés soit avec des processus connus différents pour générer les données ou des estimateurs de probabilités a posteriori du modèle proposé. Les caractéristiques du modèle proposé, l'estimation par le maximum de vraisemblance et quelques propriétés des estimateurs dans un échantillon de taille finie sont discutées. La méthode est illustrée par une application.

Mercredi 13 Juin, 4h00 – 4h15, AQ 4130

Estimation Des Paramètres De Régression Avec Des Covariables Manquantes

Cynthia STRUTHERS and D. L. McLEISH (St. Jeromes University and University of Waterloo)

Il est fréquent dans les applications de la régression qu'une ou plusieurs covariables soient non observées pour certains des sujets expérimentaux à cause design (par exemple ils sont chers ou difficiles d'obtenir) ou par un certain mécanisme aléatoire de censure. Spécifiquement, supposez que Y est une variable dépendante, probablement multivariée, avec une fonction de densité $f(y|x, v; \alpha)$ conditionnelle sur les covariables (x, v) où x et v sont des vecteurs et α est un vecteur des paramètres inconnus. Nous considérons le problème d'estimer les paramètres d' α lorsque les données sur le vecteur v de covariable sont disponibles pour toutes les observations tandis que les données sur la covariable x sont manquantes pour certaines des observations. Nous supposons D .

Mercredi 13 Juin, 4h15 – 4h30, AQ 4130

Changement Brusque Ou Pas? – « Du Bâton Rompu Au Câble Courbé »

Grace CHIU, Richard ROUTLEDGE and Richard LOCKHART (Simon Fraser University)

Dans le domaine des analyses de point de rupture et de la régression segmentée, le modèle du « bâton cassé » est largement utilisé pour décrire les courbes de régression continues qui peuvent différer radicalement de chaque côté d'un point de rupture inconnu. Dans ma recherche de doctorat, je développe une théorie d'estimation pour un modèle lisse de transition que j'appelle « câble courbé ». Ce modèle a été suggéré initialement par Tishler et Zang (1981) comme outil numérique pour lisser la fonction du « bâton cassé », permettant l'utilisation d'un algorithme numérique plus rapide et plus stable pour estimer la rupture inconnue. Le « câble courbé » comporte deux lignes droites reliées par une courbe quadratique au milieu. En modélisation statistique, c'est une alternative au « bâton cassé » qui a été considérablement critiqué dans beaucoup d'applications en raison de son présupposé de rupture. Au lieu de cela, le « câble courbé » est plus flexible et incorpore une région possible de transition douce. En fait, le « bâton cassé » est simplement un « câble courbé » avec un segment courbé infinitésimal. Une théorie exhaustive d'estimation pour le modèle de « câble courbé » était non disponible jusqu'à ce que je réussisse à identifier quelques propriétés intéressantes, quand la relation sous-jacente à (1) chacun des trois segments, ou (2) aucun segment quadratique. Dans cette présentation, je mentionnerai certaines de ces propriétés, et montrerai des exemples de l'application du modèle du « câble courbé ».

Mercredi 13 Juin, 4h30 – 4h45, AQ 4130

Une Approche Avec Un Modèle De Poisson Apparié Pour L'Analyse De Modèle Mixtes Binomials
Renjun MA (University of New Brunswick)

La généralisation des équations de modèle mixte linéaires de Henderson (1975, Biometrics) a été l'approche la plus largement adoptée par les non-Bayésiens aux modèles mixtes linéaires généralisés (Breslow et Clayton 1993; Lee et Nelder 1996); cependant, ces approches connues donnent des estimateurs incohérents pour les modèles mixtes binomials. Jiang (1998, JASA) a présenté une méthode de moments simulés pour l'estimation cohérente des paramètres dans les modèles mixte linéaires généralisés, mais cette approche est inefficace. Nous proposons une approche utilisant un modèle de Poisson apparié pour l'analyse d'un modèle mixte binomial basés sur notre approche BLUP aux modèles mixtes de Poisson. Notre approche donne non seulement des estimateurs cohérents pour les paramètres de régression et de dispersion, mais mène également aux estimateurs optimaux de paramètres de régression dans le sens de Godambe (1976, Biometrika).

Mercredi 13 Juin, 4h45 – 5h00, AQ 4130

Modèles De Régression Sur Des Événements Historiques Pour Les Covariables À Intervalle Censuré
Bingshu CHEN and Richard COOK (University of Waterloo)

Dans l'analyse d'événements historiques, on est souvent intéressé à étudier comment un processus de repère affecte la distribution d'un autre processus, appelé le processus de réponse. Le processus de repère pour le déclenchement d'un certain événement clinique intermédiaire peut être modélisé par un modèle simple à deux-états et la réponse peut être le temps de survie (par exemple Guggins et al., 1999). Alternativement, le processus de repère peut être représenté par un processus ponctuel et la réponse peut être représentée par un modèle stochastique plus général. L'exemple de motivation pour notre travail est un essai clinique pour des patients avec des métastases cancéreuse dans les os. Nous sommes intéressés à évaluer comment le développement, de nouvelles lésions métastatique dans les os, affecte les incidences des complications squelettiques. Ici nous étudions les méthodes qui nous permettent d'exprimer l'association par des modèles de régression. Ceci est réalisé en considérant le processus de repère (le développement de la tumeur) avec le temps interne comme covariable dépendante pour un événement récurrent défini par l'occurrence des complications squelettiques. De nouvelles tumeurs sont seulement détectées par un examen périodique radiographique rendant les processus de repères sujet aux censures par intervalles. Un algorithme EM sera adopté pour traiter ce problème des données incomplètes.

Contributed Session: Assorted

Mercredi 13 Juin, 3h30 – 3h45, AQ 4120

Rôle Du Paramètre De Forme Pour Les Estimateurs À Rétrécisseur De La Moyenne
Shahjahan KHAN (University of Southern Queensland)

Le rôle du paramètre de forme dans la détermination des propriétés de la fonction de rétrécissement et des estimateurs à rétrécisseur positifs du vecteur moyen des populations multidimensionnelles est d'intérêt majeur dans cet article. L'approche préliminaire du test pour l'estimation à rétrécisseur est employée pour définir deux estimateurs de type Stein basés sur l'information échantillonnale et l'information incertaine a priori non échantillonnée. L'impact du changement de la valeur du paramètre de forme sur la performance des estimateurs en ce qui concerne les critères de biais et de risque quadratique est étudié. L'analyse graphique de l'effet de la dimension de la population sur les propriétés ci-dessus est également fournie.

Mercredi 13 Juin, 3h45 – 4h00, AQ 4120

Interprétation De La Réduction De La Dimension
Hong GU and Hugh CHIPMAN (University of Waterloo)

L'analyse des données de grande dimension commence souvent par l'identification des sous-espaces dimensionnels inférieurs. L'analyse en composantes principales (PCA) est une technique de réduction de dimension qui identifie des combinaisons linéaires des variables qui expliquent une grande partie de la variabilité ou « reconstruisent » le mieux la variable originale. Par exemple, plusieurs lectures de températures peuvent être prises dans un procédé de production quand en fait il y a juste quelques variables fondamentales pilotant le processus. Un problème avec les composantes principales est que les combinaisons linéaires peuvent sembler tout à fait arbitraires. Pour les rendre plus interprétables, nous présentons des contraintes sur les coefficients de la combinaison linéaire. Deux classes de contraintes sont considérées - ceux dans lesquels les coefficients sont égaux à un de un nombre restreint de valeurs (contraintes homogènes), et ceux dans lesquels beaucoup de coefficients sont égaux à 0 (des contraintes de parcimonie). Les directions interprétables résultantes peuvent être calculées pour être près des composantes principales initiales, ou être calculées d'une façon pas à pas qui peut rendre les composantes plus orthogonales. Un petit jeu de données sur des caractéristiques des voitures est employé pour présenter les techniques, qui sont alors employées à une application plus substantielle du forage de données.

Mercredi 13 Juin, 4h00 – 4h15, AQ 4120

Réduction Du Biais De L'Estimateur De La Taille D'Une Population Animale

Tina LEVESQUE (Université Laval)

Les estimateurs déjà existant de la taille d'une population animale fermée obtenus par des expériences de capture-recapture comportent des biais importants lorsque les probabilités de capture sont faibles. Alors, il serait intéressant de pouvoir développer une méthode permettant de réduire le biais de ces estimateurs. Lors de cette présentation, le modèle de Petersen pour deux occasions de capture sera présenté en premier lieu suivi du modèle M_t et du modèle de retrait. Ensuite, une correction du biais pour de petits échantillons sera introduite. Enfin, des résultats de simulations seront présentés. Ce travail a été réalisé sous la direction de M. Louis-Paul Rivest, professeur de statistique à l'Université Laval.

Mercredi 13 Juin, 4h15 – 4h30, AQ 4120

Une Méthode De Monte Carlo Non Itérative Pour L'Échantillonnage Multidimensionnel Et L'Intégration

Liqun WANG and James C. FU (University of Manitoba)

Nous proposons un algorithme non itératif pour générer un échantillon aléatoire d'une distribution multidimensionnelle donnée. Cette méthode exige seulement la connaissance de la fonction de densité (à une constante de normalisation près). Elle est basée sur la discrétisation aléatoire de l'espace échantillonnal et l'inversion multidimensionnelle des probabilités. Il est facile de la mettre en application, elle est efficace en terme de temps de calcul et applicable à presque toutes les distributions multidimensionnelles provenant de vraies applications. En raison de sa simplicité et de son efficacité, cet algorithme permet un choix flexible des distributions a priori, qui est important dans l'analyse bayésienne non conjuguée pour des problèmes réels. Cette approche est appliquée à beaucoup d'exemples typiques de la littérature récente.

Mercredi 13 Juin, 4h30 – 4h45, AQ 4120

Limites Percentiles Basées Sur Des Transformations Normales Pour Des Distributions De Durée De Vie

Zhenlin YANG (National University of Singapore)

En utilisant une transformation de puissance, les distributions usuelles de durée de vie sont transformables de telle sorte qu'elle soit exactement ou s'approche de la normalité. Les limites percentiles exactes ou approximatives sont ainsi obtenues par une simple transformation inverse. Lorsque la transformation est inconnue et qu'elle est estimée à partir des données, une pratique courante est de remplacer la transformation inconnue par son estimateur et les inférences habituelles sont obtenues comme si la transformation estimée étaient la vraie. Dans cet article, nous prouvons que les limites percentiles suivant une estimation de la transformation n'ont pas une probabilité d'inclusion correcte, même lorsque la taille de l'échantillon n est grande. Une simple correction analytique est donnée pour assurer que les limites percentiles corrigées ont une probabilité d'inclusion correcte quand n est grand. Des simulations de Monte Carlo prouvent que les limites percentiles corrigées performant bien quand n n'est pas grand. En comparant les limites percentiles fréquentistes existantes pour une distribution particulière de durée de vie telle que Weibull, gamma, lognormal ou inverse gaussien, les nouvelles limites percentiles sont équivalentes ou meilleures. Les nouvelles limites percentiles sont attrayantes en pratique à cause de leur simplicité et de leur généralité.

Mercredi 13 Juin, 4h45 – 5h00, AQ 4120

Approches Linéaires Et Non Linéaires De Modèle Mixte À L'Analyse De « Cosinor »
Susan MIKULICH, Gary O. ZERBE and Thomas J. CROWLEY (University of Colorado)

La recherche utilisant l'analyse de cosinor suggère que les activités humaines reflètent des aspects de la biologie fondamentale des désordres psychiatriques multiples. Le modèle de cosinor, appliqué aux données avec mesures répétées gérées par le cycle circadien et d'autres rythmes biologiques, est un modèle non linéaire dans les paramètres d'amplitude et d'acrophase caractérisant chaque harmonique. Le modèle de cosinor a également une forme intrinsèquement linéaire, c.-à-d. il peut être transformé en un modèle linéaire où l'amplitude et l'acrophase pour chaque harmonique peuvent être calculés comme des fonctions non linéaires des coefficients de régression linéaire estimés. Les études précédentes évaluant l'activité ont employé des méthodes désuètes pour estimer et comparer ces paramètres circadiens non linéaires. Ici une approche utilisant un modèle mixte pour l'analyse de cosinor est considérée, où les paramètres d'effet fixes peuvent entrer non linéairement comme acrophase et amplitude pour chaque harmonique ou linéairement, après transformation, aux coefficients de régression. En outre, les effets aléatoires peuvent entrer non linéairement en tant que déviations spécifiques aux sujets des acrophases et des amplitudes ou linéairement en tant que déviations spécifiques aux sujets des coefficients de régression. Nous considérons également le cas hybride, où les effets fixes entrent non linéairement tandis que les effets aléatoires entrent linéairement. Le concept des effets aléatoires linéaires en combinaison avec des effets fixes linéaires ou non linéaires peut être étendu pour inclure des limites harmoniques de haut niveau en tant qu'effets aléatoires (c.-à-d. ajustement de covariance de Rao dans le modèle mixte) pour améliorer la précision et la puissance à détecter des différences dans les groupes. L'application de la méthode delta aux modèles mixtes linéaires a produit des résultats qui étaient souvent identiques aux résultats des modèles mixtes non linéaires. En conséquence, en appliquant la méthode delta, l'analyse linéaire traditionnelle de cosinor peut souvent être employée pour estimer et comparer les paramètres d'intérêt non linéaires. Un des objectifs de cet article est de clarifier quand les deux méthodes sont identiques et lorsqu'elles sont différentes. Support: NIDA accorde DA11015 et DA09842

Environmental Toxicology

Jeudi 14 Juin, 8h30 – 9h00, Halpern
Models and Methods for Time-Dependent Bioassays
John HUBERT (University of Guelph)

In aquatic toxicity it is becoming more common to establish a dose-response relationship from a quantal bioassay which has been monitored over time. This response-dose-time relationship is essential for risk assessment issues. When there are associated covariables, there are at least two methods for finding the point and interval estimators for certain key parameters (such as the ED50). These methods will be explained through real data sets, and recommendations for their use will be clarified.

Jeudi 14 Juin, 9h00 – 10h00, Halpern
Statistical Issues in Canadian Pulp and Paper Environmental Effects
Abdel EL-SHAARAWI and Jim SHERRY (National Water Research Institute)

The Fisheries Act of 1992 required the Canadian pulp and paper mills to conduct an Environmental Effects Monitoring (EEM) study by April 1, 1996. Following the submission of the first report the mills are required to complete a cycle of EEM studies every three years. The study cycle consists of field (adult fish survey), laboratory (sublethal toxicity tests), and fish consumption (tainting) components. The laboratory component consists of a series of standard toxicity tests on aquatic organisms. A test is a dose - response experiment in which the responses of individuals exposed to a series of the pulp mill effluent concentrations are measured. An overview of the scope and objectives of the program will be presented along with the statistical issues related to its design and data analysis. Data sets will be used for illustrations.

Contributed Session: Assorted

Jeudi 14 Juin, 8h30 – 8h45, AQ 4150
Robustesse Des Procédures Basées Sur La Profondeur
Ivan MIZERA (University of Alberta)

La robustesse et la stabilité des procédures basées sur la profondeur de demi-espace (Tukey) (des contours de profondeur et des estimateurs maximum de profondeur dans l'emplacement, la régression, et d'autres modèles) et leurs conséquences possibles pour l'analyse de données seront discutées. Les critères étudiés incluent la continuité ponctuelle, la robustesse qualitative (continuité faible), le point de panne, et d'autres.

Jeudi 14 Juin, 8h45 – 9h00, AQ 4150

Estimation Flexible D'Une Distribution À Décalage De L'Effet D'Expositions Instantanées

Bernard RACHE, M. Abrahamowicz. and A.J. Sasco

Les études épidémiologiques doivent souvent tenir compte d'une période possible de décalage, c.-à-d. une période de temps entre une exposition et le changement ultérieur du risque. Les méthodes précédentes pour modéliser des décalages ont été plus ou moins empiriques, comptant davantage sur des hypothèses a priori que sur les analyses des données appropriées. Dans cette étude, nous abordons la question d'estimation de la distribution de différentes périodes de décalage pour des données classiques censurées à droite de l'intervalle de temps initial jusqu'à un événement. Le défi est que les décalages sont latents, dans le sens qu'un changement dans le risque n'est pas directement observable mais doit être inféré à partir d'événements observés, qui peuvent se produire plusieurs années plus tard. Nous proposons ici une méthode en trois étapes pour estimer la distribution des décalages : a) modéliser des changements dans le temps des taux de risque (HR), b) estimer les limites de la distribution de décalage, et c) reconstruire la distribution des différents décalages. La première étape utilise un modèle de régression non paramétrique basé sur une généralisation du modèle proportionnel de risques (PH) qui tient compte des changements non monotones du taux de risque dans le temps, en utilisant la fonction quadratique spline(1). Les changements dans le temps en HR est testée par le test du rapport de vraisemblance(LRT) qui compare l'ajustement aux données entre le modèle flexible et le modèle conventionnel de PH. Considérant qu'un décalage dans l'effet d'exposition ne peut pas exister sans changement de HR dans le temps, les étapes suivantes sont effectuées seulement si le LRT rejette l'hypothèse nulle de PH au niveau 0,05. La deuxième étape des analyses se concentre sur l'estimation des limites inférieures et supérieures de la distribution de décalage, t_1 et t_2 , de l'estimation de la fonction HR. t_1 et t_2 sont d'abord estimés en se basant sur la pente lisse locale de la fonction estimée dans cinq sous-ensembles aléatoires différents, mutuellement exclusif des données initiales. Le procédé quintuple de validation croisée est alors utilisé pour identifier les valeurs de t_1 et de t_2 qui maximisent la vraisemblance partielle de la validation croisée. Dans la troisième étape, la distribution des valeurs de décalage dans l'intervalle (t_1, t_2), est reconstruit à partir de la fonction estimée HR. L'approche proposée est évaluée par des données simulées et appliquées aux données de Framingham pour expliquer l'effet du décalage de la cessation de fumer sur le risque de maladie coronaires.

Jeudi 14 Juin, 9h00 – 9h15, AQ 4150

Évaluation Probabiliste De La Qualité D'Étude En Méta-Analyse

Shagufta SULTAN and Robert PLATT (McGill University)

La méta-analyse est un ensemble de procédures statistiques conçues pour accumuler des résultats expérimentaux et de corrélations à travers des études indépendantes qui traitent un ensemble de questions de recherches semblables. Elle utilise des résumés statistiques de différentes études comme points de repères. La qualité des études dans la méta-analyse peut varier considérablement. Afin de tenir compte de ces variations, la variation est décrite en assignant des scores de qualité aux études. Une façon normale d'incorporer des scores de qualité à une analyse est d'accorder un poids à chaque étude dépendamment de son score de qualité. Cette méthode a été employée par plusieurs auteurs et a été critiquée à cause de sa subjectivité. Pour améliorer à cette approche, Tritchler (1999) a présenté un modèle de probabilité pour l'effet de la qualité sur une mesure d'effet globale. Il a dérivé un certain nombre de méthodes de résumés et les a comparées pour des données simulées et les a également appliquées sur un jeu de données réel pour des modèles à effets fixes. Nous avons exécuté deux évaluations de l'approche de Tritchler. Nous avons comparé le modèle de probabilité à plusieurs autres modèles pour l'ajustement de la qualité dans une étude de simulations. Dans cette étude nous avons utilisé une distribution des scores de qualité qui reflète plus convenablement les scores de qualité dans des données de méta-analyse. En général, le modèle probabiliste donne peu de biais et de bonnes probabilités d'inclusion pour les intervalles de confiance. De plus, nous avons comparé le modèle probabiliste à plusieurs autres méthodes pour l'ajustement de la qualité en utilisant une étude empirique de 31 méta-analyses. Nous prouvons que l'ajustement, pour une étude de qualité utilisant l'approche de Tritchler, donne des différences importantes comparées à d'autres approches d'ajustement de la qualité.

Jeudi 14 Juin, 9h15 – 9h30, AQ 4150

**Activités De Collaboration Entre Les Organismes Gouvernementaux
De Statistiques Sur La Santé Canadien Et Américain**

Jane GENTLEMAN and Gary CATLIN (National Center for Health Statistics)

Le département statistique de Santé Canada et le centre national du gouvernement des États-Unis pour les statistiques de santé ont maintenant deux échanges annuels. Le but de ces rencontres est de faciliter la collaboration et la coopération entre les deux groupes et d'encourager l'interaction continue entre ces échanges. Ces contacts ont comme conséquence la synergie entre les deux groupes, dont les mandats et les missions se chevauchent considérablement. Tous les ans, les membres du personnel des agences, visitent l'autre agence, alternant entre Washington D.C et Ottawa, pour deux jours de discussion formelle et informelle. Cet entretien décrit certaines des activités communes qui ont résulté de ces échanges. En particulier, des plans pour une enquête téléphonique commune binationale seront décrits. L'étude sera faite, en même temps aux États-Unis et au Canada. Un questionnaire d'enquête de 20 minutes est développé conjointement par les deux groupes, et le test cognitif sera fait dans les deux pays. Statistique Canada rassemblera les données des deux pays en utilisant les opérations de collecte des données CATI. Les coûts du projet seront partagés par les deux agences. Les objectifs de l'étude incluent : (1) étude méthodologique sur des aspects scientifiques et pratiques de faire des enquêtes communes multinationales ; (2) développer et tester des questions identiques pour faire partie des , des enquêtes par entrevue de santé nationales des deux pays, de ce fait mettant en valeur la comparabilité des évaluations et facilitant l'analyse de collaboration ; et (3) choisir des parties d'étude tout en demeurant dans le cadre commun de l'organisation de mondiale de la santé pour surveiller et évaluer la mise en place de la santé pour tous, ce qu'a pour but le développement d'une enquête commune pour l'état de santé, qui sera utilisée dans différents pays.

Jeudi 14 Juin, 9h30 – 9h45, AQ 4150

Une Examen Du Bootstrap Pour Le Modèle De Risque Relatif De Cox

Wenyu JIANG and John D. KALBFLEISCH (University of Waterloo)

Dans le contexte du modèle de risque relatif de Cox, nous proposons des méthodes bootstraps pour donner des estimations par intervalles pour les paramètres de régression basés sur la statistique de score de vraisemblance partielle studentisée. La distribution de cette statistique est estimée en rééchantillonnant ses termes ou en rééchantillonnant les sujets qui ont échoué durant l'intervalle tout en gardant les paramètres de risque fixe. Nous effectuons des études de simulation pour comparer ces méthodes aux procédures suggérées par Efron et par Tibshirani qui utilise la distribution bootstrap d'une statistique de « Wald » basée sur l'estimateur de vraisemblance partiel. Les méthodes basées sur la fonction de score partiel donnent des intervalles de confiance plus précis que ceux qui sont basées sur l'estimateur de vraisemblance partiel. De plus, les méthodes de rééchantillonnage appliquées à la fonction de score sont plus avantageuses au niveau des calculs que les approches plus classiques.

Nonlinear Timeseries

Jeudi 14 Juin, 10h30 – 10h50, Halpern

Concepts De Dépendance Non Linéaire Et Non Stationnaire En Séries Chronologiques

Dag Bjarne TJOSTHEIM (University of Bergen, Norway)

La différence entre l'analyse de séries chronologiques linéaires et non linéaires est illustrée par deux exemples. Le premier exemple concerne l'utilisation de concepts de dépendance non linéaire dans l'alignement de séries chronologiques. Les mesures de dépendance sont basées sur l'entropie et les fonctions de distance de Hellinger, et sont appliquées aux chronologies en varve. Le second exemple illustre la dépendance non linéaire entre des séries chronologiques non stationnaires. Il est montré comment la théorie des chaînes de Markov peut être utilisée pour obtenir une théorie asymptotique pour des estimés non paramétriques dans un tel contexte. La motivation vient du problème de co-intégration non linéaire en économétrie, mais cette méthode peut être appliquée à des systèmes d'entrée-sortie plus généraux des processus non stationnaires.

Jeudi 14 Juin, 10h50 – 11h10, Halpern

Séries Chronologiques Non Linéaires: Quelques Aspects Pour La Détection De Modèles Et Tester Leur Adéquation

Reg J. KULPERGER and Hao YU (University of Western Ontario)

Les processus ARHC et ARHCG sont parmi les séries chronologiques non linéaires les plus utilisées, avec plusieurs applications en séries chronologiques économétriques. Les principales hypothèses ou caractéristiques de ces modèles sont (i) un processus i.i.d. déterminé par les erreurs, habituellement gaussienne et (ii) la structure du terme de variance hétéroscédastique conditionnelle. Est-ce que ces hypothèses peuvent être testées statistiquement avec une quantité suffisante de données? Pour ce, nous examinerons le processus de FRE des résidus ajustés. Le processus gaussien limite est obtenu, et il constitue un pont brownien auquel des termes de perturbation ont été ajoutés. Il est surprenant de noter que cette limite est plus facile à obtenir que le processus résiduel FRE correspondant à un processus linéaire tel un processus AR. Une autre question, aussi reliée, est l'estimation lorsque le modèle est mal spécifié. Des travaux précédents sur un processus FRE aléatoirement pondéré seront revus. Ce processus est utile pour détecter la non linéarité dans des séries chronologiques

Jeudi 14 Juin, 11h10 – 11h30, Halpern

Modèles De Séries Chronologiques À Paramètres Variables

Wai Keung LI (Hong Kong University)

Les deux dernières décennies ont connu une croissance rapide de la variété de modèles de séries chronologiques non linéaires. Cela inclus les modèles de seuil, les modèles de coefficients fonctionnels, les modèles de type ARHC(G), les modèles à branchements markoviens, les modèles à index unique, et les modèles de séries chronologiques de type mixte. Tous ces modèles peuvent être regroupés sous la bannière des séries chronologiques à paramètres variables. La modélisation de tels modèles peut être entreprise en utilisant une approche semi-paramétrique. Dans cette présentation nous allons faire un survol de quelques développements récents, en mettant l'accent sur la modélisation. Quelques exemples réels seront donnés.

Jeudi 14 Juin, 11h30 – 11h50, Halpern

Tests Non Paramétriques Pour L'Hétéroscédasticité Conditionnelle Dans Les Séries Chronologiques

Min CHEN (Chinese Academy of Sciences)

Une caractéristique dominante des modèles ARHC (autorégressif hétéroscédastique conditionnel) et des modèles ARHCG (autorégressif hétéroscédastique conditionnel généralisé) est la variabilité de la variance conditionnelle. Cependant, il n'est pas facile de déterminer si une série chronologique particulière possède cette caractéristique. Deux tests paramétriques ont été proposés pour détecter cette dernière, mais ils dépendent fortement de la structure paramétrique spécifiée par l'utilisateur et de l'hypothèse de normalité des erreurs. Dans cette présentation nous décrirons une approche non paramétrique pour détecter l'existence de variabilité pour la variance conditionnelle.

Diagnostic Testing

Jeudi 14 Juin, 10h30 – 11h00, Images

Modèles, Méthodes Et Inférences Pour La Prévalence Et La Précision De Tests Basés Sur Des Tests De Dépistage Imparfais-Dépendants

Wesley O. JOHNSON (University of California-Davis)

Une hypothèse commune posée lors de l'estimation de la sensibilité et de la spécificité de deux (ou plus) tests de diagnostics binaires en l'absence d'un standard est celle de l'indépendance conditionnelle entre les tests selon un statut de la maladie donné. Même si elle est raisonnable dans certains cas, cette hypothèse est souvent non fondée ou non testée et peut mener à des résultats biaisés. Nous proposons une classe de modèles hiérarchiques pour l'estimation de la précision de deux tests et la distribution de la prévalence en l'absence d'un standard, lorsque plusieurs populations interchangeables avec différentes prévalences de maladies sont disponibles pour échantillonnage. Ceci a pour effet de diminuer l'importance de l'hypothèse d'indépendance conditionnelle entre les tests. Nous mettrons l'accent sur l'estimation des distributions de prévalence dans des populations. De plus, nous discuterons d'un modèle plus simple lorsqu'un nombre restreint de populations non interchangeables sont disponibles pour échantillonnage. (Ce travail est fait conjointement avec Tim Hanson, Marios Georgiadis, Ramanpreet Singh et Ian Gardner)

Jeudi 14 Juin, 11h00 – 11h30, Images

Efficacité De La Mise En Commun D'ADN Pour L'Estimation De Fréquences Conjointes D'Allèles Et Pour La Mesure De Déséquilibre De Liaison

Ruth PFEIFFER (National Cancer Institute)

La mise en commun d'échantillons d'ADN pour analyse remplit deux objectifs : le criblage de grands nombres d'individus pour estimer la prévalence de variantes génétiques, et l'identification des individus aux variantes. Nous étendons des méthodes d'analyse d'échantillons d'ADN mis en commun à l'estimation de prévalence conjointe des variantes à deux (ou plus) loci. Cette probabilité conjointe est la base de la compréhension des risques de maladies causés par l'exposition conjointe aux variantes génétiques, tel les effets d'une variante commune sur une mutation pénétrante rare. Les prévalences conjointes d'allèles à différents loci sont aussi requises pour estimer la mesure normale D de déséquilibre de liaison, un paramètre fondamental en génétique des populations et dans la détermination de la puissance des études d'association. Nous montrons comment estimer les fréquences d'allèles et le coefficient d'association allélique des échantillons mis en commun. De plus, nous quantifions les économies espérées en nombre d'analyses requises lorsque nous mettons les échantillons en commun versus le nombre d'analyses requises si chaque individu est testé individuellement. Dans le but de simplifier les protocoles en laboratoire, nos estimés sont basés sur un plan de mise en commun à une étape. De plus, nous laissons une marge d'erreur dans les tests et nous fournissons des estimateurs simples dans le cas de variantes rares. Nous prenons en considération les analyses pouvant déterminer si un individu est porteur de certains allèles (statut de porteur) de même que les analyses pouvant déterminer les génotypes conjoints. La réduction dans la taille de l'échantillon attendue obtenue par la mise en commun d'échantillons est démontrée pour les estimations de prévalence et pour l'identification d'individus avec quelques mutations rares, et nous étudions les caractéristiques d'opération de ces procédures par l'intermédiaire de simulations.

Jeudi 14 Juin, 11h30 – 12h00, Images

Estimation Bayésienne De Taille D'Échantillon Pour Études Impliquant Deux Tests Diagnostiques

Nandini DENDUKURI (St. Mary's Hospital)

Les calculs de taille d'échantillons pour l'estimation de la prévalence de maladies ou des mesures de précision de diagnostiques, telles que la sensibilité et la spécificité, sont usuellement basés sur des méthodes fréquentistes de proportions. Cette approche implique l'énonciation d'hypothèses par rapport aux valeurs exactes des paramètres d'intérêt, inconnus ou partiellement connus lors de la planification de l'étude. De plus, les incertitudes entourant toutes les autres valeurs des paramètres sont habituellement ignorées lors du calcul de taille d'échantillon pour un paramètre donné, menant à des tailles d'échantillon beaucoup trop petites. Nous étendons l'approche bayésienne de Rahme, Joseph et Gyorkos (2000) à l'estimation de taille d'échantillon lorsque deux (ou plus) tests imparfaits sont utilisés. Nous utilisons trois critères différents basés sur les intervalles de crédibilité a posteriori marginaux: 1) pour obtenir une couverture moyenne pré-déterminée pour une longueur fixe, 2) pour obtenir une longueur moyenne pré-déterminée pour une couverture fixe, 3) pour obtenir une couverture minimale pour une longueur pré-déterminée. La présence d'un second test permet de diminuer considérablement la taille de l'échantillon par rapport à une situation où un seul test est disponible. Par contre, l'augmentation de l'incertitude a priori dans n'importe lequel des paramètres va considérablement augmenter le besoin en taille échantillonnale. Dans certains cas, même un échantillon infiniment grand ne suffit pas à obtenir la précision voulue. Nos méthodes seront appliquées à une étude impliquant deux tests non standard pour l'infection aux strongyloïdes.

Contributed Session: Assorted

Jeudi 14 Juin, 10h30 – 10h45, AQ 4150

Intervalle De Confiance Pour La Dose Sujet À Une Certaine Augmentation De La Réponse À Cette Dose

Xiaoli HOU (ASA)

Des expériences avec réponse à une dose sont largement répandues dans la recherche scientifique. La dose d'un composé associé à une certaine augmentation de la réponse à cette dose est habituellement critique dans ce type d'expériences et se doit d'être estimé. Une méthode générale est présentée pour établir un intervalle de confiance pour la dose sujet à une certaine augmentation de la réponse à cette dose.

Jeudi 14 Juin, 10h45 – 11h00, AQ 4150

Contrôle Du Trafic Basé Sur Un Concept D'Ergodicité

Otto MOESCHLIN (Fern Universität)

Une contrôle optimal des feux de circulation signifie un débit maximal par unité de temps aussi bien que des temps d'attente minimaux et juste pour les participants du trafic. Comme prototype pour des formes bien plus compliquées de l'organisation du trafic comme des jonctions et des ronds points est la commande dite du goulot d'étranglement asymétrique, qui signifie que les processus d'arrivée des deux côtés sont différents. L'approche donnée ici est orientée sur une théorie de la décision; elle utilise un théorème de point fixe pour des opérateurs sur des espaces spéciaux de mesures de probabilité qui assure l'ergodicité du processus de queues.

Jeudi 14 Juin, 11h00 – 11h15, AQ 4150

Propriétés De Structure D'Un Processus De Bandit

Xikui WANG (University of Manitoba)

L'exposé sera concentré sur l'exploration des propriétés de structure de la stratégie optimale pour un processus de bandit avec des réponses retardées. L'avantage d'utiliser ces propriétés dans le calcul et la simulation est également discuté.

Jeudi 14 Juin, 11h15 – 11h30, AQ 4150

Estimation D'Une Probabilité De Bernoulli À Partir D'Échantillons Groupés Avec Des Applications Aux Taux D'Infection Par Un Virus Transmis Par Les Moustiques

Brad BIGGERSTAFF (US Centers for Disease Control and Prevention)

Les méthodes de vraisemblance sont employées pour estimer une probabilité p de Bernoulli où des observations ont été divisées en groupes et la réponse agrégée de chaque groupe est enregistrée comme un succès ou un échec. La méthodologie actuelle suppose que la taille des groupes soit égale, souvent une hypothèse peu réaliste dans la pratique, ainsi des méthodes sont développées pour permettre aux tailles des groupes de différer. Une forme pratique et itérative de l'estimateur du maximum de vraisemblance (MLE) de p est donnée, et sa variance asymptotique est dérivée. Les 3 intervalles de confiance habituels basés sur la vraisemblance ---Wald, le rapport de vraisemblance, et score--- sont calculés. Des estimateurs simplifiés ignorant le groupement ou utilisant la taille moyenne des groupes sont également considérés, et tous les estimateurs et intervalles de confiance sont évalués dans une étude de simulations. Les méthodes sont démontrées pour une étude du taux d'infection du virus du « West Nile » dans les moustiques échantillonnés dans Queens, New York, 1999.

Jeudi 14 Juin, 11h30 – 11h45, AQ 4150

Optimalité Des Tests Non Paramétriques Pour L'Association Entre Le Temps De Survie Et Les Covariables Continues

Karen LEFFONDRÉ, J. LELLOUCH, C. COM-NOUGUÉ, and T. MOREAU (McGill University Health Center)

Liu et al. (JASA, 1993) ont proposé une classe de tests du logarithme des rangs pondérée pour des hypothèses alternatives ordonnées dans l'analyse de survie. Nous proposons d'étudier cette classe de tests non paramétriques dans le cas de covariable continue. Les tests non paramétriques, qui remplacent les variables dépendantes et les variables indépendantes par leurs rangs, ont comme avantage de ne pas dépendre du choix de l'échelle de mesure de la covariable, et ils sont moins sensibles aux valeurs aberrantes qui se trouvent dans l'espace de covariable, que le modèle semi-paramétrique de Cox. Une reformulation des tests statistiques, faite en utilisant le rang de la covariable, nous permet de dériver un nouveau modèle multivarié et de présenter un nouveau procédé pour identifier l'alternative pour laquelle chaque test de la classe est optimal. Ceci diffère des études habituelles faite sur l'optimalité, c'est-à-dire nous étudions l'alternative pour laquelle un test donné est optimal, au lieu d'étudier quel test est optimal pour une alternative donnée. En utilisant cette nouvelle approche, nous prouvons que les notions d'ordre du logarithme des rangs et les tests de Gehan sont optimaux pour certaines proportions de risque et des cotes de risques alternatives respectivement. Une étude de simulation compare la taille et la puissance de ces deux tests avec ceux du rang des logits et avec ceux du tests de scores de Cox, sous les alternatives identifiées par notre approche. Les résultats de simulation sont en accord avec les résultats d'optimalité théoriques et suggèrent que l'ordre du test des logarithme des rangs puisse être un sérieux concurrent pour le rang des logits et pour le test de score de Cox, même pour des risques proportionnels alternatifs.

Jeudi 14 Juin, 11h45 – 12h00, AQ 4150

Neo-Adjuvant Hormone Therapy + Radiation For T3, Margin Positive Prostate Cancer
Eric AGYEKUM, C. LUDGATE et al (University of Victoria)

Increasing patient selection has resulted in a decrease in post operative positive margins, however 40-80% of patients are likely to have residual cancer following radical surgery (Solway, CUOG). Following prostatectomy the prostate specific antigen (PSA) level should become undetectable. A rise in PSA signifies cancer recurrence. Radiation therapy to the prostate bed has become a standard treatment for definitely positive margin disease, however the results in terms of PSA disease free survival following radiation have varied from 53 to 98 % depending on risk factors. This appears better than surgery alone, with 75% of patient developing a rising PSA by 5 years (Partin). The survival benefit of post operative RT however remains unproven, and survival benefit may be limited to those treated with undetectable PSAs and/or a low Gleason score. This presumes that all residual disease is in the prostatic fossa, but if the cancer has already grown through the capsule, then it is behaving in an aggressive way, and the failure of local treatment could be due to cancer spread. The Messing study, combining immediate hormone therapy for node positive disease, shows a control and survival advantage for immediate hormone therapy. The EORTC study of locally advanced disease (Bolla), also showed an advantage for early HT in high risk patients. At the Vancouver Island Cancer Centre from 1992 onwards, there were delays of several months in starting radiation therapy due to resource limitations. Concerns over the possibility of disease progression during the radiotherapy wait times led to the use of hormone therapy to try and prevent disease progression. (Ludgate). We present the results of patients managed in this way.

ABRAHAM, 2, 16
ADJENGUE, 10, 33, 41
AGEYKUM, 14
AGYEKUM, 32, 93, 143
AHMED, 10, 33, 41
ALAVI, 27, 81, 129
ALEXANDER, 21, 70, 117
ALI, 15
ANDERSON, 23, 37, 46, 73, 120
ANGERS, 2, 24, 28, 75, 84, 122, 133
ARCARO, 17, 59, 104
ARNOLD, 24, 76, 123
BANKIER, 13, 51, 95
BARMİ, 24, 75, 122
BARON, 21
BEDRICK, 13
BEGG, 17, 61, 105
BÉLAND, 18, 63, 109
BELL, 10, 33, 41
BELLAVANCE, 2
BELLHOUSE, 10, 18, 20, 33, 41, 69, 116
BENTKUS, 26, 79, 127
BÉRARD, 18, 63, 108
BERGERUD, 25, 78, 125
BERNARD, 22, 30, 71, 88, 118, 138
BESAG, 14, 53, 97
BHATTACHARYA, 26
BIGGERSTAFF, 31, 92, 142
BILBY, 23, 72, 119
BINGHAM, 20, 67, 113
Bingshu CHEN, 29, 86, 134
BLEUER, 27, 81, 129
BORCHERS, 25, 77, 124
BOUDREAU, 17, 61, 106
BOURBEAU, 18, 64, 109
BRADSHAW, 25, 77, 124
BRAUN, 10, 33, 42
BRILLINGER, 10, 34, 42
Bruce SMITH, 12, 20, 30, 39, 48
BULL, 25, 64, 110
BURTON, 2, 16, 28
CADIGAN, 10, 34, 42
CAHILL, 28, 84, 132
CARRIERE, 17, 60, 105
CATLIN, 30, 89, 139
CHAPMAN, 17, 60, 104
CHEN, Gemai, 30
CHEN, Hanfeng, 24, 76, 123
CHEN, Hong, 26, 79, 127
CHEN, Jiahua, 24, 76, 123
CHEN, Min, 31, 90, 140
CHEN, Ming-Hui, 9
CHENG, Shaowei, 20, 67, 113
CHIPMAN, 9
CHIPMAN, 10, 29, 34, 42, 86, 134
CHIU, Grace, 28, 85, 133
CHOI, Yun-Hee, 15
CHONG, Amos, 25
CHOULAKIAN, 17, 61, 105
CHUNG, Kelly, 26, 79, 127
CINTORA, 15
COHEN, 27, 82, 130
COOK, 17, 60, 105
COOK, 23, 29, 74, 86, 121, 134
Courchesne, 2
COWEN, 10, 34, 43
CRESSIE, 13, 51, 94
CROWLEY, 19
CROWLEY, Thomas J., 29, 87, 136
CZADO, 24, 75, 122
DALANG, 28, 83, 131
DALE, 14, 54, 98
DAMIANAKOS, 15
DARLINGTON, 24
DATTA, 19, 65, 111
DEAN, 14, 38, 47, 54, 97
DELAMPADY, 28, 84, 133
DENDUKURI, 31, 91, 141
DESMEULES, 26, 79, 127
DIAZ, 25, 76, 123
DICK, 13, 52, 95
DIDIER, 17, 59, 104
DOBBIE, 19, 65, 111
DORNER, 16, 57, 101
DRAGOMIR, 10, 35, 43
DRESSEL, 27, 82, 130
DU, Juan, 25
DUCHESNE, 18, 64, 109
DUCHESNE, 20, 69, 115
DUDOIT, 16, 58, 102
DUNN, 11, 35, 43
DUNSIGER, 25
Elizabeth A. THOMPSON, 9
ELLISON, 18, 77, 124

EL-SHAARAWI, 30, 88, 137
ELTINGE, 9, 19
ESCOBAR, 16, 57, 102
ESTERBY, 30
EVANS, 18, 26
FALL, 16, 57, 101
FARRELL, 2, 13, 20
FATTEDAD, 26, 79, 126
FAY, 13, 52, 59, 65, 95, 103, 111
FEIST, 23, 72, 119
FU, Audrey, 15
FU, C., 29, 87, 135
GENEST, 30
GENTLEMAN, 30, 89, 139
GILL, 11, 35, 44
GOLD, 25
GOMBAY, 24, 75, 122
GONEN, 17, 60, 104
GOTO, 26, 79, 127
GREENWOOD, 13
GRUNWALD, 24, 74, 121
GU, Hong, 29, 86, 134
GUSTAFSON, 18, 62, 107
HAMILTON, 23
HANSON, 31, 91, 140
HAREZLAK, 15, 56, 100
HASAN, 15
HAZIZA, 16, 59, 103
HE, Fangliang, 14
HE, Fangliang, 24, 25
HIDIROGLOU, 16, 17, 59, 103
HIDIROGLOU, Mike, 21, 70, 116
HOLMES, 26, 79, 127
HONG, Yongmiao, 20, 69, 115
HOOPER, 30
HOU, Xiaoli, 31, 92, 141
HU, Jack, 20, 67, 113
HUANG, 15
HUBERT, 30, 88, 137
HUNTER, 21, 70, 116
IBRAHIM, 9
JACROUX, 27, 80, 128
JIANG, Wenyu, 30, 89, 139
JOCELYN, 17, 59, 104
JOHNSON, 31, 91, 140
JONES, 24, 74, 121
Julian BESAG, 53, 97
K. SONG, 24, 74, 121

KALBFLEISCH, 24, 30, 76, 89, 123, 139
KARUNAMUNI, 24, 75, 122
KHAN, 29, 86, 134
KHAN, 15
KHOSHNEVISAN, 28
KIM, 11, 25, 35, 44
KIM, Jeong-ae, 25
KIM, Jong Sung, 15, 55, 99
KOULIS, 15
KOVACEVIC, 14, 53, 97
KOVACEVIC, 11, 27, 44
KULPERGER, 30, 90, 140
KURAMOTO, 15
LAHIRI, 26, 80, 127
LAIRD, 19, 66, 112
LARDJANE, 11, 35, 44
LAURITZEN, 16, 57, 102
LAZZERONI, 16, 58, 103
LEBLANC, 24, 75, 122
LEE, Jack, 11
LEE, Sophia, 15
Lefebvre, 2
LEFEBVRE, 25
LEFFONDRÉ, 32, 92, 142
LELE, 14, 53, 96
LELLOUCH, 32, 92, 142
LERTZMAN, 16, 57, 101
LESPERANCE, 2
LEVERT, 25
LEVESQUE, 29, 87, 135
LEWINGER, 11, 36, 45
LI, Wai Keung, 31, 90, 140
LI, Wei, 20, 67, 113
LI, William, 20, 67, 113
LI, Xiaoqiang, 25, 76, 123
LIE, 11, 36, 45
Liqun WANG, 29, 87, 135
LIU, Jinnan, 15
LOCKHART, 28, 85, 133
LOGAN, 18, 62, 107
LOHR, 19, 65, 111
LU, Xuwen, 15, 56, 100
LUAN, 24
LUDGATE, 32, 93, 143
LUNDEVALLER, 11, 36, 46
MA, Renjun, 29, 86, 134
MAAG, 25
MACDONALD, 15, 24

MACKAY, 20, 67, 113
Macmahon, 25
MACNAB, 23
MAJUMDAR, 26
Makhzoum, 2
MANDAL, 17, 60, 105
MANTEL, 14, 53, 97
MANTEL, 28
MCKELVEY, 25, 77, 124
MCLEISH, 20, 28, 69, 115
MCLEOD, 20
MEEKER, 16, 57, 102
MEIER, 19, 37, 64, 110
MIKULICH, 29, 87, 136
MIREA, 25, 64, 110
MIYAOKA, 11, 37, 46
MIZERA, 30, 88, 137
MOESCHLIN, 31
MOREAU, 32, 92, 142
MORGAN, 26
MUKERJEE, 24, 75, 122
MURPHY, 15, 55, 66, 99
MURPHY, 19, 112
NAWA, 24
NELSON, 19
NIYONSENGA, 11, 37, 46
NOBREGA, 28
NORCROSS, 27, 82, 131
NORRIS, 21, 71, 118
NOTZ, 27, 80, 128
NOUGUÉ, 32, 92, 142
ORLICH, 15, 55, 99
PANAGEAS, 17, 60, 104
PAUL, 11, 38, 47
PERKINS, 28
PESS, 23, 72, 119
PFEIFFER, 31, 91, 141
PINKSE, 23, 73, 120
PINTO, 12, 38, 47
PRASAD, 19, 65, 111
PUTERMAN, 26
QIAN, Hong, 25
QIU, Zhenguo, 24, 74, 121
QU, Annie, 12, 38, 47
QUINN, 26, 78, 126
RACHE, 30, 88, 138
RAJARSHI, 14, 53, 97
RAMSAY, 12, 38, 48
RAO, J.N.K., 7, 17, 20, 21, 59, 69, 70, 103, 116
RAO, Sunil, 25, 76, 123
RECK, 26, 80, 128
REED, 25, 29, 77, 124
REID, 19, 66, 112
REINERT, 26, 79, 127
REMPEL, 25
RICHARDSON, 16, 37, 57, 64, 102, 110
RICHARDSON, 19, 64, 110
RIVEST, 13, 23, 87, 135
ROBERTS, 21, 27, 52, 81, 96, 129
ROBINSON, 12, 39, 48
ROSE, 15
ROSENTHAL, 13, 52, 96
ROUTLEDGE, 23
ROY, 23
ROYCE, 18
ROZGA, 20, 68, 115
RUBIN, 19, 27, 81, 129
SAÏD, 12, 39, 48
SALEH, 12, 39, 48
SALTER, 15, 18, 25
SATAGOPAN, 17, 61, 105
SCHAFFER, 18, 62, 107
SCHIOPU-KRATINA, 28, 84, 132
SCHWARZ, 16, 26, 57, 78, 101, 126
SHAH, 21, 71, 117
SHERIDAN, 15
SHERRY, 30, 88, 137
SHI, Ellen, 24
SHRIVER, 20, 67, 113
SIT, 25
SITTER, 19, 21, 28, 69, 84, 116, 132
SLADE, 21, 71, 117
SMITH, Alex, 17, 60, 104
SMITH, Bruce, 12, 20, 30, 39, 48
SMITH, David, 19, 65, 111
SMITH, Isabelle, 26, 79, 127
SMITH, S., 25, 27
SPIEGELMAN, 18, 62, 107
SPINELLI, 18
STEEL, 23, 72, 119
STEELE, 21
STEINER, 20, 67, 113
STENGOS, 23, 73, 120
STEPHENS, Matthew, 16, 58, 103
STOICA, 13, 53, 96
STRUTHERS, 28, 85, 133
STRUTHERS, 19

- STUFKEN, 27, 81, 128
SUTHERLAND, 16, 57, 101
SWARTZ, Tim, 2
SYRJALA, 27, 83, 131
TAN, Ming, 24, 74, 121
TAVARE, 21, 71, 117
TAYLOR, 13, 52, 96
TERPSTRA, 20, 68, 114
THOMAS, 10, 16, 18, 29, 33, 41, 57, 64, 83, 87,
102, 109, 131, 136
THOMPSON, Elizabeth, 16
THOMPSON, Elizabeth A., 9
THOMPSON, Mary, 13, 22, 71, 118
THOMPSON, Mary, 17
TJOSTHEIM, 30, 90, 139
TOOZE, 24, 74, 121
TROUPÉ, 14, 55, 99
TSAO, 14, 54, 98
TURNER, 12, 39, 49
TYLER, 14, 54, 98
UDEVITZ, 15, 55, 100
URBANOVICH, 26, 79, 126
URRUTIA, 12, 40, 49
VAILLANT, 14, 55, 99
VAQUERA, 24, 76, 123
VENKATRAMAN, 17, 61, 105
VERBEL, 17, 61, 105
VILLASEÑOR, 24, 76, 123
WALSH, 28
WANG, Peiming, 28, 85, 133
WANG, Xikui, 31, 92, 142
WANG, Yuanyuan, 16, 57, 101
WANG, Zilin, 12, 40
WEI, Wei, 17, 60, 105
WELDON, 12, 40, 50
WELSH, 19, 65, 111
WENCZLER, 15
WHITRIDGE, 18, 63, 108
WIENS, 14, 54, 98
WILLIAMS, 25
WOOD, 23, 72, 119
WORLSEY, 23, 72, 119
WU, Changbao, 28, 84, 132
WU, Jeff, 21, 69, 116
WYSOCKI, 18, 64, 109
XING, Baifang, 15
XIONG, Jian, 15
YANG, Zhenlin, 29, 87, 135
YI, Grace, 23
YIN, Frank, 15
YOUNG, 26, 79, 126
YU, Eugenia, 15
YU, Hao, 30, 90, 140
YUNG, Wesley, 17, 59, 103, 104
ZAREPOUR, 12, 40, 50
ZERBE, 29, 87, 136
ZHANG, S., 24, 75, 122
ZHANG, Shenghai, 15, 56, 101
ZHANG, Xiaoling, 12, 40, 50
ZHANG, Ying, 20, 68, 114
ZHANG, Z., 26, 78, 125
ZHOU, Julie, 2, 14
ZHU, Hongtu, 24, 74, 121
ZIDEK, 13, 51, 94
ZWIERS, 15