SSC 2000 – Programme – ASSQ/ISQ

Université d'Ottawa University of Ottawa
June 4 2000 — June 7 2000

Contents

Welcome

The organizing committee of the Year 2000 SSC/ASSQ/ISQ conference welcomes you to Ottawa.

Program	Duncan Murdoch (chair)
	Charmaine Dean (biostatistics)
	Mike Hidiroglou (survey methods)
	Hélène Bérard (survey methods)
	Françoise Tarte (colloque)
Local Arrangements	André Dabrowski (chair)
	Ed Hughes
	Anne-Marie Houle
	John Nash
	Rama Nair

The committees also wish to thank a number of individuals for their help in preparing for this conference; Julie Deschamps, Thierry Bédard, Charles Patrick, Brian Allen and Paul Cabilio. We also thank all the organizers and chairs of the workshops, sessions and conferences.

Sponsors

The organizers are pleased to acknowledge the major support of Statistics Canada for the conference, and the joint support of the CRM (centre de recherches mathématiques), the Fields Institute and PIMS (Pacific Institute for the mathematical sciences) for scientific sessions. We thank them for their assistance.

In addition the following have generously supported the conference financially or through the donation of resources. Their contributons were very welcome.

Glaxo-Wellcome
The Faculty of Science, University of Ottawa.
The Faculty of Medicine, University of Ottawa.
The Faculty of Graduate and Postdoctoral Studies, University of Ottawa.
The Statistical Society of Ottawa.

Exhibitors

The following exhibitors will be available Monday and Tuesday in the Terminus area (near the main conference halls and Café Solstice) from approximately 10h30 to 17h00. Please drop by!! In addition, SAS Institute will demonstrate its product on Tuesday at 12h00 in MNT203.

Mathsoft
Statistics Canada (2)
Minitab
Nelson Thomson Learning
John Wiley and Sons
McGraw-Hill Ryerson
Sas Insitute

General Information

Services

- **Conference Office** The Conference Office is found in room UCU205 (Jock-Turcot University Centre). You can obtain supplies, fax service and other services here.
- **Food** Most on-campus food services are found in the Jock-Turcot University Centre, although there are small stands and cafeterias throughout the campus.

Unicentre lower level	Pivik corner store
	Presto take-out food
	Nox student bar
Unicentre second level	Cafeteria
Cumberland Ave.	Nostalgica restaurant

  Commercial Restaurants can be found along King Edward (Mexican, pizza, hamburger restaurants) and Laurier Avenue (Pub, Subway, coffee shop, vegetarian).

- **Airport Bus** Tickets for the airport bus will be sold at registration and at the conference office. Airporter busses will leave from the front entrance to Pavillon Lamoureux Hall, on the Jean-Jacques Lussier Rd. side. These will take you directly to Ottawa airport. They will leave at 15h10 and 15h30 on Wednesday, only.

TBT	=	Pavillon Tabaret Hall
UCU	=	Centre universitaire Jock Turcot Unicenter
MRT	=	Pavillon Morisset Hall
MNT	=	Pavillon Montpetit Hall
NAC	=	Centre national des arts National Arts Centre
LMX	=	Pavillon Lamoureux Hall
KED	=	Mathematics and Statistics Department, 585 King Edward.

Committees and Meetings

Saturday June 3		
18h00-23h00	SSC executive meeting	Novotel hotel
Sunday June 4		
7h30-16h00	SSC Meetings	LMX 475/477
	7h30 Finance	B. Allen
	9h00 Publications	M.L. McLaren
	10h00 SSC Board	J. Kalbfleisch
Monday June 5		
7h00-8h30	Committee meetings	UCU Cafeteria B section
	0700 Biostatistics Executive	C. Dean
	0700 Survey Methods Executive	M. Hidiroglou
	0730 Program	P. Cabilio
12h00-13h30	Committee meetings	UCU Cafeteria B section
	1230 CJS Board	C. Genest
	1230 Public Relations	M.L. McLaren
	1230 Liaison	C. Patrick
	1230 Professional Development	S. Bartlett
	1230 Regional & Society Cooperation	C. Léger
1700	Biostatistics AGM	MNT 203
1700	Survey Methods AGM	MNT 202
Tuesday June 6		
7h30-8h30	Committee Meetings	UCU Cafeteria B section
	Publications	M.L. McLaren
	Biostatistics Executive	C. Dean
	Statistical Consulting	J. O'Hara-Hines
12h00-13h30	Committee meetings	UCU Cafeteria B section
	Finance	B. Allen
	Bilingualism	F. Bellavance
	Research	J. Petkau
	Statistical Education	S. Brown
	Caucus for Women in Statistics	C. Struthers
	Professional Development	S. Bartlett
15h30-17h00	ASSQ AGM	MRT205
17h00-18h30	SSC AGM	MNT202
Wednesday June 7		
7h30	SORA AGM R. Kulperger	UCU Cafeteria B section
12h00-13h30	NSERC S. Esterby	UCU Cafeteria B section
12h00-13h30	Business and Industrial Statistics	UCU Cafeteria B section
	— organizational meeting	B. Abraham
15h10	Bus to airport	LMX front door
15h30	Bus to airport	LMX front door

Job Fair

See the Poster Session or conference office for posting of appointments. All appointments will occur in the following locations.

Sunday	LMX 112 115 116 117 118 223, MRT 224B: 8h-18h
Monday	LMX 112 115 116 117 118 223, MRT 224B: 8h-18h
Tuesday	LMX 112 8h-11h30, 14h30-18h
	LMX 117 8h-11h00, 14h00-17h
	LMX 223 8h-17h30
	MRT 224B: 8h-18h
Wednesday	LMX 112 115 116 117 118 223, MRT 224B: 8h-18h

Outline of Events

Day	Time	Place	Event
Sunday	8:00–18:00		Job Fair Interviews
Sunday	9:30–11:00	MNT 202 hallway	Registration
Sunday	9:00–16:00	MNT 202	Biostatistics Workshop
Sunday	10:00–16:00	MNT 203	Survey Methods Workshop
Sunday	15:00–17:00	MNT 202 hallway	Registration
Sunday	16:00–19:00	MRT 218	CJS/CRM Read Paper Session
Sunday	18:30–20:00	TBT Rotunda	Registration
Sunday	19:00–22:00	TBT Chapel	Reception
Monday	8:00–9:00	TBT Rotunda	Registration
Monday	8:30–10:00	TBT Chapel	Session 1: Welcome and SSC Presidential Invited Address
Monday	10:00–10:30	UCU Promenade	Coffee
Monday	10:30–12:00	UCU Promenade	Session 2: Poster Session
Monday	10:30–17:00	UCU Terminus	Exhibitors' booths
Monday	13:00-14:00	MNT 202 hallway	Registration
Monday	13:30-15:00	MNT 202	Session 3: Confidentiality
Monday	13:30-15:00	MNT 201	Session 4: Longitudinal and Life-History Data Analysis
Monday	13:30-15:00	MRT 218	Session 5: Statistics in Business and Industry
Monday	13:30-15:00	MNT 203	Session 6: Positive Dependence and Ordering for Point Processes
Monday	13:30-15:00	MNT 207	Session 7: Graduate Student Presentations
Monday	13:30–13:45	MRT 205	Session 8: Opening of the Colloque
Monday	13:45–15:00	MRT 205	Session 9: Special session on Survey Sampling
Monday	15:00–15:30	UCU Terminus	Coffee
Monday	15:30–17:00	MNT 202	Session 10: A Panel on Training of Survey Statisticians
Monday	15:30–17:00	MNT 201	Session 11: Invited Address by Norman Breslow
Monday	15:30–17:00	MRT 218	Session 12: Panel Discussion on Accreditation
Monday	15:30–17:00	MNT 201	Session 13: Mathematical Modelling of TCP Control Flows Protocols
Monday	15:30–17:00	MNT 207	Session 14: Stochastic Processes and Time Series
Monday	15:30–17:00	MRT 205	Session 15: Survey Methodology
Monday	17:00–18:00	MNT 203	Biostatistics Section Annual General Meeting
Monday	17:00–18:00	MNT 202	Survey Methods Section Annual General Meeting
Monday	17:00–19:00	Café Solstice	Professional Development Mixer
Monday	18:30	UCU Terrasse	SSO Graduate Student Barbeque
Tuesday	8:30–10:00	MNT 207	Session 16: Bayesian and Monte Carlo Methods
Tuesday	8:30–10:00	MNT 201	Session 17: Design Issues
Tuesday	8:30–10:00	MNT 203	Session 18: Strong Limit Theorems
Tuesday	8:30–10:00	MNT 202	Session 19: Two-phase Estimation
Tuesday	8:30–10:00	MRT 205	Session 20: Statistical Software
Tuesday	10:00–10:30	TBT Rotunda	Coffee
Tuesday	10:30–12:00	TBT Chapel	Session 21: Gold Medal Address
Tuesday	10:30–17:00	UCU Terminus	Exhibitors' booths

Day	Time	Place	Event
Tuesday	12:00–13:30	MNT 203	SAS Technological Demonstration
Tuesday	13:30–15:00	MNT 202	Session 22: Edit and Imputation
Tuesday	13:30–15:00	MNT 203	Session 23: Biostatistical Applications of GEEs
Tuesday	13:30–15:00	MRT 218	Session 24: Statistics in Business
Tuesday	13:30–15:00	MNT 204	Session 25: Case Study: Data Mining
Tuesday	13:30–15:00	MNT 201	Session 26: Science and Statistics
Tuesday	13:30–15:00	MRT 205	Session 27: Survey Methodology
Tuesday	15:00–15:30	UCU Terminus	Coffee
Tuesday	15:30–17:00	MNT 202	Session 28: Time Series Methods Applied in Statistical Agencies
Tuesday	15:30–17:00	MNT 203	Session 29: Y2K National Health Statistics
Tuesday	15:30–17:00	MRT 218	Session 30: Directional Statistics
Tuesday	15:30–17:00	MNT 204	Session 31: Case Studies: Mixtures Plus
Tuesday	15:30–17:00	MNT 207	Session 32: Theoretical Survey Methods
Tuesday	15:30–17:00	MRT 205	ASSQ Annual General Meeting
Tuesday	17:00–19:00	MNT 202	SSC Annual General Meeting
Tuesday	19:00–1:00	NAC Tent	Banquet (cash bar at 19:00, dinner at 19:30)
Wednesday	8:30–10:00	MNT 207	Session 33: Applied Survey Methods - I
Wednesday	8:30–10:00	MNT 203	Session 34: Statistical Methods for Correlated Data
Wednesday	8:30–10:00	MNT 201	Session 35: Nonparametric Ranking Methods
Wednesday	8:30–10:00	MRT 218	Session 36: MCMC session with Peter Green
Wednesday	8:30–10:00	MNT 202	Session 37: Risk Estimation and Applications
Wednesday	8:30–10:00	MRT 205	Session 38: Methods for Teaching Statistics
Wednesday	10:00–10:30	UCU Terminus	Coffee
Wednesday	10:30–12:00	MNT 202	Session 39: Complex designs
Wednesday	10:30–12:00	MNT 207	Session 40: Graduate Student Presentations
Wednesday	10:30–12:00	MRT 218	Session 41: Probability and Markov Chains
Wednesday	10:30–12:00	MNT 203	Session 42: Large Multi-Centre Clinical Trials
Wednesday	10:30–12:00	MNT 201	Session 43: The Bootstrap, Wavelets, Nonlinear Regression, and Dynamical Systems
Wednesday	10:30–12:00	MRT 205	Session 44: Estimation, biostatistics and time series
Wednesday	12:00–13:30	UCU Cafeteria B	NSERC Luncheon Meeting
Wednesday	13:30–15:00	MNT 202	Session 45: Pierre-Robillard and CJS Awards
Wednesday	13:30–15:00	MNT 201	Session 46: Applied Survey Methods — II
Wednesday	13:30–15:00	MNT 207	Session 47: Correlated Data
Wednesday	13:30–15:00	MNT 203	Session 48: Nonparametric Bayesian Inference
Wednesday	13:30–15:00	MRT 218	Session 49: Probability, Design and Mathematical Statistics
Wednesday	13:30–15:00	MRT 205	Session 50: Survey Methodology
Wednesday	15:10	LMX Front door	Bus to airport
Wednesday	15:30	LMX Front door	Bus to airport

Scientific Programme

Sunday, June 4th

9:00–16:00, Montpetit Hall 202 Biostatistics

Biostatistics Workshop
Organizer: Hau Lei & Charmaine Dean
Chair: Charmaine Dean

Richard Cook & Jerry Lawless, University of Waterloo.
The Analysis of Event History Data

10:00–16:00, Montpetit Hall 203 Survey Methods

Survey Methods Workshop
Organizer and chair: Mike Hidiroglou

Sharon Lohr, Ohio State University.
Sampling: Design and Analysis

16:00–19:00, Morisset Hall 218 Canadian Journal of Statistics & Centre de recherche mathématique Special Session

CJS/CRM Read Paper Session
Organizer and chair: Christian Genest

Feifang Hu & John D. Kalbfleisch, National University of Singapore & University of Waterloo.
The Estimating Function Bootstrap

Tom DiCiccio, Cornell University.
Discussant

Christian Leger, Université de Montréal.
Discussant

Nancy Reid, University of Toronto.
Discussant

Rob Tibshirani, Stanford University.
Discussant

Jim Zidek, University of British Columbia.
Discussant

Monday, June 5th

8:30–10:00, Tabaret Hall Chapel

Special Session

Session 1: Welcome and SSC Presidential Invited Address
Organizer and chair: John D. Kalbfleisch

8:30 Jim Zidek, University of British Columbia.
Air Pollution and Health Risk: Some Statistical Problems and Solutions

10:30–12:00, Unicentre Jock-Turcot Promenade

Poster Session

Session 2: Poster Session

S. Ejaz Ahmed & Richard J. Mcintosh, University of Regina.
An Asymptotic Approximation for the Birthday Problem

S. E. Ahmed & A. I. Volodin, University of Regina.
On the Convergence Rates of the Bootstrapped Means

Naomi Altman* & Julio Villarreal, Cornell University.
Self-Modeling Regression with Random Effects using Penalized Regression Splines

Denise Babineau, Simon Fraser University.
Estimating the Population Size of Clams Using A Model Assisted Sampling Design and Analysis

Cynthia Bocci, University of Ottawa.
Linear Regression with Spatially Correlated Data

David R. Brillinger, University of California, Berkeley.
Space Debris: Flux in a Two Dimensional Orbit

Khaled El Emam & Shesh Rai, National Research Council of Canada & St. Jude Children's Research Hospital.
Evaluating Predictions of Faulty Software Using Thresholds

Claudia Émond, Chantal Mérette, Andrée Brassard & Chantal Caron, Université Laval.
Statistical Linkage Analysis Led to Strong Evidence of a Susceptibility Gene on Chromosome 11 for Gilles de la Tourette Syndrome

Krish Ghosh*, M. Prendergast, C. Soong, G. Matsueda, Bristol-Myers Squibb.
Dose-Time-Response Nonlinear Model of tPA-Mediated Plasma Clot Lysis in the Presence of Candidate Drugs

Zhenguo Qiu, York University.
BUGS Program for the Analysis of Multi-Level Repeated Ordinal Data using a Bayesian Hierarchical Model

K. Schulze, P.A. Groome, B. O'Sullivan, B. Grice, C. Goh, S.F. Hall, F. Liu, W.J. Mackillop, D. Payne, D. Rothwell, J. Waldron, P. Warde & B. Cummings, Queen's University and University of Toronto.
Measures for Comparing Head and Neck Cancer Staging Schemes

Carl James Schwarz, Simon Fraser University.
The Siren Call of the Schaeffer Estimator

Javid Shabbir, Quaid-I-Azam University.
Comparison of Ratio-Type Estimators under a Super Population Model

Philip J. Smith, U.S. Centers for Disease Control and Prevention.
A Method for Forming Nonresponse Adjustment Cells, Applied to Children without Vaccination Histories from Providers in the National Immunization Survey

ACPD/CSPP, SSC.
SSC Ad-Hoc Committee on Professional Development

George Tomlinson* & Keith O'Rourke, University of Toronto & University of Ottawa/Ottawa Hospital.
Combining Biased and Unbiased Estimates in a Meta-Analysis

Rolf Turner, University of New Brunswick.
Predicting Thickness of Pipe Walls Subjected to Flow Accelerated Corrosion

Marlos Viana, University of Illinois at Chicago.
Statistical Methods in Symmetry Studies

Zilin Wang, University of Western Ontario.
Diagnostic Checking for Linear and Nonlinear Time Series Data Fittings: An Application of Reordered Residuals Processes

Swarna Weerasinghe & Lamont Sweet, Dalhousie University.
Annual Trends in Rates of Asthma Hospital Admissions in Nova Scotia, Canada: Influence of Hospital Bed Reductions

Julie Zhou, University of Victoria.
Integer-Valued, Minimax Robust Designs for Approximately Linear Model with Correlated Errors

13:30-15:00, Montpetit Hall 202

Survey Methods
Invited Paper Session

Session 3: Confidentiality
Organizer and chair: Jean-Louis Tambay

13:30 John Horm, National Center for Health Statistics.
National Center for Health Statistics Approaches to the Release of Microdata: Data Perturbation and the Research Data Center

14:00 Sallie Keller-McNulty & Elisabeth A. Unger, Los Alamos National Laboratory and Kansas State University.
Cancelled

14:00 Jeanine Bustros, Statistics Canada.
Access to Microdata Files at Statistics Canada

13:30-15:00, Montpetit Hall 201

Biostatistics
Contributed Paper Session

Session 4: Longitudinal and Life-History Data Analysis
Chair: Fangliang He

13:30 Peter X.-K. Song, York University.
Marginal Dispersion Models for Longitudinal Data

13:45 Peiming Wang* & Martin L. Puterman, Nanyang Technological University, Singapore & University of British Columbia.
A Two-state Hidden Markov Poisson Model

14:00 Bingshu Chen & Richard J. Cook, University of Waterloo.
Test of Treatment Effects Robust to Dependent Censoring

14:15 Rhonda J. Rosychuk* & Mary E. Thompson, University of Alberta and University of Waterloo.
A Semi-Markov Model for Binary Longitudinal Responses Subject to Misclassification

14:30 W. Wei & R.J. Cook, University of Waterloo.
Efficiency Consideration in the Design of Clinical Trials with Recurrent Events

13:30-15:00, Morisset Hall 218

Invited Paper Session

Session 5: Statistics in Business and Industry
Organizer and chair: Bovas Abraham

13:30 Hugh Chipman, University of Waterloo.
Tree Models in Industrial Statistics

14:00 Mike Brajac, General Motors of Canada.
Design of Experiments Applied to the Development of an Alternative Fuel Injection System

14:30 Derek Bingham, University of Michigan, Ann Arbor.
Discussant

13:30-15:00, Montpetit Hall 203

Invited Paper Session

Session 6: Positive Dependence and Ordering for Point Processes
Organizer and chair: Gail Ivanoff

13:30 Robert M. Burton, Oregon State University.
Positively and Negatively Dependent Point Processes: Structures, Constructions, and Central Limit Theorems

14:15 Mathieu Plante, University of Ottawa.
Self-exciting Point Processes: Vector Representation and Kernel Monotonicity

13:30-15:00, Montpetit Hall 207

Contributed Paper Session

Session 7: Graduate Student Presentations
Chair: J. Terry Smith

13:30 M.K. Chung & K.J. Worsley, McGill University.
The Detection of Local Volume Change in Shape Analysis

13:45 Sanjoy K. Sinha*, Christopher A. Field & Bruce Smith, Dalhousie University.
Robust Spectrum Estimation in an ARMA(p, q) Model with Application to the Sea Level Data

14:00 Rong Zhu, University of British Columbia.
Continuous-Time Stochastic Processes Based on Generalizations of Binomial Thinning Operators for Modelling Count Data

14:15 Wayne Horn, Myron Hlynka & Percy H. Brill, University of Western Ontario & University of Windsor.
Laplace Transforms of Order Statistics of Erlang Random Variables

14:30 Minnie Horace*, Swarna Weerasinghe, David MacLean & Meng Tan, Dalhousie University.
Cardiovascular Disease Risk Factors among Bermudans in Bermuda: Comparison with Canadians Living in Nova Scotia Utilizing Existing Data

13:30–13:45, Morisset Hall 205 colloque "Méthodes et applications de la statistique 2000"

Session 8: Opening of the Colloque
Organizer and chair: Françoise Tarte

Yvon Fortin, directeur général de l'Institut de la statistique du Québec.
Welcome to the Colloque

13:45–15:00, Morisset Hall 205 colloque "Méthodes et applications de la statistique 2000"
Special Session

Session 9: Special session on Survey Sampling
Organizer and chair: Françoise Tarte

13:45 Paul-André Salamin, Office fédéral de la statistique, Suisse.
Inquiries Coordination for Businesses

15:30–17:00, Montpetit Hall 202 Survey Methods
Panel Discussion

Session 10: A Panel on Training of Survey Statisticians
Organizer: Karla Nobrega
Chair: J.N.K. Rao

David Bellhouse, University of Western Ontario.
Panellist

Louis-Paul Rivest, Université Laval.
Panellist

David Binder, Statistics Canada.
Panellist

Louise Bourque, l'Institut de la statistique du Québec.
Panellist

Tom Goss, Goss Gilroy Inc..
Panellist

15:30–17:00, Montpetit Hall 201 Biostatistics
Special Session

Session 11: Invited Address by Norman Breslow
Organizer and chair: Charmaine Dean

15:30 Norman Breslow, University of Washington.
The Value of Long Term Follow-up: Lessons from the National Wilms Tumor Study

15:30–17:00, Morisset Hall 218
Panel Discussion

Session 12: Panel Discussion on Accreditation
Organizer and chair: S. Bartlett

15:30 Ken McRae, Agriculture and Agri-Food Canada.
Introduction by SSC Professional Development Committee

15:45 Rolf Turner, University of New Brunswick.
Pros of accreditation

16:00 Steven Smith, Fisheries & Oceans Canada.
Cons of accreditation

16:15 Michel Lord, SAIC Canada.
Another experience: the physics model

16:30 Ken McRae, Agriculture and Agri-Food Canada.
Summary and open discussion

15:30–17:00, Montpetit Hall 201

Invited Paper Session

Session 13: Mathematical Modelling of TCP Control Flows Protocols
Organizer and chair: David McDonald

15:30 François Baccelli, École Normale Supérieure.
TCP is $(\max, +)$

16:30 David McDonald, University of Ottawa.
Discussant

16:40 Francois Theberge, University of Ottawa.
Discussant

16:50 Normand Glaude, Microlegend Corp..
Discussant

15:30–17:00, Montpetit Hall 207

Contributed Paper Session

Session 14: Stochastic Processes and Time Series
Chair: Reg Kulperger

15:30 Adriana Climescu, University of Ottawa.
Cramer-Hida Decomposition Applied to Problems of Random Signal Detection

15:45 Dano Kako, Université de Sherbrooke.
Estimation of Parameters of Multilevel Linear System of Stochastic Differential Equations

16:00 Ying-Sheng Hu, University of Houston.
The Optimal Approximation Idea of Classical CuSum Procedure for Level-Shift Problem

16:15 Arjun Gupta & Jie Chen, Bowling Green State University.
Information Theoretic Approach to Multiple Change-Points Analysis

16:30 Gemai Chen* & Min Chen, University of Regina and Chinese Academy of Sciences.
Kolmogorov Test of Nonlinearity in Time Series: Percentage Points and Power Study

16:45 A.K. Basu, Calcutta University.
Sequential Estimation, Confidence Sequences and Tests with Power One for Parameters of AR(P) and RCAR(1) Models

15:30–17:00, Morisset Hall 205 colloque "Méthodes et applications de la statistique 2000"
Contributed Paper Session

Session 15: Survey Methodology
Chair: Hélène Bérard

15:30 Yves Béland, Suzanne Lessard, Marianna Morano, Orane Saint-Denis & Sylvain Thivierge, Statistics Canada.
Design of the Canadian Community Health Survey (CCHS)

15:45 Sylvie Laroche, Statistics Canada.
Overview of the Longitudinal Study of Canadian Immigrants

16:00 Denis Malo, Statistics Canada.
Responses Obtained By Proxy: The Effect on Data Quality for the 1998 National Census Test

16:15 Serge Chevalier, Régie régionale de la santé et des services sociaux de Montréal-Centre.
Call Sheets Data of Telephone Survey - Utilisation and Validation

16:30 Jean-François Bastien & Laurent Roy, Statistics Canada.
Television Viewing: A New Method for Collecting Data

16:45 Jean-Sébastien Brien & Robert Courtemanche, Institut de la statistique du Québec.
Cancelled

Tuesday, June 6th

8:30–10:00, Montpetit Hall 207

Contributed Paper Session

Session 16: Bayesian and Monte Carlo Methods
Chair: Duncan Murdoch

8:30 François Perron, Université de Montréal.
Beyond Accept-Reject Sampling

8:45 Gary Sneddon, Memorial University of Newfoundland.
Generating Probability Distributions of Analyses using Ensembles of Forecasts and Bayesian Statistics

9:00 Fatemah Alqallaf, University of British Columbia.
Bayesian Cross-Validation Choice and Assessment of Statistical Models

9:15 Paul Gustafson* & Lawrence J. Walker, University of British Columbia.
Bayesian Analysis of Longitudinal Multinomial Data arising from Developmental Research

9:30 Nathan Johnson, University of British Columbia.
Bayesian Regression with Interactions and Smooth Effects

9:45 Keyue Ding & Rohana J. Karunamuni*, University of Alberta.
A Linear Empirical Bayes Method on the Problem of Multivariate Calibration

8:30–10:00, Montpetit Hall 201

Biostatistics
Contributed Paper Session

Session 17: Design Issues
Chair: Pina D'Angelo

8:30 Ka Ho Wu, Siu Hung Cheung & Man Lai Tang, The Chinese University of Hong Kong.
Sequential Multiple Testing of Several Binomial Proportions to a Specified Standard

8:45 Peter Kupchak, University of Toronto.
Optimal Designs for Detecting Drug Interactions

9:00 Jean-Pierre Bélisle & François Bellavance, École des Hautes Études Commerciales.
Active Control Equivalence Studies : A Review of Statistical and Ethical Issues

9:15 Rollin Brant, University of Calgary.
Design Considerations for Construction and Calibration of Health Measures

9:30 Benny Zee, Queen's University.
Design Issues of a Phase III Randomized Trial incorporating a Phase II Portion

9:45 J. Terry Smith, Queen's University.
Setting Standards: Experiences and Statistical Issues

8:30–10:00, Montpetit Hall 203

Invited Paper Session

Session 18: Strong Limit Theorems
Organizer and chair: André Dabrowski

8:30 Walter Philipp, University of Illinois, Urbana-Champaign.
Some Limit Theorems for Independent and Weakly Dependent Random Variables

9:15 Herold Dehling, Rijksuniversiteit Groningen.
Some New Limit Theorems for (Dependent) U-statistics

8:30–10:00, Montpetit Hall 202

Survey Methods
Special Session

Session 19: Two-phase Estimation
Organizer and chair: Mike Hidiroglou

8:30 Wayne Fuller, Iowa State University.
The Use of Regression in Multi-Phase Sampling

8:30–10:00, Morisset Hall 205

colloque "Méthodes et applications de la statistique 2000"
Invited Paper Session

Session 20: Statistical Software
Chair: Sylvain Végiard

8:30 Jean Hardy, Hardy Consulting Services.
Taking Advantage of the New Version of SAS

10:30–12:00, Tabaret Hall Chapel

Special Session

Session 21: Gold Medal Address
Organizer and chair: Jane Gentleman

10:30 Jerry Lawless, University of Waterloo.
Survival Analysis and Likelihood Methods

13:30–15:00, Montpetit Hall 202

Survey Methods
Invited Paper Session

Session 22: Edit and Imputation
Organizer and chair: Pat Whitridge

13:30 Michael Bankier*, Paul Poirier, Martin Lachance & Patrick Mason, Statistics Canada.
A Generic Implementation of the Nearest-Neighbour Imputation Methodology

14:00 Claude Poirier, Statistics Canada.
The Generalized Edit and Imputation System and its Recent Developments

14:30 Kara Perritt, US Department of Agriculture.
A Look into AGGIES, An Automated Edit and Imputation System

13:30–15:00, Montpetit Hall 203 Biostatistics
Invited Paper Session

Session 23: Biostatistical Applications of GEEs
Organizer and chair: John Koval

13:30 John Neuhaus, Alistair Scott & Chris Wild, University of California, San Francisco & University of Auckland.
The Analysis of Case-Control Family Data

14:00 Jeanette O'Hara Hines, University of Waterloo.
Handling MAR Drop-outs in Regression Models Analyzed Using GEE's

14:30 Brajenda C. Sutradhar, Memorial University of Newfoundland.
Efficiency Aspects of Working Correlations Based GEEs in Longitudinal Studies

13:30–15:00, Morisset Hall 218 CFW
Invited Paper Session

Session 24: Statistics in Business
Organizer and chair: Danielle Morin

13:30 Johanne Thiffault, Ministère des Transports du Québec.
Quality Control in Statistical Surveys

14:00 Alec Whitmore, McGill University.
Statistical Aspects of Surveillance in Business

14:30 Anne-Marie Croteau, Concordia University.
Structural Equation Modeling Applied to Strategic Alignment of Information Technology

13:30–15:00, Montpetit Hall 204
Case Studies

Session 25: Case Study: Data Mining
Organizer and chair: Peter Macdonald

13:30 Gary Saarenvirta, The Loyalty Group.
Introduction

13:45 Jamie Powers, Arlene Zarubick, Dean Rutty, Geoff Smale & Brian Mangal, University of Guelph.
Title TBA

13:57 Swetlana Ljubicic, Melissa Naglic, Peter Macdonald, McMaster University.
Title TBA

14:09 Mark Robinson, Fatemah Alqallaf, Dana Scott Aeschliman & Aditya Sharma, University of British Columbia.
Mining Magic

14:21 Shahab Raza, Abdul Rauf Khan & A. B. M. Idris, Carleton University.
Title TBA

14:33 Jiaqiong Xu, Yuanyuan Wang, Jagadish Rangrej, Serje Robidoux & Anais Badour, University of Waterloo.
Title TBA

Shizuko Takahashi Calvo, Rodrigo Juarez Y Ruiz, Patricia Medina Monroy & Lan Nhu Pham, University of Waterloo.
Poster presentation

14:45 Gary Saarenvirta, The Loyalty Group.
Discussion

13:30–15:00, Montpetit Hall 201

Invited Paper Session

Session 26: Science and Statistics
Organizer and chair: Agnes Herzberg

13:30 I. U. Ingold, National Research Council of Canada.
Vitamin E: Or Why We Don't Go Rancid

14:15 R. J. Le Roy, University of Waterloo.
Uncertainty, Precision and Rounding: How Spectroscopists Use/Abuse Regression

13:30–15:00, Morisset Hall 205 colloque "Méthodes et applications de la statistique 2000"

Special Session

Session 27: Survey Methodology
Chair: Pierre Lavallée

13:30 Yves Tillé, École Nationale de la Statistique et de l'Analyse de l'Information, France.
Sampling, Unequal Probabilities and Balancing

15:30–17:00, Montpetit Hall 202 Survey Methods
Invited Paper Session

Session 28: Time Series Methods Applied in Statistical Agencies
Organizer and chair: Marietta Morry

15:30 Zhao-Guo Chen* & Ka Ho Wu, Statistics Canada & Chinese University of Hong Kong.
Survey Error Modeling in the Presence of Benchmarks

16:00 Pierre Cholette, Statistics Canada.
Interpolating Monthly Values from Data covering Various Fiscal Periods

16:30 Ruth E. Detlefsen* & Michael Z. Shimberg, U.S. Census Bureau.
North American Industrial Classification System (NAICS) Time Series of U.S. Retail and Wholesale Estimates

15:30–17:00, Montpetit Hall 203 Biostatistics
Invited Paper Session

Session 29: Y2K National Health Statistics
Organizer and chair: Ping Yan

15:30 Jane Gentleman, U.S. National Center for Health Statistics.
The U.S. National Health Interview Survey

16:00 Lorna Bailie, Statistics Canada.
Investing in Health Surveys, Investing in Health

16:30 William Ross, Health Canada.
Statistics in the Health Protection Branch of Health Canada

15:30–17:00, Morisset Hall 218
Invited Paper Session

Session 30: Directional Statistics
Organizer and chair: Peter Kim

15:30 Rudolf Beran, University of California Berkeley.
Superefficient Estimation of Directional Trend

16:00 Peter Jupp, University of St Andrews.
Geometry of Score Tests and Wald Tests

16:30 Frits Ruymgaart, Texas Tech University.
A New Method to Solve Noisy Abel Equations

15:30–17:00, Montpetit Hall 204

Case Studies

Session 31: Case Studies: Mixtures Plus
Organizer and chair: Peter MACDONALD

15:30 Peter MACDONALD & Karen WHALEN, McMaster University & Department of Fisheries and Oceans.
Introduction

15:45 Jerome ASSELIN, Abu LATIF, Rachel MACKAY & Mark ROBINSON, University of British Columbia.
Floundering in Mixtures

15:57 Abdulkadir HUSSEIN, A. DE LEON, R. HUANG, X. CHEN & Ella HUSZTI, University of Alberta.
Analysis of Yellowtail Length and Age Mixture Data using Hidden Markov Chains

16:09 Martin PERRY & Dan KEHLER, Dalhousie University.
Title TBA

16:21 Rinku SUTRADHAR, Ellen JIANG & Yun-Hee CHOI, University of Waterloo.
Title TBA

16:33 William C. LIU & David R. BELLHOUSE, University of Western Ontario.
Double Sampling for Stratification and Estimation of Conditional Means and Marginal Distributions

16:45 Peter MACDONALD & Karen WHALEN, McMaster University & Department of Fisheries and Oceans.
Discussion

15:30–17:00, Montpetit Hall 207

Survey Methods
Contributed Paper Session

Session 32: Theoretical Survey Methods
Chair: Wayne FULLER

15:30 Yong YOU, Statistics Canada.
A Nonlinear Hierarchical Modeling Approach for Census Undercoverage Estimation

15:45 Yong YOU & J.N.K. RAO*, Statistics Canada & Carleton University.
Small Area Estimation with Unmatched Sampling and Linking Models

16:00 D. HAZIZA & J.N.K. RAO, Carleton University.
Design-Based Inference for Domain Totals under Imputation for Missing Data

16:15 M. TAUSI* & J.N.K. RAO, Carleton University.
Estimating Function Resampling Variance Estimators under Stratified Multistage Sampling

16:30 Changbao Wu, University of Waterloo.
Using Empirical Likelihood Methods to Obtain Positive Weights in Regression Estimation for Complex Surveys

Wednesday, June 7th

8:30–10:00, Montpetit Hall 207 Survey Methods
Contributed Paper Session

Session 33: Applied Survey Methods - I
Chair: Changbao Wu

8:30 Anjela Tzontcheva, University of Toronto.
Modified Regression Trees for Complex Survey Data

8:45 A. Demnati, J. Lawson, B. Harbour, J. Young & G. Robbins, Statistics Canada.
Estimation in a Traffic Network Survey

9:00 Sanping Chen, Paul Hunsberger & Claude Poirier, Statistics Canada.
Real-Time Editing: Some Technical Issues

9:15 Ante Rozga, University of Split.
The Effect of Political Expectations on Forecasting

8:30–10:00, Montpetit Hall 203 SurvMethBiostatistics
Invited Paper Session

Session 34: Statistical Methods for Correlated Data
Organizer and chair: Georgia Roberts

8:30 D. Roland Thomas & Yves J. Decady, Carleton University.
Analysing Categorical Data with Multiple Responses Per Subject

9:00 Lisa M. LaVange, Quintiles.
Applying Sample Survey Methods to Clinical Trials Data

9:30 Richard T. Burnett, Michael Jerrett & Daniel Krewski, Health Canada, McMaster University & University of Ottawa.
On the Use of Spatial Analysis Methods in Assessing the Relationship Between Community Air Pollution and Mortality

8:30–10:00, Montpetit Hall 201
Invited Paper Session

Session 35: Nonparametric Ranking Methods
Organizer and chair: Mayer Alvo

8:30 Philip L.H. Yu, Hong Kong University.
A Maximum Entropy Approach to Recovering Information from Ranking Data

9:15 M. Mountassir, University of Ottawa.
A Weighted Combination of Nonparametric Statistics for the Two-Sample Bivariate Problem

8:30–10:00, Morisset Hall 218

Special Session

Session 36: MCMC session with Peter Green
Organizer and chair: Duncan Murdoch

8:30 Peter Green, University of Bristol.
Bayesian Analysis of Heterogeneity using Mixtures and Related Models

8:30–10:00, Montpetit Hall 202

Biostatistics
Contributed Paper Session

Session 37: Risk Estimation and Applications
Chair: Rhonda Rosychuk

8:30 Lin Liu*, Chu-In Charles Lee & Jianan Peng, Memorial University of Newfoundland.
Cancelled

8:45 P.D. Faris, R.F. Brant & W. Ghali, University of Calgary.
Risk Adjustment with Binary Outcomes when Covariate Information is Incomplete

9:00 Jason Loeppky, Simon Fraser University.
Migration of Elephant Seals

9:15 Giuseppina D'Angelo, Diane Potvin, Jacques Turgeon, Phoenix International Life Sciences.
Carryover Effects in Bioequivalence (BE) Studies

9:30 Yingwei Peng, Memorial University of Newfoundland.
Estimating Baseline Distribution in Proportional Hazards Cure Models

9:45 Jeff Bakal & Dongsheng Tu, Queen's University.
Nonparametric Regression Analyses to Identify Prognostic Factors for Women with Early Breast Cancer

8:30–10:00, Morisset Hall 205

colloque "Méthodes et applications de la statistique 2000"
Invited Paper Session

Session 38: Methods for Teaching Statistics
Chair: Christian Genest

8:30 Marc Duchesne, Via Systems.
Methods for Teaching Statistics

10:30–12:00, Montpetit Hall 202

Survey Methods
Invited Paper Session

Session 39: Complex designs
Organizer and chair: Mike Hidiroglou

10:30 J.N.K. Rao, Carleton University.
Undoing Complex Survey Data Structures: Some Theory of Inverse Sampling

11:00 John L. Eltinge, U.S. Bureau of Labor Statistics.
Composite Estimation for Complex Survey Data

11:30 Ralph Folsom & A.C. Singh, Research Triangle Institute.
The Generalised Exponential Model for Sampling Weight Calibration

10:30–12:00, Montpetit Hall 207

Contributed Paper Session

Session 40: Graduate Student Presentations
Chair: Peter Kim

10:30 N'deye Rokhaya Gueye, Université de Montréal.
Equivariant Estimation of Multivariate Parameters from a Mixture of Normal Distributions with Certain Constraints

10:45 Steven X. Wang, University of British Columbia.
Combining Information by using Relevance Weighted Likelihood Method

11:00 Connie Winchester, Dalhousie University.
On Estimation of the Four-Parameter Kappa Distribution

11:15 Jin Zhang, York University.
Powerful Goodness-of-Fit Tests

10:30–12:00, Morisset Hall 218

Invited Paper Session

Session 41: Probability and Markov Chains
Organizer and chair: Jeffrey Rosenthal

10:30 Peter Bickel, UC Berkeley.
Bounds on Loglikelihood Derivatives in Hidden Markov Models and Various Applications

11:00 Priscilla Greenwood, University of British Columbia.
Markov Chains and $1/f$ Noise

11:30 Richard Tweedie, University of Minnesota.
Drift Conditions and Invariant Measures for Markov Chains without Irreducibility or Continuity Conditions

10:30–12:00, Montpetit Hall 203

Biostatistics
Invited Paper Session

Session 42: Large Multi-Centre Clinical Trials
Organizer and chair: Jayanti Mukherjee

10:30 Wayne Taylor, Clinical DataFax Systems.
Data Management Challenges for Large Multicenter Trials

11:00 Robin Roberts, McMaster University.
The CAPRIE Trial - A Case Study of an Academic/Pharmaceutical Industry Collaboration

11:30 David Henry, Bristol-Myers Squibb.
Organization of Large Multi-centre Trials from the Perspective of a Pharmaceutical Company Statistician

10:30–12:00, Montpetit Hall 201

Contributed Paper Session

Session 43: The Bootstrap, Wavelets, Nonlinear Regression, and Dynamical Systems
Chair: Brajendra Sutradhar

10:30 Mahmoud Zarepour, University of Ottawa.
On Development of the Weighted Bootstrap

10:45 Patrick J. Farrell* & T.W.F. Stroud, Acadia University & Queen's University.
Balanced Bootstrap Samples: A New Method for Unbalanced Data

11:00 Angelo J. Canty, Concordia University.
Examining Pivotality using Bootstrap Recycling

11:15 Alwell J. Oyet & Li Mei Sun, Memorial University of Newfoundland.
Wavelet Methods and Designs for Curve Estimation

11:30 J. Huh & K. C. Carrière, University of Alberta.
Efficiency of Aggregate Data in Non-linear Regression

11:45 Salim Lardjane, ENSAI.
Lyapunov Exponent Estimation for One-Dimensional Chaotic Dynamical Systems

10:30–12:00, Morisset Hall 205 colloque "Méthodes et applications de la statistique 2000"

Contributed Paper Session

Session 44: Estimation, biostatistics and time series
Chair: Jimmy Baulne

10:30 Christian Genest, Université Laval.
Is There Anything Left to Say about Pearson's Correlation Coefficient?

10:45 Serge B. Provost, The University of Western Ontario.
Density Estimation on the Half-line

11:00 Meriem Saïd, Université Laval.
Interval Censoring in Survival Analysis - Application to an HIV Study

11:15 M'hamed Mesfioui, Université Laval.
A Poisson Approximation for Heterogeneous Epidemic Processes

11:30 Elmiloud Bensaïd, Université Laval.
Weighted Local Adjustments of Polynomials to a Cumulative Distribution Function and its Partial Derivatives

11:45 Christian Genest, Jean-François Quessy & Bruno Rémillard, Université Laval & Université du Québec à Trois-Rivières.
Applications of a Serial Version of Kendall's Process

13:30–15:00, Montpetit Hall 202

Special Session

Session 45: Pierre-Robillard and CJS Awards
Organizer and chair: Mary Lesperance

13:30–15:00, Montpetit Hall 201

Survey Methods
Contributed Paper Session

Session 46: Applied Survey Methods — II
Chair: Julie Trépanier

13:30 M.-N. Parent & M. Simard, Statistics Canada.
Sampling for the Unified Enterprise Surveys (UES)

13:45 Caroline Cauchon, Alain Côté, Michelle Simard, Statistics Canada.
Estimation for the Unified Enterprise Surveys (UES).

14:00 Jean-Sébastien Provençal & Hélène Bérard, Statistics Canada.
The Annual Wholesale Trade Survey: the Transition to the Unified Enterprise Survey

14:15 Yves Béland, Suzanne Lessard & Marianna Morano, Statistics Canada.
Creation of the Dual Frame for the New Canadian Community Health Survey

13:30–15:00, Montpetit Hall 207

Biostatistics
Contributed Paper Session

Session 47: Correlated Data
Chair: Noel Cadigan

13:30 Wenqing He, University of Waterloo.
Analysis of Multivariate Failure Time Data

13:45 Yinshan Zhao, University of British Columbia.
An Analysis of Familial Correlation in Expression of Neurofibromatosis

14:00 J.C. Loredo-Osti & B. R. Smith, McGill University, Montreal General Hospital Research Institute, & Dalhousie University.
Likelihood in QTL Analysis under a Mixed Model of Inheritance in Non-Experimental Populations

14:15 J. Sunil Rao & Jiming Jiang, Case Western Reserve University.
On Mixed Model Selection

14:30 Renjun Ma, University of Ottawa.
An Optimal Linear Estimating Function For Latent Variable Models

14:45 Mei-Ling Ting Lee & G. A. Whitmore, Harvard Medical School and McGill University.
Distribution-Free Inferences about Latent Disease Progression and Survival

13:30–15:00, Montpetit Hall 203

Invited Paper Session

Session 48: Nonparametric Bayesian Inference
Organizer and chair: Mahmoud Zarepour

13:30 Hemant Ishwaran, The Cleveland Clinic Foundation.
Approximate Dirichlet Process Computing for Finite Normal Mixtures

14:00 George Tomlinson & Mike Escobar*, University of Toronto and Harvard University.
Analysis of Densities

14:30 Jayaram Sethuraman, Florida State University.
Properties and Approximations of Dirichlet Processes

13:30–15:00, Morisset Hall 218

Contributed Paper Session

Session 49: Probability, Design and Mathematical Statistics
Chair: Serge Provost

13:30 A.W. Kemp, University of St Andrews.
The Secretary Problem with Correlated Rank and Order

13:45 John F. Brewster* & Robert G. McLeod, University of Manitoba.
Design Issues in a Blocked Fractional Factorial Split-Split-Plot Experiment – A Case Study from the Aerospace Sector

14:00 Denis Larocque, Université du Québec à Trois-Rivières.
An Affine-Invariant Multivariate Sign Test for Cluster Correlated Data

14:15 Hongtu Zhu, University of Victoria.
Assessment of Local Influence for Models with Incomplete Data

14:30 D.J. Dupuis, Dalhousie University.
Fitting Bivariate Extreme Value Models Robustly

14:45 Young-Ho Cheong, University of Western Ontario.
Approximating the Distribution of Ratios of Quadratic Forms

13:30–15:00, Morisset Hall 205 colloque "Méthodes et applications de la statistique 2000"
Invited Paper Session

Session 50: Survey Methodology
Chair: Diane Leroux

13:30 Mario Pimparé & Johanne Thiffault, Ministère des transports du Québec.
Exploring the Path of an Origin-Destination Survey

Abstracts

Biostatistics Workshop

Sunday, June 4th, 9:00–16:00, Montpetit Hall 202

The Analysis of Event History Data

Richard Cook & Jerry Lawless, University of Waterloo

The Biostatistics Workshop on the Analysis of Event History Data will be led by Professors Richard Cook and Jerry Lawless from the University of Waterloo. Event history analysis deals with events which occur over time to individuals, and has widespread application in biostatistics, epidemiology, and other areas such as demography, economics, and sociology. Examples of events for humans include onset of disease, clinical outcomes related to disease or its treatment, completion of an education program, gaining or losing employment, utilization of health or social services, marriage, divorce, birth and death.

Survival analysis is the most familiar type of event history analysis, and deals with the time to occurrence of some specific event. More generally, we may want to consider the occurrence of several types of event, relationships among the events, and factors which influence the occurrence and timing of the events. This workshop will provide an introduction to important types of event history analysis, beginning with survival analysis and progressing to the analysis of multiple events and multi-state models.

The emphasis will be on methods which can be implemented using existing software. S-Plus will be used in the examples, but other packages (e.g. SAS) will be considered, and both the background and implementation of methodology will be discussed. Applications and examples will be drawn from bronchial disease, cancer treatment, HIV/AIDS epidemiology, neurosurgery, and organ transplantation. The workshop will be aimed at making these techniques accessible to graduate students, biostatisticians out and in of academia, and established researchers. The workshop will also devote a period of time to the review of current or newly developed methodology for the established researcher.

The workshop is sponsored by Glaxo Wellcome Inc. and the Biostatistics Section.

Survey Methods Workshop

Sunday, June 4th, 10:00–16:00, Montpetit Hall 203

Sampling: Design and Analysis

Sharon Lohr, Ohio State University

The morning session of the workshop will be an introduction to sampling topics, based on Sharon Lohr's recently published text "Sampling: Design and Analysis" (Duxbury Press). The afternoon session will cover more advanced topics on variance estimation. The text brings the field of survey sampling into the real world, with true-life examples drawn from a variety of disciplines and surveys. With an emphasis on evaluating real surveys and solving real problems, it is equal parts instructional tool and reference tool and reference guide—as useful for the reader new to the subject as it is for the survey practitioner. Participants are encouraged to bring their own copies of the text to the workshop, or to purchase copies there. The text is not included in the registration fee.

CJS/CRM Read Paper Session

Sunday, June 4th, 16:00–19:00, Morisset Hall 218

The Estimating Function Bootstrap

Feifang Hu, University of Singapore, & John D. Kalbfleisch, University of Waterloo

The authors propose a bootstrap procedure which estimates the distribution of an estimating function by resampling its terms using bootstrap techniques. Studentized versions of this so-called estimating function (EF) bootstrap yield methods which are invariant under reparametrizations. This approach often has substantial advantage, both in computation and accuracy, over more traditional bootstrap methods and it applies to a wide class of practical problems where the data are independent but not necessarily identically distributed. The methods allow for simultaneous estimation of vector parameters and their components. The authors use simulations to compare the EF bootstrap with competing methods in several examples including the common means problem and nonlinear regression. They also prove asymptotic results showing that the studentized EF bootstrap yields higher order approximations for the whole vector parameter in a wide class of problems.

Session 1: Welcome and SSC Presidential Invited Address

Monday, June 5th, 8:30, Tabaret Hall Chapel

Air Pollution and Health Risk: Some Statistical Problems and Solutions

Jim Zidek, University of British Columbia

In this talk I will present some basic issues that I and my co-investigators have confronted over a period of years in a program of research centered at UBC. A number of these issues relate to the fundamental problem of proving a causal link between air pollution and morbidity as well as morbidity. The existing link has been established primarily through the demonstration of a consistent statistical association over time and space. Although medical researchers are clinically testing new theories of causation, the dominance of statistical evidence is likely to persist for some time to come. And it is that evidence that underlies the controversial criteria for particulate air pollution that are now being hotly debated in the United States. In this talk I will discuss a hierarchical Bayesian approach that incorporates computer simulation models (pCNEM/PM), to better estimate human exposure to air pollution. I will highlight the difficulties presented by measurement error and the role of spatial prediction of pollution fields in addressing that problem. The predictive distribution can be coupled with the pCNEM/PM to form a "hybrid" predictor of human exposure that takes account of human behaviour and its influence on exposure. [On high ozone days people may tend to stay in air-conditioned indoor environments and paradoxically experience lower levels of exposure than the ambient levels (on which criteria are set) would predict.] I will indicate how a health effects impact model developed at UBC can be coupled with the output from the hybrid model. Inferences made about the association between air pollution and health then reflect uncertainties about the actual exposures. And reflecting those uncertainties is essential to avoid conclusions that will be discounted by regulators and those impacted by their regulations.

Session 2: Poster Session

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

An Asymptotic Approximation for the Birthday Problem

S. Ejaz Ahmed & Richard J. Mcintosh, University of Regina

It is known that for a class of 23 students the probability that at least two students have the same birthday is more than 0.5. Suppose that the number of days in the calender tends to infinity. For a fixed number p with $0 < p < 1$ we give an asymptotic formula and a simple proof, not using Stirling's formula, for the minimum class size to insure a probability of at least p that two or more students have the same birthday.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

On the Convergence Rates of the Bootstrapped Means

S. Ejaz Ahmed & A.I. Volodin, University of Regina

For a sequence of random variables $\{X_n, n \geq 1\}$, the convergence rate (that is Baum Katz/Hsu-Robbins/Spitzer type result) is obtained for bootstrapped means. In this investigation, the sequence of random variables need not be necessary independent or identically distributed. Further, no assumptions are made concerning either the marginal or joint distributions of the random variables.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Self-Modeling Regression with Random Effects using Penalized Regression Splines

Naomi Altman, Cornell University, & Julio Villareal, Cornell University & University of California, San Diego

Self-modeling regression is a semi-parametric method for describing a family of similar curves. The overall shape of the curve is estimated nonparametrically, but differences among the curves are described through a parametric model. In designed experiments in which the response is a curve, modeling the parameters through a random effects model is often desirable. In this paper, we describe a self-modeling regression model in which the nonparametric curve is defined by a penalized regression spline. Since the spline can be estimated as a linear random effects model, this allows us considerable simplification in both computation and inference. This simplicity can then be exploited to extend the model to generalized regressions.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Estimating the Population Size of Clams Using A Model Assisted Sampling Design and Analysis

Denise Babineau, Simon Fraser University

For several decades, North Beach near Masset, British Columbia has been used for commercial and non-commercial harvesting of clams. In the early 1990's, there was some concern among environmentalists that the health of the stock seemed to be failing. To address this concern, the size of the population needed to be accurately estimated. Usually, standard sampling methods, like those outlined in Gillespie and Kronlund (1999), are used to estimate the population size of clams in a certain area. However, because methods like simple random sampling, stratified random sampling, two-stage sampling and stratified two stage sampling can be applied to any population, such methods fail to incorporate any biological knowledge that may be discovered concerning the distribution of clams. To use such information to an advantage, a model assisted sampling design and analysis must be carried out.

This type of analysis was used to estimate the population size of clams on North Beach and surrounding beaches using a three stage sampling design. Once the beaches of interest were chosen, transects were randomly sampled along each beach. Distances were then systematically sampled along each transect. To sample the clams, plots of sand were randomly sampled at each distance. Two methods of estimation were then used to estimate the number of clams on each transect considered. The first method used a straight line interpolation while the second method used a cubic smoothing spline. A preliminary analysis of the data also showed that the population density for each distance considered had an approximate Poisson distribution. This information was also integrated into the two separate analyses as well. To estimate the overall population size of clams for each beach considered, both a ratio estimator and a simple inflation estimator were used. These results were then used to obtain a sampling design that gave the best estimate of population size using optimum numbers of transects, beach elevations, and samples of clams.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Linear Regression with Spatially Correlated Data

Cynthia Bocci, Université d'Ottawa

In this paper, we pursue a regression model for data whose dependence is captured through spatially correlated errors. Parameters include both the regression and spatial coefficients. The estimation of the spatial parameters by some common techniques appears ineffective for this model. An algorithm is proposed to obtain parameter estimates and compared to other methods of estimation through simulation. Furthermore, a procedure for the optimal placement of additional sites is suggested by minimizing the asymptotic covariance matrix of the parameter estimates.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Space Debris: Flux in a Two Dimensional Orbit

David R. Brillinger, University of California, Berkeley

The debris circling the Earth has become an important environmental problem. In particular it creates nonnegligeable risk for spacecraft and satellites. This paper discusses a method to bound the rate of passage of objects through an arc segment in the plane of motion, given random initial conditions for the orbiting object. An advantage of the approach of this paper is that the initial conditions may be arbitrary and that moments beyond the first may be computed in a similar manner.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Evaluating Predictions of Faulty Software Using Thresholds

Khaled El Emam, National Research Council of Canada

An increasingly common practice in high maturity software organizations is to collect metrics that characterize the structure of software, and use these metrics to predict which components are most likely to contain a fault. For example, one can measure coupling: the number of connections amongst the components. Components with high coupling are more difficult to understand, and hence more likely to have a fault inserted during development. Threshold models can be developed for this purpose, whereby coupling up to a certain threshold value has no impact on fault-proneness, beyond which the probability of a fault increases rapidly. The justification for thresholds derives from a cognitive theory. This study evaluate thresholds for some popular metrics on two object-oriented telecommunications systems. We apply a technique proposed in an epidemiological context for identifying and testing threshold effects in dose-response relationships. Our results indicate that there are no threshold effects for object-oriented metrics, hence disconfirming the posited theory.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Statistical Linkage Analysis Led to Strong Evidence of a Susceptibility Gene on Chromosome 11 for Gilles de la Tourette Syndrome

Claudia Émond, Chantal Mérette, Andrée Brassard & Chantal Caron, Université Laval

Gilles de la Tourette syndrome is a chronic tic disorder characterized by sudden and recurrent occurrence of multiple motor and vocal tics with an average age of onset of 7 years old. Family, twin and adoption studies have demonstrated strong evidence for a genetic basis of this disorder. A sample made of one large multigenerational pedigree densely affected by Tourette's syndrome was ascertained in La Malbaie an area of Eastern Quebec. Clinical and genetic information were collected at the Centre de recherche Université Laval Robert-Giffard for 120 family members. Statistical genetic linkage analysis and association studies were performed with the aim of detecting one or more genes involved in the aetiology of this childhood disorder.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Dose-Time-Response Nonlinear Model of tPA-Mediated Plasma Clot Lysis in the Presence of Candidate Drugs

Krish Ghosh, M. Prendergast, C. Soong & G. Matsueda, Bristol-Myers Squibb

Analysis of in-vitro human plasma clot lysis is used to evaluate drug candidates as inhibitors of fibrin clot formation or dissolution. A study was developed to evaluate candidate drugs on thrombin-mediated clot formation and alteplase (tPA)-mediated clot lysis. The experiment was conducted at different concentrations of tPA in the presence of DMSO and increasing concentrations of inhibitor at different time points. The interaction of tPA and inhibitor led to inhibition of the overall response. A four-dimensional nonlinear statistical model was developed to estimate parameters of inhibitor compounds. SAS/JMP and SigmaPlot were used for such evaluations.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

BUGS Program for the Analysis of Multi-Level Repeated Ordinal Data using a Bayesian Hierarchical Model

Zhenguo Qiu, York University

For the multi-level data with an unequal number of ordinal outcomes: Oral Practice Examinations (OPEs), Ming Tan et al. (Statist. Med. 18, 1983-1992 (1999)) proposed a Bayesian hierarchical proportional odds model to make inference on the association of OPE outcomes with other factors, such as the length of training, didactic experience, and to estimate the variance components as well. The implementation of this model is complicated and computational intensively using common mathematical software. In this paper we suggest a reparameterized model using BUGS program to implement the above data analysis and also discuss the corresponding MCMC convergence.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Measures for Comparing Head and Neck Cancer Staging Schemes

K. Schulze, P.A. Groome, B. O'Sullivan, B. Grice, C. Goh, S.F. Hall, F. Liu W.J. MacKillop, D. Payne, D. Rothwell, J. Waldron, P. Warde & B. Cummings, Queen's University & University of Toronto

Malignant tumours can be classified by the tumour extent (T) & lymph node involvement (N). To make the 20 possible combinations useful they must be grouped. Our goal was to quantitatively compare 8 different stage grouping schemes to check that 1) the T–N subgroups are consistent within their group, 2) the groups are diverse, 3) the percent variance explained (PVE) is high, & 4) the groups are balanced in number. For (1) we developed a measure that compared the distance between survival of the individual T–N curves and their group curve. For (2) we calculated a score measuring how evenly the group curves were spaced and the span of their survival. Methods of Mittleböck & Schemper (1996) were used for PVE (3). We used a sample expected–versus–observed measure for balance (4). Measures are demonstrated using tonsil cancer data.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

The Siren Call of the Schaeffer Estimator

Carl James Schwarz, Simon Fraser University

The Schaeffer estimator is often used in fisheries management to estimate abundance. Typically, fish are marked over a period of time, and recaptured over a period of time (e.g. weekly). The observed distribution of the marked recoveries is used to expand the recoveries of the unmarked fish in a seemingly intuitive fashion. However, all of the variations presented in the literature have a common flaw - they assume equal catchability either in the marking or recovery phases. Under these conditions, the simple Petersen estimator should be used since it is more precise and also consistent. If this assumption is false, then the Schaeffer estimator can be severely biased and the stratified-Petersen should be used in its place.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Comparison of Ratio-Type Estimators under a Super Population Model

Javid Shabbir, Quaid-I-Azam University, Pakistan

A ratio-type estimator is modified. Its comparison is made with ratio estimator and Chakrabarty's estimator (1979) with respect to bias and variance under a model. The modified estimator is found to be more attractive in case of its efficiency and less bias than the mean and ratio estimators and is comparable with Chakrabarty's estimator.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

A Method for Forming Nonresponse Adjustment Cells, Applied to Children without Vaccination Histories from Providers in the National Immunization Survey

Philip J. Smith, Centers for Disease Control and Prevention, Atlanta

For forming cells to adjust for unit nonresponse, it is often effective to group sample units according to either their response propensities or their predictive probabilities. To combine these two approaches, we use response propensity as the basis for an initial set of cells and then use predictive probability to refine each cell (separately). A straightforward algorithm determines the total number of adjustment cells and the extent of refinement, including the possibility of basing the cells entirely on response propensity or entirely on predictive probability. We apply the method to the National Immunization Survey (NIS), which uses an RDD survey to elicit households' reports on the vaccinations of sample children, but then may be unable to obtain vaccination histories from some children's providers. For the 1998 samples in the 78 Immunization Action Plan areas of the NIS, two or three adjustment cells, based on predictive probability, are generally best, in the sense of bias reduction. The results suggest that the NIS encounters little nonresponse bias from the provider-verified vaccination histories that it cannot obtain.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

SSC Ad-Hoc Committee on Professional Development

ACPD/CSPP, Statistical Society of Canada

The Committee on Professional Development was established as an ad hoc committee by the Executive Committee on 19 October 1997. Its mandate is to consider ways of bringing non-academic statisticians into the Society, of developing continuing education courses and workshops for professional statisticians, and generally consider ways in which the Society can better serve the non-academic statisticians. The recent Strategic Plan for the SSC has put particular emphasis on professional development activities suggesting the the Society must try to satisfy the needs of practicing statisticians including balancing between the needs of academic and practicing statisticians in the publications that the Society offers. The committee aims to inform SSC members of its current activities and plans for the future.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Combining Biased and Unbiased Estimates in a Meta-Analysis

George Tomlinson, Toronto General Hospital & University of Toronto, & Keith O'Rourke, Ottawa Hospital & University of Ottawa

After completion of the data collection stage of a meta-analysis, the relevant studies can sometimes be classified as randomized (and on average giving unbiased estimates) or nonrandomized (and on average giving perhaps biased estimates). Choice of which studies to include in the meta-analysis is now a very important issue. One approach is to examine only the unbiased studies. A Bayesian, knowing that the nonrandomized studies contain bias, might place a carefully elicited prior on the bias and include all studies in the meta-analysis, hoping to arrive at a nearly unbiased estimate with smaller variance. We show that this second approach is preferable only under conditions where more is known about the bias than is realistic, and recommend the combination of only unbiased estimates.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Predicting Thickness of Pipe Walls Subjected to Flow Accelerated Corrosion

Rolf Turner, University of New Brunswick

In a recent consulting project I was asked to compute prediction intervals for the thicknesses (which appear to decay linearly) of the walls of pipes carrying heavy water in a nuclear power plant. The aim was to forecast when the thickness of a pipe wall would fall below a specified level that provides for a reasonable safety margin. Although this problem is basically elementary, there are several intriguing non-standard aspects to it. These include the random effect of the initial thickness of the walls of different pipes, and the fact that prediction intervals were required for individual pipes rather than for the ensemble. There is also a problem of how to incorporate a separate estimate of measurement error. In this presentation, I will elaborate on these problems and how they were solved.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Statistical Methods in Symmetry Studies

Marlos Viana, University of Illinois, Chicago

Symmetry studies are aimed at the statistical/probabilistic assessment of experimental hypotheses related to symmetry. This is in contrast to the well-established "comparative studies", in which the experimental evaluation of a scientific question is directly translated (often incompletely or inadequately) into the statistical evaluation of a parametric hypothesis comparing two or more means (e.g., analysis of variance). In this seminar I will discuss the notion of representing and quantifying symmetry and argue that the statistical assessment of typical symmetry-related questions (in clinical studies, for example) is either fully absent or mistakenly viewed as a question of statistical association or correlation. Selected examples of anatomical and functional symmetries, temporal symmetries (e.g., post caloric nystagmus) and circular symmetries (e.g., corneal topography images) will be discussed within that argument. I will then comment that a class of natural tools to assess symmetry is based on the notion of invariance under a well-defined set of actions, such as those defined by permutations. The amount of symmetry is characterized by (indexes related to) the amount of permutations leaving the system invariant. This leads to the notion of random permutations and its potential as a method of exploring the data from experimental symmetry studies. I will discuss preliminary applications of analytical properties (vs. simulation) of random permutations to the analysis of data from familial/bilateral indexes and to the analysis of multivariate data with circular coordinates, such as those obtained from corneal topography maps.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Diagnostic Checking for Linear and Nonlinear Time Series Data Fittings: An Application of Reordered Residuals Processes

Zilin Wang, University of Western Ontario

Using a reordered residuals process (RRP), we developed a model selection criterion to diagnose the specifications of ARMA and ARCH processes. Based on the theoretical findings in Kawczak (1998) that the RRP from an independently and identically distributed innovation of an ARMA model converges to a centered Gaussian process, one would expect distance between zero and the RRP of a correctly fitted model should be less than that of a wrongly specified model because misspecification yields local dependence. By comparing the maximum absolute values of the RRPs from different fittings, one can locate the correct order to be the corresponding order of the first minimum number among all the maximum distances (FMIN). Because the FMIN criterion is designed to pick up the local dependence of wrongly fitted models around the true model, it can significantly avoid the overparametrization. The results from most of the Monte Carlo experiments show that FMIN outperforms or performs at least as well as AIC for both ARMA and ARCH models. In addition, the performance of the FMIN in the ARCH process is affected by functional form of the random weights and the lags for the observation to be included in the partial sum of the process.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Annual Trends in Rates of Asthma Hospital Admissions in Nova Scotia, Canada: Influence of Hospital Bed Reductions

Swarna Weerasinghe & Lamont Sweet, Dalhousie University

The Maritime provinces have the highest reported rates of asthma morbidity of all the Canadian provinces. Specifically the rates reported for Nova Scotia is twice more than the Canadian average. In the last decade, access to hospitals in Nova Scotia declined annually as a result of health care budget limitations. Assessment of effects of asthma using hospital admissions data as a marker can be a challenging task since access to hospitals varied annually, especially due reductions of hospital beds available in NS. This paper presents a method to estimate annual trends in asthma admission that account for the changes in hospital beds using total beds available as a surrogate measure for asthma beds. Two phase, population average, Poisson regression analysis were performed to estimate the annual variation of asthma hospital admissions accounted for the annual variation of hospital beds. The analytical methods presented in this paper are not limited to the investigation of annual trends in asthma but can be applied to any other disease, where hospital admissions can be used as a marker for exacerbation or as an outcome measure of incidences of temporal trends. Strong association between hospital bed reductions and asthma admissions were found. Hospital admission rates when adjusted for available beds show a remarkable increasing trend. However, in all health regions slight downward trend over the years can be seen.

Monday, June 5th, 10:30–12:00, Unicentre Jock-Turcot Promenade

Integer-Valued, Minimax Robust Designs for Approximately Linear Model with Correlated Errors

Julie Zhou, University of Victoria

We consider finite design spaces and integer-valued discrete designs to construct minimax designs which are robust against both departures from assumed linear regression response and departures from the assumption of uncorrelated errors. Simulated annealing algorithm is developed to carry out the minimization problems to search for minimax designs. In particular minimax designs are studied for approximately linear models with errors following moving average MA processes. Results are obtained for $MA(1)$ and $MA(2)$ error processes in this paper, and those results can be extended to $MA(q), (q \geq 3)$ processes. Examples of minimax designs are given for approximately polynomial regression and approximately second order multiple regression.

Session 3: Confidentiality

Monday, June 5th, 13:30, Montpetit Hall 202

National Center for Health Statistics Approaches to the Release of Microdata: Data Perturbation and the Research Data Center

John Horm, National Center for Health Statistics

The National Center for Health Statistics (NCHS) in the United States is the agency with the primary responsibility for collecting and disseminating health data. The NCHS operates under strict confidentiality regulations that do not permit releasing identifiable microdata files. Any microdata files released outside the agency are considered to be in the public domain because once files are released, the NCHS is unable to further protect them. Recently the NCHS has developed a Research Data Center, which allows access to additional internal microdata files through two mechanisms: remote access and onsite. In both cases, strict controls and procedures have been developed and implemented to ensure that respondent confidentiality is not compromised. During the first year of operation, the NCHS RDC hosted over 20 researchers who successfully prepared work for presentations at professional meetings, peer-review journal articles, and completed two doctoral dissertations.

Monday, June 5th, 14:00, Montpetit Hall 202

Access to Microdata Files at Statistics Canada

Jeanine Bustros, Statistics Canada

The fast paced environment, the new technology and the continuing need for information has increased the pressure on the statistical systems to develop new means by which the data collected is made accessible to the research community. Under these new rules, Statistics Canada had to complement the access to their traditional Public Use Microdata Files and aggregate tables by allowing access to unscreened microdata files. These access methods have been developed taking into consideration the Statistics Canada Act, which prevents the disclosure of any information that could possibly be related to any individual person. This presentation will focus on the framework and the implementation of two new ways to access unscreened microdata files within Statistics Canada, namely the Research Data Centres and the Remote Access.

Session 4: Longitudinal and Life-History Data Analysis

Monday, June 5th, 13:30, Montpetit Hall 201

Marginal Dispersion Models for Longitudinal Data

Peter X.-K. Song, York University

In this talk I will present a class of marginal models based on Jorgensen's dispersion models which provides a synthesis and an extension of Liang and Zeger's marginal models. The extension enables us to analyse other types of longitudinal data that Liang and Zeger's models do not cover, such as longitudinal directional data and longitudinal compositional data. For the part of parameter estimation, the dispersion models lead to an generalization of Liang and Zeger's GEE due to the nonlinear score functions in marginal expectations. Also I will illustrate the theory with an analysis of longitudinal continuous proportional data from an ophthalmology study on the use of intraocular gas in retinal surgeries.

Monday, June 5th, 13:45, Montpetit Hall 201

A Two-state Hidden Markov Poisson Model

Peiming Wang, Nanyang Technological University, Singapore & Martin L. Puterman, University of British Columbia

This talk discusses a two-state hidden Markov Poisson model for analyzing longitudinal data of epileptic seizure counts, which allows for the rate of the Poisson process to depend on covariates through an exponential link function and to change according to the states of a two-state Markov chain with its transition probabilities associated with covariates through a logit link function. The proposed model in this talk relaxes the stationary requirement of the Markov chain, allows for overdispersion relative to the usual Poisson regression model and for correlation between repeated observations. It provides a plausible analysis for the longitudinal data of epileptic seizure counts. Maximum likelihood estimation using the EM and quasi-Newton algorithms is discussed. Monte Carlo studies investigate the reliability of the estimation method, the choice of starting probabilities for the Markov chain, and some finite-sample behaviors of the maximum likelihood estimates, which suggests that (1) the estimation method is accurate and reliable for both single-subject and multiple-subject longitudinal count data as long as the total number of observations is reasonably large, and (2) the choice of starting probabilities for the Markov process has little impact on the parameter estimates.

Monday, June 5th, 14:00, Montpetit Hall 201

Test of Treatment Effects Robust to Dependent Censoring

Bingshu Chen & Richard J. Cook, University of Waterloo

In the analysis of event history data, it is typically assumed that censoring is independent of the event process. It is widely acknowledge when this assumption is not true that a naive analysis will generate biased

and/or inefficient parameter estimates. Here we develop a robust method to test for treatment effects in the presence of dependent censoring. The survival time of subject is controlled in a multiplicative intensity model. For censored subjects, the real terminating times are unknown and considered as missing values. The EM algorithm will be applied to address this missing data. Simulation studies will be discussed which examine the performance of the association tests.

Monday, June 5th, 14:15, Montpetit Hall 201

A Semi-Markov Model for Binary Longitudinal Responses Subject to Misclassification

Rhonda J. Rosychuk, University of Alberta Mary E. Thompson, University of Waterloo

We consider a two-state continuous-time semi-Markov model for an unobservable alternating binary process. An observable process, thought to be closely related to the unobservable true process, is observed at discrete time points and may misclassify the state of the true process. To estimate parameters associated with the sojourn time distributions and probabilities of misclassification, a minimum Pearson chi-square type estimating function is proposed based on joint probabilities when the true process is in equilibrium. The equilibrium joint probabilities for three consecutive observations are calculated via a recursive approximation. We demonstrate the methodology on parasitic infection data where exponential and gamma sojourn time distributions are considered for the uninfected and infected states, respectively.

Monday, June 5th, 14:30, Montpetit Hall 201

Efficiency Consideration in the Design of Clinical Trials with Recurrent Events

W. Wei & R.J. Cook, University of Waterloo

Clinical trials in which treatment comparisons are made based on recurrent events are becoming increasingly common. Use of efficient procedures for estimating treatment effects is critical to ensure optimal use of resources, timely study completion, and rapid dissemination of results. We compare several methods of analysis for recurrent events which are commonly reported in the literature. Relative efficiency and power considerations are explored under a mixed Poisson regression model. We also consider issues surrounding the analysis of heterogeneous recurrent event data in the presence of selection criteria. These findings have implications on both the design and analysis of future studies.

Session 5: Statistics in Business and Industry

Monday, June 5th, 13:30, Morisset Hall 218

Tree Models in Industrial Statistics

Hugh Chipman, University of Waterloo

Tree models provide a flexible, adaptive means of relating several input variables to one or more response variables. In the context of industrial data, they offer an alternative to the usual linear model and analysis of variance. In particular, tree models are well suited to identify interactions between variables, and partition the space of input variables into regions where simple models describe the response. Recent tree methods will be outlined, and their application to data arising from designed experiments illustrated. Emphasis will be placed on search algorithms for identification of interesting trees, and methods for interpreting the trees identified by the search.

Monday, June 5th, 14:00, Morisset Hall 218

Design of Experiments Applied to the Development of an Alternative Fuel Injection System

Mike Brajac, General Motors of Canada

The value of experimental design has clearly been identified in industry. This paper explores the application of experimental design to the development of an Alternative Fuel Injection System. The investigation consists of two experiments: the first experiment was conducted on actual development vehicles and was not successful, the second experiment, conducted in a laboratory environment, was very successful in providing engineers with knowledge and insight. Reasons for the difference in the outcomes between the two experiments and the lessons learned will be discussed.

Session 6: Positive Dependence and Ordering for Point Processes

Monday, June 5th, 13:30, Montpetit Hall 203

Positively and Negatively Dependent Point Processes: Structures, Constructions, and Central Limit Theorems

Robert M. Burton, Oregon State University

The class of objects under consideration is that of point processes on the line or on general Euclidean space. There is a multitude of ways that positive [negative] dependence may be defined. Here the focus is on the strong notions of association and the FKG inequalities, and their variants. Point processes come with the natural order of inclusion. A point process is associated if increasing functions are positively correlated. The FKG inequalities are defined in terms of product densities and put higher probabilities on clusters of points. Examples that are under discussion include self-exciting [self-inhibiting] processes, cluster processes, Poisson point processes subordinated to a random measure, and some examples arising in the theory of neural networks. The central limit theorem and the invariance principle follow from relatively mild moment conditions. If time permits, a connection with a problem in financial mathematics concludes the presentation.

Monday, June 5th, 14:15, Montpetit Hall 203

Self-exciting Point Processes: Vector Representation and Kernel Monotonicity

Mathieu Plante, University of Ottawa

Simple point processes on the half-line admit several representations as sequential random vectors with real entries. In some cases, a closed partial order $\prec$ on the space of point process realizations may be induced from the coordinate partial order $\leq$ on $\mathbf{R}^\infty$ via one such representation. When this holds, a point process is associated ($\prec$) iff its corresponding $\prec$-specific representation is associated ($\leq$) as a random element of $\mathbf{R}^\infty$. It is also known of (sequential) random vectors that if their finite dimensional projections exhibit the property of monotone kernels, then they are associated ($\leq$) (Lindqvist, 1988). On the other hand, point processes which are defined as self-exciting with respect to $\prec$ have been shown, under fairly general hypotheses, to be associated ($\prec$) when $\prec$ is one of three natural partial orders (Kwieciński and Szekli, 1996). In this presentation we relate, for each of these three orders, the self-exciting property of a point process to the conjectured kernel monotonicity of its representation vector.

Session 7: Graduate Student Presentations

Monday, June 5th, 13:30, Montpetit Hall 207

The Detection of Local Volume Change in Shape Analysis

M.K. Chung & K.J. Worsley, McGill University

Continuous 3D vector displacement data from the non-linear deformation required to map one geometric object to another is available at each point in the object. Hotelling's T^2 random field has been used in detecting local shape changes.(Cao and Worsley, 2000) In this paper, we present a new method based on the dilatation random field or the trace of the displacement gradient matrix which measures the relative local volume change in the non-linear deformation. The results are applied to the detection of the region of brain growth or atrophy in magnetic resonance scans of 29 subjects over time.

Monday, June 5th, 13:45, Montpetit Hall 207

Robust Spectrum Estimation in an ARMA(p, q) Model with Application to the Sea Level Data

Sanjoy K. Sinha, Christopher A. Field & Bruce Smith, Dalhousie University

Instead of following a pure ARMA process, most oceanographic data follow a contaminated (or mixed) process which is an ARMA process mixed with some harmonic components. The classical ML method of spectrum estimation in an ARMA process may be severely affected by the harmonic components mixed with the original ARMA process. Here robust spectrum estimation in an ARMA(p, q) process has been considered in order to avoid the effect of such harmonic contamination. The estimation procedure divides naturally into two parts, in the first of which we discuss the robust estimation of the power spectrum of an ARMA(p, q) process assuming the orders p and q are known a priori. In the second part, a robust model selection criterion has been proposed for choosing the appropriate orders of the process. The proposed robust estimation and model selection technique have been applied to some actual data sets on sea level fluctuations at Halifax, Sydney and Yarmouth harbors of Nova Scotia, Canada. The analysis shows that robust M-estimate of the power spectrum enables us to see important features of sea level such as harbor seiches in the presence of large tidal components.

Monday, June 5th, 14:00, Montpetit Hall 207

Continuous-Time Stochastic Processes Based on Generalizations of Binomial Thinning Operators for Modelling Count Data

Rong Zhu, University of British Columbia

We study a class of continuous-time Markov processes which have non-negative integer margins. These processes have been developed via a representation from a generalization of the binomial thinning operator. The marginal distributions of the class of processes include the Poisson distribution, negative binomial distribution, power series distribution, and logarithmic series distribution, which are of special interest for modelling count data. Time series models which arise from the special case of equally spaced observations from these processes are compared with known results in the literatures. Furthermore, we investigate the dependence structures and the probabilistic properties of these processes for statistical inference. These models are motivated from the modelling for unequally spaced longitudinal count data.

Monday, June 5th, 14:15, Montpetit Hall 207

Laplace Transforms of Order Statistics of Erlang Random Variables

Wayne Horn, University of Western Ontario, Myron Hlynka & Percy H. Brill, University of Windsor

We present several methods to calculate the Laplace transform of order statistics of Erlang random variables. All methods are based on the probabilistic interpretation of the Laplace transform. A Markov chain analysis is included. Applications to queueing and other stochastic models are discussed.

Monday, June 5th, 14:30, Montpetit Hall 207

Cardiovascular Disease Risk Factors among Bermudans in Bermuda: Comparison with Canadians Living in Nova Scotia Utilizing Existing Data

Minnie Horace, Swarna Weerasinghe, David MacLean & Meng Tan, Dalhousie University

Diseases of the circulatory system are the most common causes of mortality and morbidity in Bermuda as well as in Nova Scotia, Canada. While the extend of cardiovascular disease and CVD risk factors are well known in Nova Scotia, there is scant information on the extend of the disease and its risk factors in Bermuda. Utilizing data from the Bermuda Diabetes Epidemiology Survey (1996), the first project to collect data on cardiovascular disease risk factors in Bermuda, this study is intended to analyze the prevalence of cardiovascular disease, disease risk factors and disease awareness levels among Bermudans. The impact of

various socioeconomic and demographic factors on cardiovascular disease rates and risk factor levels on the island nation will also be examined. These results will then be compared with Nova Scotia's, utilizing data from the Nova Scotia Health Survey (1995). Comparability of the data from the two surveys is facilitated partly by the similar survey instrument and methodologies used. Also, socioeconomic indicators for the two populations are fairly similar except for the percent of gross domestic product spent on health services and programs, which is higher in Nova Scotia.

Session 9: Special session on Survey Sampling

Monday, June 5th, 13:45, Morisset Hall 205

Inquiries Coordination for Businesses

Paul-André Salamin, Office fédéral de la statistique, Suisse

An inquiry coordination system is being created for the review of the Federal Statistical Office's Register of Businesses (Switzerland). The objective of this system is to divide the workload among the units inquired as fairly as possible. Moreover, the system must be able to deal with different types of units, i.e. standard units and special units. Standard units are businesses, while special units are sometimes constructed for special inquiries. Finally, the system must also allow to draw and update samplings for panels. Most of the inquiry coordination systems use fixed random numbers. According to our approach, the coordination is uniquely determined by the workload, which is defined as the number of times a unit was inquired. In this talk, we will present algorithms that allow for the drawing of sampling series coordinated by the workload. We show that, under certain conditions, the marginal distributions induced by the coordinated sampling are the same as those from simple random surveys or Bernoulli surveys. Those conditions require constraints in dealing with the different types of units and in considering the population changes.

Session 11: Invited Address by Norman Breslow

Monday, June 5th, 15:30, Montpetit Hall 201

The Value of Long Term Follow-up: Lessons from the National Wilms Tumor Study

Norman Breslow, University of Washington, Seattle

Wilms tumor is an embryonal tumor of the kidney that affects approximately one child in every 10,000. During the 20th century, cure rates increased from 10% to 90% as first radiation and then chemotherapy joined surgical removal of the diseased kidney as standard treatment. The National Wilms Tumor Study Group (NWTSG) was established in 1969 and currently enrolls nearly 80% of the 550 cases estimated to occur annually in North America. Its focus has been the identification of patient subgroups at high or low risk of relapse, and the substitution of combination chemotherapy for radiation therapy, with a primary goal to reduce long term complications while producing the maximum number of cures. The NWTSG Data and Statistical Center, located in Seattle since the start of the study, has played a major role in this effort. Systematic follow-up of the increasing number of surviving patients has documented the long term "costs of cure" and the wisdom of reserving the most toxic treatments for those who actually need them. Secondary cancers, for example, which once affected 1.6% of Wilms tumor survivors by 15 years from diagnosis, have been much reduced for more recent patients because 60% of them no longer receive radiation therapy. Other studies have documented the effects of abdominal radiation on skeletal growth and reproductive outcomes and of doxorubicin on congestive heart failure. Systematic study of the NWTSG database has challenged prevailing theories for the genetic origins of Wilms tumor and led to new hypotheses for investigation by molecular biologists. Key factors that facilitated these statistical contributions include a compulsive effort to maintain continuity in data collection and follow-up, a thorough understanding of the basic biology and a continuing search for ways to use clinical and epidemiologic data to answer questions of basic biological significance.

Session 12: Panel Discussion on Accreditation

Monday, June 5th, 15:30, Morisset Hall 218

Introduction by SSC Professional Development Committee

Ken McRae, Agriculture and Agri-Food Canada

Accreditation of statisticians is one way to ensure that statistics is done by professionals with the proper statistical training. A panel discussion will be held to explore the positives and negatives of an accreditation process for statisticians in Canada. Experiences gained from an attempt to accredit physics professionals a few years back will also be discussed. A discussion period will follow.

Session 13: Mathematical Modelling of TCP Control Flows Protocols

Monday, June 5th, 15:30, Montpetit Hall 201

TCP is $(\max, +)$

François Baccelli, École Normale Supérieure

A $(\max, +)$-linear system admits a representation by a vector of state variables (like the departure times of packets from a different nodes) which evolve according to a vectorial $(\max, +)$-linear evolution equation. The subject of this presentation is the representation of the Transfer Control Protocol (TCP) as $(\max, +)$-linear system. TCP is a window flow control used to avoid congestion in IP networks. A specified window size means only a fixed number of packets can be on route to their destination at a given time. An acknowledgment to the sender that a packet has arrived releases the next packet into the network. We will indicate some consequences of the fact that TCP is $(\max, +)$.

Session 14: Stochastic Processes and Time Series

Monday, June 5th, 15:30, Montpetit Hall 207

Cramer-Hida Decomposition Applied to Problems of Random Signal Detection

Adriana Climescu, University of Ottawa

The detection of a non-Gaussian stochastic signal in additive and dependent Gaussian noise can be viewed as the canonical detection problem for active sonar in a reverberation-limited environment. As superpositions of causal filters are acting on the noise and signal-plus-noise processes, the use of the Cramer-Hida decomposition is particularly appropriate. Also, an extension of the method to situations for which the noise is no longer Gaussian is studied. A likelihood formula is obtained for noises that are non-anticipative transformations of the sum of a Wiener process and an independent Poisson martingale.

Monday, June 5th, 15:45, Montpetit Hall 207

Estimation of Parameters of Multilevel Linear System of Stochastic Differential Equations

Dano Kako, Université de Sherbrooke, dano.kako@dmi.usherb.ca

We investigate the problem of parametric estimation for the N^L-dimensional process characterized by multilevel linear system of stochastic differential equations. We find the Maximum Likelihood Estimator, prove criteria for consistency and Local Asymptotic normality(LAN), when the size-N of each level of the hierarchy tends to infinity.

We compute the first two moments of all average processes and introduce another estimation based on those moments. We study the asymptotic properties of these last estimators.

Monday, June 5th, 16:00, Montpetit Hall 207

The Optimal Approximation Idea of Classical CuSum Procedure for Level-Shift Problem

Ying-Sheng Hu, University of Houston

In many applications, typically in quality control practice, detecting upwards/downwards shift(s) in the level of a process is often desired. This talk presents an optimal approximation point of view for level-shift, or change-point (abrupt) problem. The classical CuSum approach provided a statistical tool for dealing with single level-shift problem. For given ordered observations, the CuSum procedure forms a sample change process based on the cumulative sums of the deviations of each sample value from the sample mean. Under the null hypothesis of no change, the sample change process converges to the standard Brownian bridge stochastic process in distribution. But, if there is a change in mean the CuSum procedure specifies the change point, such that the sample change test process reaches its maximum. The optimal approximation approach is, on the other hand, seeking the unique optimal step-function in L2 for the data. For single level shift problem, the author has shown the equivalence between the two approaches. More interestingly, the optimal approximation point of view leads to an applicable and efficient solution for detecting multiple level-shifts (at once) via wavelets.

Monday, June 5th, 16:15, Montpetit Hall 207

Information Theoretic Approach to Multiple Change-Points Analysis

Arjun K. Gupta & Jie Chen, Bowling Green State University

The testing and estimation of multiple covariance change points for a sequence of m-dimensional ($m > 1$) Gaussian random vectors by using Schwarz Information Criterion (SIC) have been studied. We will estimate the number of change points as well as their locations. The consistency of the estimator is proved. The unbiased SIC is also obtained. Then asymptotic null distribution of the test statistic is derived. The result is applied to the weekly prices of Exxon and General Dynamics stocks ($m = 2$) from 1990 to 1991, and changes are successfully detected.

Monday, June 5th, 16:30, Montpetit Hall 207

Kolmogorov Test of Nonlinearity in Time Series: Percentage Points and Power Study

Gemai Chen, University of Regina & Min Chen, Chinese Academy of Sciences

Percentage points for carrying out the Kolmogorov test of nonlinearity in time series (An and Cheng, 1991) are provided in this paper. These points make it possible to apply the test for both small and large sample sizes, and for different error distributions such as the normal, student t and gamma distributions. The excellent power of the test for various types of nonlinearity is also demonstrated.

Monday, June 5th, 16:45, Montpetit Hall 207

Sequential Estimation, Confidence Sequences and Tests with Power One for Parameters of AR(P) and RCAR(1) Models

A.K. Basu, Calcutta University

In this paper, we have obtained confidence sequences using Hajek-Renyi Inequality and law of iterative logarithm for the auto-regressive parameters in a stable first order model. A sequential procedure with power one has also been proposed. We have also indicated a generalisation to confidence sequences for

several parameters in a p^{th} order model and parameter of a RCAR model. A sequential estimator based on least square estimator is studied here. It is shown that the sequential estimator is asymptotically risk efficient and stopping rule is shown to asymptotically consistent, as the cost per observation tends to zero.

Session 15: Survey Methodology

Monday, June 5th, 15:30, Morisset Hall 205

Design of the Canadian Community Health Survey (CCHS)

Yves Béland, Suzanne Lessard, Marianna Morano, Orane St-Denis & Sylvain Thivierge, Statistics Canada

A new study is under development which is intended to fill the principal statistical gaps in the determinants of health, and the use of the Canadian public health system, down to the level of socio-medical areas. The Canadian Community Health Survey (CCHS) will have its principal objective to produce estimates for more than 130 socio-medical areas of the country. The sample of 130 000 respondents will be selected starting from two frames: an areal frame (interviews in person) and random digit dialing telephone interviews. Data acquisition will begin in September 2000 and will be done monthly over a one year period in order to spread out the workload of the interviewers and to eliminate the seasonal effects on certain characteristics connected to health.

This paper will describe the sampling design of the CCHS. The design of the questionnaire, the distribution of the sample among the areas, the selection of the the multiple frames, the challenges connected to the estimate, and the strategy of dissemination of the results will all be discussed.

Monday, June 5th, 15:45, Morisset Hall 205

Overview of the Longitudinal Study of Canadian Immigrants

Sylvie Laroche, Statistics Canada

The longitudinal study among Canadian immigrants was created to obtain information about the immigrants coming to Canada. In particular, we wish to study the integration of those immigrants to the Canadian society, the factors that influence this integration, and the services used by the immigrants to facilitate their integration. The study will be divided into three waves: the first, six months after the immigrant's arrival in Canada; the second, one year and a half later; and the third, about two years after the second. We must also coordinate an initial interview with the immigrant at his arrival in the country in order to obtain his address and to prepare a strategy. The population targeted includes every immigrant coming to Canada. The expected sampling size of the third wave is 5,000 respondents. The interviews should begin in October 2000 or shortly thereafter. A pilot scheme was made in 1996-97, including the initial interview and the tryout of the first wave questionnaire. In this presentation, we will discuss the results obtained by this pilot scheme and present an overview of the survey.

Monday, June 5th, 16:00, Morisset Hall 205

Responses Obtained By Proxy: The Effect on Data Quality for the 1998 National Census Test

Denis Malo, Statistics Canada, maloden@statcan.ca

For household surveys, it is well known that responses are often obtained by proxy. In that situation, a person represents another member of his household. This method is very advantageous concerning costs related to data collection and diminution of total non-response rate. But what is the impact of this method

on data quality?

The presentation summarizes the work done to attempt to measure the effect of this mechanism used for the 1998 National Census Test. Firstly, a picture of the typical respondent will be given. Then, non-response rates for several variables will be compared for the respondent and the member(s) of his household that he represented. Finally, two variables particularly interesting in this context will be analysed in more detail.

Monday, June 5th, 16:15, Morisset Hall 205

Call Sheets Data of Telephone Survey - Utilisation and Validation

Serge Chevalier, Régie régionale de la santé et des services sociaux de Montréal-Centre Direction de la santé publique, scheval2@santepub-mtl.qc.ca

Call sheets data contains the result of every phone call made to achieve a survey. The data includes many types of information; for our purpose only the information pertaining to the call results are studied here (completed questionnaire, refusal, no answer, etc.). The available data permit the computation of the productivity rate (associated with the cost of the survey and the number of respondents), the response rate (a measure of the quality of the gathered data), and the weighting of the raw data. Theoretical elements will be presented as will the validation scheme of the data. Results indicate discrepancies between the productivity rate and the response rate computed before and after validation. Elements of the solution will be discussed.

Monday, June 5th, 16:30, Morisset Hall 205

Television Viewing: A New Method for Collecting Data

Jean-François Bastien, Statistique Canada, bastjea@statcan.ca & Laurent Roy,Statistics Canada, roylaur@statcan.ca

Statistics Canada's mandate includes collecting and publishing information about culture in Canada. As one element of the Culture Statistics Program, Statistics Canada and BBM are working together to provide information on television viewing of Canadians. Until fall 1998, diaries were used as the only tool for collecting TV viewing data. A new data collection method using electronic meters is being tested and planned to be implemented in the future for big urban centers. In November 1998, in Vancouver, the two methods for collecting data were used at the same time on two independent samples, thereby giving us a unique opportunity to compare parallel results from the two methodologies. The purpose of our study is to integrate data collected by the two methodologies in order to produce estimates that are comparable between regions and over time.

Session 16: Bayesian and Monte Carlo Methods

Tuesday, June 6th, 8:30, Montpetit Hall 207

Beyond Accept-Reject Sampling

François Perron, Université de Montréal

We propose a new estimator for estimating a quantity μ, via Monte Carlo simulations, where $\mu = \int f dP$ and P is the target probability measure. The estimation is based on the accept-reject sampling but the estimator is not the sample mean of f evaluated at the accepted observations. In fact, our new estimator involves the rejected observations. This estimator outperforms the usual accept-reject algorithm in terms of reducing the variance. Properties of the estimator are derived. A Rao-Blackwellised version of the estimator is also produced, but this version requires enormous computing time. Numerical results are provided.

Tuesday, June 6th, 8:45, Montpetit Hall 207

Generating Probability Distributions of Analyses using Ensembles of Forecasts and Bayesian Statistics

Gary Sneddon, Memorial University of Newfoundland

Ensemble forecasting is used in numerical weather prediction to give an improved estimate of the atmospheric state and to improve measures of forecast accuracy. However, there are open issues in interpreting the ensemble and specifying the large, complex covariance matrices in the model. We present an approach to this problem by representing the prior distribution as a mixture of Gaussian distributions, generate an ensemble from the posterior and use this ensemble to construct a kernel approximation to the posterior distribution.

Tuesday, June 6th, 9:00, Montpetit Hall 207

Bayesian Cross-Validation Choice and Assessment of Statistical Models

Fatemah Alqallaf, University of British Columbia

This talk will be concerned with application of a cross-validation criterion to the choice and assessment of statistical models, in which observed data are partitioned, with one part of the data compared to predictions conditional on the model and the rest of the data. We develop three methods, gold, silver, and bronze based on the idea of splitting data in the context of measuring prediction error; however, they can also be adapted for model checking. The gold method uses analytic calculations for the posterior predictive distribution; however, the silver method avoids this mathematical intensity, instead simulating many posterior samples, and the bronze method reduces the amount of sampling to speed up computation. We also consider the Bayesian p-value in which the posterior distribution can be used to check model adequacy, in the context of cross-validation with repeated data splitting. Application to examples is detailed, using the discussed methodologies of estimation and prediction.

Tuesday, June 6th, 9:15, Montpetit Hall 207

Bayesian Analysis of Longitudinal Multinomial Data arising from Developmental Research

Paul Gustafson & Lawrence J. Walker, University of British Columbia

Longitudinal multinomial data frequently arise in developmental psychology. This talk illustrates a Bayesian approach to the analysis of such data. Rather than infusing latent structure, we construct a prior distribution for the multinomial parameters which reflects the longitudinal nature of the observations. This permits posterior assessment of developmental hypotheses concerning the multinomial parameters, as we exemplify with the analysis of some data on the development of moral reasoning. We explore the posterior distribution by sampling, using Markov chain Monte Carlo methods.

Tuesday, June 6th, 9:30, Montpetit Hall 207

Bayesian Regression with Interactions and Smooth Effects

Nathan Johnson, University of British Columbia

The B-WISE (Bayesian regression with interactions and smooth effects) method for regression is meant to improve upon existing regression schemes in the areas of implementation and model interpretability. In the B-WISE scheme the effect of a predictor variable can be modelled using either a linear function or a cubic spline with fixed knots. A prior distribution over the model space penalizes models according to their roughness and the number of interactions present. In the interests of model parsimony, a predictor is allowed to interact with at most one other predictor. In this presentation I will introduce the B-WISE method, and show how it can be extended so that predictors can interact with any number of other predictors and so that categorical predictors can also be utilized.

Tuesday, June 6th, 9:45, Montpetit Hall 207

A Linear Empirical Bayes Method on the Problem of Multivariate Calibration

Keyue Ding & Rohana J. Karunamuni, University of Alberta

The well-known problem of multivariate calibration involves making inferences about an unknown vector X from a single random observed response vector Y. Two well-known competing estimators of X are the classical and inverse estimators. They can be obtained by direct or inverse regression and are supported by the maximum likelihood and Bayesian approaches, respectively. Here we exhibit another estimator, which is a convex combination of classical estimators, derived by exploiting the techniques of linear empirical Bayes methods. We show that the proposed estimator has better performance over the classical estimator in the sense of having a smaller generalized mean squared error. We also compare our estimator with the inverse estimator as well.

Session 17: Design Issues

Tuesday, June 6th, 8:30, Montpetit Hall 201

Sequential Multiple Testing of Several Binomial Proportions to a Specified Standard

Ka Ho Wu, Siu Hung Cheung and Man Lai Tang, Chinese University of Hong Kong

A single step procedure testing simultaneously the equality of several binomial proportions to a prespecified standard was given in Kulkarni and Shah (1995). In this article, we derive sequential testing procedures which are more powerful than the single step procedure. Our proposed procedure is an exact method. Thus the p-values are computed solely based on binomial formulas, rather than evoking Poisson or normal approximations. As a result, our method will be valid for all possible configurations of population proportions, avoiding possible unreliable testing conclusions from using approximation techniques. Relevant formulas to compute power are also provided to enable one to determine required sample sizes to achieve a predetermined level of power. The proposed procedure is illustrated by an example excerpted from a carcinogenicity study.

Tuesday, June 6th, 8:45, Montpetit Hall 201

Optimal Designs for Detecting Drug Interactions

Peter Kupchak, University of Toronto

We consider a logistic dose-response model that incorporates a parameter for interaction between two drugs. Design measures satisfying various optimality criteria are proposed, and minimax designs are investigated. The goal is to have available designs which can be easily implemented in practice, can efficiently detect the nature and the degree of interaction between two pharmacological or toxicological agents, and remain robust to misspecifications of the parameters in the model.

Tuesday, June 6th, 9:00, Montpetit Hall 201

Active Control Equivalence Studies : A Review of Statistical and Ethical Issues

Jean-Pierre Bélisle & François Bellavance, École des Hautes Études Commerciales

This presentation will review some statistical and ethical issues raised by clinical trials which are designed to show that an experimental treatment is as effective as a standard treatment, but not necessarily more effective. We will discuss some failings commonly observed in such equivalence studies, the use of hypothesis testing versus the use of confidence interval estimation, and the determination of an appropriate sample size. We will then examine two ethical questions: the determination of an equivalence criteria, i.e. of an acceptable true difference in the effectiveness of the 2 treatments; and the type of information that should be given to potential participants.

Tuesday, June 6th, 9:15, Montpetit Hall 201

Design Considerations for Construction and Calibration of Health Measures

Rollin Brant, University of Calgary

In this paper I explore design considerations for studies where subjective self-report by patients is taken as a benchmark for validation of more objective measures of health and illness. Following the same framework (Brant et al, 1999) established for analysing calibration designs involving between subject comparative assessment (Redelmeier et al, 1996), I consider that problem of scale construction from the perspective of weight development for composite scores based on of individual items or sub-scores. In addition, I explore the relative utility of cross-sectional (between subject) and longitudinal (within subject) designs.

Tuesday, June 6th, 9:30, Montpetit Hall 201

Design Issues of a Phase III Randomized Trial incorporating a Phase II Portion

Benny Zee, Queen's University

Recent development of a class of cancer treatment such as matrix metalloproteinase inhibitor (MMPI) has been shown to possess anti-angiogenic and anti-metastatic effects that may prolong survival but do not lead to tumour size reduction. Conventional Phase II study design in cancer drug development using tumor response alone as endpoint may not be appropriate. In this paper, we will discuss a few design issues for Phase II study done within a Phase III framework. The sample size for the Phase III study will be estimated given the Phase II information within the same study. A bias estimate for the log hazard rates will be incorporated in the sample size calculation and the impact of stopping the study early at the Phase II stage will be discussed.

Tuesday, June 6th, 9:45, Montpetit Hall 201

Setting Standards: Experiences and Statistical Issues

J. Terry Smith, Queen's University

The question of setting standards arises in a variety of contexts outside of traditional quality management, for example, reference values in clinical chemistry, physical fitness standards for physically demanding occupations, limits on pollutants in effluent from manufacturing plants, and characterization of abnormal human motion in the performance of a task. This talk will draw on the author's experience with consulting projects in these areas. Statistical issues will be identified and some problems, not all statistical, will be explored.

Session 18: Strong Limit Theorems

Tuesday, June 6th, 8:30, Montpetit Hall 203

Some Limit Theorems for Independent and Weakly Dependent Random Variables

Walter Philipp, University of Illinois at Urbana-Champaign

This is a report on some recent results, obtained jointly with I. Berkes and R. Tichy, on pair correlations for independent and some weakly dependent random variables. If $\{X_j\}$ is a sequence of random variables with uniform distribution over [0,1], the pair correlation function counts the number of pairs (j,k) with $1 < j,k < N$ such that the distance of X_j and X_k differs by less than t. This is a U-statistic with kernel being the indicator of the interval $[-t,t]$. We prove Glivenko Cantelli type theorems for these U-statistics when $t = t(N)$ tends to 0 at certain rates. We also show that for polynomially decreasing rates $t(N)$ our results are sharp save for powers of $\log N$.

Tuesday, June 6th, 9:15, Montpetit Hall 203

Some New Limit Theorems for (Dependent) U-statistics

Herold Dehling, Rijksuniversiteit Groningen

In our talk we will survey some recent results concerning the limit behaviour of U-statistics with dependent observations. We will investigate laws of large numbers, central limit theorems and empirical processes. The observations are supposed to arise as functionals of an absolutely regular process, as is often encountered in dynamical systems.

Session 19: Two-phase Estimation

Tuesday, June 6th, 8:30, Montpetit Hall 202

The Use of Regression in Multi-Phase Sampling

Wayne Fuller, Iowa State University

In some surveys information is collected at different degrees of detail for samples that are subsamples of the original sample. The subsamples are called phases when estimation information is collected at levels other than the final subsample. Regression estimators and the use of regression in variance estimation for multiple phase samples will be discussed. Consideration will be given to samples in which the first phase data are of subject matter interest as well as a source of data for estimation. Replication variance estimation will be discussed and examples will be presented.

Session 20: Statistical Software

Tuesday, June 6th, 8:30, Morisset Hall 205

Taking Advantage of the New Version of SAS

Jean Hardy & Jacques Pagé, Services Conseils Hardy Inc.

Version 8 of the SAS System hit the street recently, with new statements, functions and features. Among all those enhancements, some will have significant impact on statisticians and will be reviewed in this session: easier access to DBMS's, better integration with Internet, a comprehensive and flexible set of tools to adapt SAS output (including PROC TABULATE) and export results toward other applications. New procedures for sampling, kriging analysis and analysis of survey data are also reviewed. Finally, we examine issues related to migrating V6 files and programs to the new version.

Session 21: Gold Medal Address

Tuesday, June 6th, 10:30, Tabaret Hall Chapel

Survival Analysis and Likelihood Methods

Jerry Lawless, University of Waterloo

Survival analysis is one of the most widely used branches of statistical science, and likelihood is the basis for much statistical methodology. This talk will survey the historical interplay between survival analysis and likelihood, and describe some applications. It will be shown how survival analysis has motivated new developments, including partial, empirical, and semiparametric likelihood.

Session 22: Edit and Imputation

Tuesday, June 6th, 13:30, Montpetit Hall 202

A Generic Implementation of the Nearest-Neighbour Imputation Methodology

Michael Bankier, Paul Poirier, Martin Lachance & Patrick Mason, Statistics Canada

A Nearest-neighbour Imputation Methodology (NIM) was introduced in the 1996 Canadian Census to carry out the hot deck Edit and Imputation (E&I) for the demographic variables. For the first time, qualitative and quantitative variables could be imputed simultaneously. The NIM used a data-driven approach with single-donor imputation to determine the best imputation action. For the 2001 Canadian Census, the NIM will be extended to more variables in accordance with the long-term objective to progressively move all census variables to the NIM for the 2006 Canadian Census. This expansion of NIM's use provides the opportunity to make a generic implementation of the methodology. The generic implementation will have new features under consideration, such as the evaluation of many numeric variables in a single edit rule and the capability of treating non-linear propositions. The flexibility of the generic implementation also includes the possibility of using the NIM in a wider variety of surveys.

Tuesday, June 6th, 14:00, Montpetit Hall 202

The Generalized Edit and Imputation System and its Recent Developments

Claude Poirier & Sanping Chen, Statistics Canada

The Generalised Edit and Imputation System (GEIS) was developed at Statistics Canada to meet the imputation requirements of the Canadian economic surveys. Being entirely driven by linear edit rules, GEIS processes numeric data that are assumed to be continuous and non-negative. Linear programming techniques are used to localize fields to be imputed and search algorithms are used to perform automatic imputations. Amongst its imputation methods, there are the donor approach, the imputation by estimators, and the prorating. Recent developments were initiated to make the system more practical. The desired modifications were identified based on past experiences with up to 30 economic surveys. The modules are being redesigned to be used independently from each other, the job submission is being made more user-friendly in order to process several sub-populations together, and the system will be modified to interact with SAS data sets in addition to Oracle databases. This paper describes the current functionality, the planed changes and the new architecture, which will provide the users with a more flexible product. Future directions are also described.

Tuesday, June 6th, 14:30, Montpetit Hall 202

A Look into AGGIES, An Automated Edit and Imputation System

Kara Perritt, U.S. Department of Agriculture

The National Agricultural Statistics Service (NASS) currently uses several edit and imputation tools to ensure that data reported by respondents are consistent and complete. Such tools include manual editing and imputation, interactive micro level editing with manual imputation, batch micro level editing with manual imputation, interactive macro level editing with manual imputation, and, for some surveys and censuses, automated computer imputation after editing has occurred. Taken together, these tools can manage even the most complex edit and imputation schemes and aggregate level impact can be evaluated. However, since these tools are not integrated, maintenance is costly, redundancy is apparent and neither editing nor imputation is always performed in a consistent manner across, and even within, State offices. The Research and Development Division of NASS is developing and evaluating an automated edit and imputation system called the Agricultural Generalized Imputation and Edit System (AGGIES). Using methodology based on GEIS, Statistics Canada's Generalized Edit and Imputation System, AGGIES is programmed in SAS with object-oriented features that make for a user-friendly environment. The system, designed to edit non-negative, continuous values, is structured in a modular fashion and offers NASS several potential advantages over current edit and imputation procedures. First, the editing and imputation functions are fully automated eliminating the need for the extensive manual data reviews currently done. Second, these functions are performed objectively allowing for more consistency throughout the editing process. Third, it is written

in a language that is heavily used throughout the agency. Thus, integration with tools currently being used will be simplified. Finally, the system can be easily applied to any number of surveys and censuses, thereby conserving resources to the development and maintenance of a single system. Evaluations using data collected through surveys and censuses conducted by NASS have been completed by comparing AGGIES output to current processing output. First, this paper provides a system overview that briefly describes each of the following AGGIES modules: edit specification, check edits, error count report, outlier detection, error localization and imputation. Second, an evaluation discussion of AGGIES as a core tool in NASS's complete edit and imputation strategy will be presented.

Session 23: Biostatistical Applications of GEEs

Tuesday, June 6th, 13:30, Montpetit Hall 203

The Analysis of Case-Control Family Data

John Neuhaus, University of California, San Francisco, Alastair Scott & Chris Wild, University of Auckland, New Zealand

Genetic epidemiologists often augment an initial case-control sample with responses and covariates gathered from the family members of the initial sample persons (probands). The objective with such data is to more efficiently estimate the associations of interest in the case-control sample, to estimate associations controlled for family characteristics and propensities and to measure familial aggregation (within-family dependence) of the response. This talk describes methods to estimate the associations of interest by fitting marginal and cluster-specific generalized linear models to case-control family data. Example data from case-control family studies of cancer will illustrate our methods and results.

Tuesday, June 6th, 14:00, Montpetit Hall 203

Handling MAR Drop-outs in Regression Models Analyzed Using GEE's

R.J. O'Hara Hines, University of Waterloo

Recent work in the analysis of clustered (longitudinal) data with a combination of continuous and categorical responses has involved examination of the unbiasedness and efficiency of several methods proposed for handling drop-outs (Robins, Rotnitzky and Zhao,1995, Fitzmaurice, Molenbergs and Lipsitz, 1995 and Lipsitz, S.R., Fitzmaurice, G.M., Molenberghs, G. and Zhao, L.P., 1997). This work has generated further interest in methods for handling drop-outs in clustered data using generalized estimating equations. Progress will be reported on exploring the strengths and weaknesses of these drop-out schemes, and on examining the implications for interpretation of the results. We will also attempt to provide answers to such questions as when, from a practical point of view, drop-outs should be considered, and what are the effects on the interpretations of ignoring or adjusting for drop-outs and of misspecification of the drop-out model on conclusions of analyses. The data set used for demonstration models the size and stage of nests of young bass over the lifetime of the nest. A nest which disappears before maturation is considered a form of drop-out.

Tuesday, June 6th, 14:30, Montpetit Hall 203

Efficiency Aspects of Working Correlations Based GEEs in Longitudinal Studies

Brajendra C. Sutradhar, Memorial University of Newfoundland

In a seminal paper, Liang and Zeger (Biometrika, 1986) introduced a generalised estimating equations (GEE) approach based on a "working" correlation matrix to obtain consistent and efficient estimators of regression parameters in the class of generalised linear models for repeated measures data. Recently, in contrary to Liang and Zeger's claim, Sutradhar and Das (Biometrika, 1999) have, however, demonstrated that the "working" correlations based GEE approach usually produces less efficient estimators as compared to the regression estimators obtained by using the independence estimating equations (IEE) approach. Sutradhar

and Das have also provided a simple generalized regression estimator as a resolution to this problem. In this talk, I will discuss how to develop GEE following Sutradhar and Das to deal with multivariate longitudinal data and demonstrate the application of this new procedure by analyzing a longitudinal data set on health care utilization by families in the city of St. John's, NF.

Session 24: Statistics in Business

Tuesday, June 6th, 13:30, Morisset Hall 218

Quality Control in Statistical Surveys

Johanne Thiffault, Ministère des Transport du Québec

In the transportation field, quality control techniques and survey methodologies generally vary whether the inquiry deals with the population's mobility or the transportation of goods. In all kinds of inquiries, the principal role of the statistician is to ensure the quality of the data and its proper use. Even if there exist many ways to accomplish this type of analysis, every approach raises the same concerns for the statistician: sample representation, reduction of the observed error, and analysis of the data. This presentation will discuss two inquiries that were made over the last two years in which the Ministry of Transportation of Quebec participated. The first investigation, made in Fall 1998, investigated the mobility of the population in the Montreal area. The second dealt with road borders and was carried out in Summer 1999 for the Canadian Council of Administrators in Motorized Transportation, in conjunction with Transportation Canada and the other provinces. Among the measures included in the first inquiry, we note the integration of a number of interactive information validation functions in the software for telephone interviews, automatic verification and manual correction of the data, followups of the sample progression, and the development of a series of indicators to measure the performance of the inquirers. For the second investigation, field interviews allowed better quality control. Moreover, preliminary analyses performed on the data collected during the first few weeks allowed detection the most common errors. A bulletin for the inquirers was also created.

Tuesday, June 6th, 14:00, Morisset Hall 218

Statistical Aspects of Surveillance in Business

Alex Whitmore, McGill University

Many commercial and industrial processes require auditing, inspection or monitoring to establish their condition for purposes of reporting, planning and control. The processes may range from business accounting systems to industrial environmental monitoring. The generic term surveillance describes this activity. Surveillance data take the form of statistical readings on the underlying process and generally provide an incomplete picture of the process. Hence, statistical inferences must be drawn from the data about the true state of the process. This talk is derived from an ongoing research program that is developing a conceptual framework for surveillance systems, examining their optimal design and management, considering the implications for managers dependent on such systems, while paying particular attention to statistical concerns. This talk will present a non-technical survey of some ideas and issues connected with this subject. The talk will include a case application that highlights some of the interesting and challenging features of statistical work in this area of business and management.

Tuesday, June 6th, 14:30, Morisset Hall 218

Structural Equation Modeling Applied to Strategic Alignment of Information Technology

Anne-Marie Croteau, Concordia University

Much research in the field of Management Information Systems makes use of very sophisticated statistical models. This talk will show how statistics has become a major tool in the assessment of Business problems. A concrete research problem will be presented. An empirical study was conducted to measure the level

of strategic alignment between business strategy and information technology among 223 Canadian companies. The objective was also to identify the profiles of technological deployment specific to various types of business strategy that best support organizational performance. Using structural equation modeling, a confirmatory factor analysis approach was adopted and the estimation of multiple and interrelated dependence relationships was also performed. Questionnaires completed by both CEOs and CIOs were analyzed with a statistical tool called Partial Least Square (PLS). PLS belongs to the second generation of multivariate statistical analysis and is a regression-based technique, with roots in path analysis that can estimate and test the relationships among constructs. It produces loadings between items and constructs and estimates standardized regression coefficients (i.e., beta coefficients) for the paths between constructs. PLS analysis involves two stages: (1) the assessment of the measurement model, including the item reliability, convergent validity, and the discriminant validity, and (2) the assessment of the structural model. Together, the measurement and structural models form a network of constructs and measures. The item weights and loadings indicate the strength of measures, while the estimated path coefficients indicate the strength and the sign of the theoretical relationships.

Session 26: Science and Statistics

Tuesday, June 6th, 13:30, Montpetit Hall 201

Vitamin E: Or Why We Don't Go Rancid

I.U. Ingold, National Research Council of Canada

What is it? Why is it important? How reliable are bioassays? What form should I (or my female partner) take?

Tuesday, June 6th, 14:15, Montpetit Hall 201

Uncertainty, Precision and Rounding: How Spectroscopists Use/Abuse Regression

R.J. Le Roy, University of Waterloo

Molecular spectroscopists determine patterns of molecular energy levels by measuring, often to extremely high accuracy, many hundreds or thousands or tens of thousands of discrete absorption or emission lines due to particular energy level spacings. These are then fitted by least-squares to some analytic function based on some physical model for the system to determine numbers (sometimes several, sometimes tens, sometimes hundreds and sometimes thousands) of molecular parameters, some of which have physical significance and some of which do not. However, in the spectroscopic literature the treatment of uncertainties and the rounding of reported parameters has historically often been treated somewhat "cavalierly" (to be generous!). I will illustrate this problem, describe our recent suggestions regarding a more systematic approach to it, and raise a question concerning "averaging over models" which is still being addressed in an unsatisfying manner.

Session 27: Survey Methodology

Tuesday, June 6th, 13:30, Morisset Hall 205

Sampling, Unequal Probabilities and Balancing

Yves Tillé, École nationale de la statistique et de l'analyse de l'information

The splitting method allows a unified representation of all the unequal probability sampling methods. This splitting representation can be interpreted in a more fruitful way when each sample is defined as the vertex of a hypercube. A sampling design is then defined as a probability distribution on the vertices of the cube. The balancing constraints can thus be viewed as a hyperplane that cuts off this cube. The definition of a balancing martingale allows random selection of a vertex-sample that satisfies a set of balancing constraints.

Session 28: Time Series Methods Applied in Statistical Agencies

Tuesday, June 6th, 15:30, Montpetit Hall 202

Survey Error Modeling in the Presence of Benchmarks

Zhao-Guo Chen, Statistics Canada & Ka Ho Wu, Chinese University of Hong Kong

Data for a socio-economic variable obtained from a repeated survey are with error. Usually, some estimates of the variance of the error are obtained in the survey process and published, but estimates of the correlation of the error (equivalently, time series models fitted to the error) are rarely given. If individual unit values, or other elementary estimates such as the values for rotating panels, are available, the method to model the survey error is referred to as primary analysis. However, usually analysts have access only to the aggregated data. Methods under such conditions are referred to as secondary analysis. These methods are usually developed using very subjective assumptions. It was ignored by the literature that, with aggregated data, some auxiliary sources may often be available to help in analyzing the survey error, so we can get more data-based results. This approach will be referred to as extended secondary analysis. This paper proposes a method to model the survey error with the AR(1) model when some benchmarks from an auxiliary source are available.

Tuesday, June 6th, 16:00, Montpetit Hall 202

Interpolating Monthly Values from Data covering Various Fiscal Periods

Pierre Cholette, Statistics Canada

Data supplied to statistical agencies often refer to fiscal periods, i.e. fiscal years, fiscal quarters, fiscal months, instead of target calendar periods. This is the case of the records in the Canadian Goods and Services Tax administrative file examined here, which exhibit a multitude of fiscal periods. This paper presents a method to calendarize such data. The model assumes that, within one class (i.e. industry and province), the individual businesses share a common seasonal pattern but have their own level, trend and autocorrelated disturbance. In principle, the method can be used to interpolate values of any periodicity.

Tuesday, June 6th, 16:30, Montpetit Hall 202

North American Industrial Classification System (NAICS) Time Series of U.S. Retail and Wholesale Estimates

Ruth E. Detlefsen & Michael Z. Shimberg, U.S. Census Bureau

The U.S. Census Bureau maintains time series of monthly sales and inventory data for retail and wholesale Standard Industrial Classification (SIC) codes. New samples have been selected to provide estimates of those items for North American Industrial Classification System (NAICS) codes within the retail and wholesale sectors. The first NAICS-based monthly estimates will be released to the public for the March 2001 data month. Simultaneously, NAICS-based time series beginning with the January 1992 data month will be made available. This paper discusses the methods that will be used to develop the NAICS-based time series and addresses related quality issues.

Session 29: Y2K National Health Statistics

Tuesday, June 6th, 15:30, Montpetit Hall 203

The U.S. National Health Interview Survey

Jane F. Gentleman, National Center for Health Statistics

The National Health Interview Survey (NHIS) is one of the major health data programs in the United States. Through the NHIS, information concerning the health of the U.S. civilian noninstitutionalized population

is collected in in-person household interviews throughout the country. The NHIS has been in continuous operation since 1957. The data are collected during 50 weeks of the year and are released annually. The survey is conducted by the National Center for Health Statistics (NCHS), which is located in the Washington, D.C. area and is a component of the Centers for Disease Control and Prevention (CDC) in Atlanta. The U.S. Bureau of the Census, under contract to NCHS, has been the primary data collector for the NHIS since the survey's inception. Annually, the survey obtains information from about 40,000 households containing about 100,000 people. Topics covered by the NHIS include household composition, health status and limitation of activity, injury, health care access and utilization, health insurance, socio-demographic characteristics, income and assets, conditions related to functional limitations, health behaviors, immunization, AIDS, and other topics covered by supplementary questions that vary from year to year. This talk will describe key aspects of the design and content of the NHIS and present some results of analysis of NHIS data.

Tuesday, June 6th, 16:00, Montpetit Hall 203

Investing in Health Surveys, Investing in Health

Lorna Bailie, Statistics Canada

Interesting and exciting times are ahead as Canada moves towards greater investment in new health information. The priorities, inaugurated with the 1998 Health Information Roadmap, promise stimulating and critical new insights into the health of Canadians. Data collection on health will be extended to the community level, new concepts will be developed in the areas of mental health, nutrition, and physical measures and survey frames will be put in place to react to changing data needs. This session will highlight the main survey initiatives being explored and some of the challenges facing the expansion of health collection activities.

Tuesday, June 6th, 16:30, Montpetit Hall 203

Statistics in the Health Protection Branch of Health Canada

William Ross, Health Canada

The Health Protection Branch is a key component of Canada's public health system. It provides those programs through which Health Canada contributes to the protection of the Canadian population against health risks. These programs are wide ranging but revolve around two broad areas - (i) the safety of therapeutic, industrial and consumer products; and (ii) control and prevention of disease. New and emerging public health issues, together with rapidly changing information technologies, have created new challenges for the discipline of statistics as applied in regulatory bodies. In addition, they raise questions as to the role of statistics and statisticians in setting good health policies. In this presentation, we summarize and discuss some of these challenges as they relate to the use of modern statistical techniques. We go on to identify some opportunities for statistics (and statisticians) in the area of health protection.

Session 30: Directional Statistics

Tuesday, June 6th, 15:30, Morisset Hall 218

Superefficient Estimation of Directional Trend

Rudolf Beran, University of California, Berkeley

Consider a fixed effects one-way layout with one independent directional observation per factor level. Measured positions of the earth's magnetic pole as a function of discrete time illustrate the design. The goal is to estimate the mean direction at each factor level. The naive, classically efficient estimate of this directional trend is the raw data itself. We construct superefficient trend estimators that dominate the naive estimator substantially under quadratic loss. The reduction in risk is obtained through an adaptive shrinkage strategy in a transformed coordinate system that accomplishes a form of smoothing.

Tuesday, June 6th, 16:00, Morisset Hall 218

Geometry of Score Tests and Wald Tests

Peter Jupp, University of St. Andrews

Although score tests and Wald tests form two of the three major families of standard tests in parametric inference, they each have major disadvantages. The theme of this talk is that geometrical ideas can be useful in overcoming these disadvantages. Unlike likelihood ratio tests, score tests do not, in general, have a Bartlett correction. The first part of the talk will discuss some cubic modifications of score tests which bring the null distribution of the score test statistic close to its (large-sample) asymptotic chi-squared distribution. Geometric interpretations of the terms appearing in these corrections will be emphasised. Unlike the likelihood ratio and score tests, Wald tests have the considerable disadvantage that they are not invariant under change of parameterisation. The second part of the talk will describe some parameterisation-invariant versions of Wald tests. Their construction uses a little differential geometry.

Tuesday, June 6th, 16:30, Morisset Hall 218

A New Method to Solve Noisy Abel Equations

Frits Ruymgaart, Texas Tech University

Astronomers can measure only the projections of vectors onto a plane perpendicular to the line of sight. The distribution of the true major axis of an ellipse is then given by the solution of Abel's integral equation. Such equations have also applications in several other areas like stereology (Wicksell's unfolding problem). A new approach to solving noisy integral equations of the first kind will be applied to the family of Abel equations. The method is based on an expansion in an arbitrary orthonormal basis, coupled with exact inversion of the integral operator. The inverse appears in the Fourier coefficients of the expansion where it can be carried over to the usually well behaved basis elements in the form of the adjoint. This method is an alternative to Tikhonov regularization, regularization of the inverse of the operator itself, or a wavelet-vaguelette/ SV decomposition. The method is particularly interesting when either the kernel, the input, or both are irregular. Since knowledge of the spectral properties of the operator is not required, the method is also of interest in regular cases where these spectral properties are not sufficiently known or hard to deal with. For smooth input functions the simple basis of trigonometric functions yields input estimators whose MISE converges at the optimal rate for the entire family of Abel operators. Surprisingly, there are strong indications that for a subfamily of operators the optimal rate is not attained when the estimators are derived from a wavelet basis.

Session 32: Theoretical Survey Methods

Tuesday, June 6th, 15:30, Montpetit Hall 207

A Nonlinear Hierarchical Modeling Approach for Census Undercoverage Estimation

Yong You, Statistics Canada

An area-level nonlinear mixed effects model is considered for the province-level census undercoverage survey estimates. In particular, the sampling model is based on the survey estimate of the undercoverage count, and the linking model is a log-linear random effects model for the undercoverage rate, which is expressed as a nonlinear function of the undercoverage count. A full hierarchical Bayes (HB) approach is developed to obtain the posterior estimates of the census undercoverage count and rate. Overall model validation and model fitting at the individual observation level are also studied based on posterior predictive distributions. Our results have shown that the proposed HB estimator improves the direct survey estimator and the proposed model fits the data quite well.

Tuesday, June 6th, 15:45, Montpetit Hall 207

Small Area Estimation with Unmatched Sampling and Linking Models

Yong You, Statistics Canada & J.N.K. Rao, Carleton University

In small area estimation, a sampling model on transformed survey estimates and a linking model on transformed population means are often used. This leads to a linear mixed effects and standard results in empirical Bayes (EB) or hierarchical Bayes (HB) can be applied to get improved small area estimates. But the assumption of zero mean sampling errors may not be valid in small samples due to nonlinear transformation. To avoid this difficulty, we propose a sampling model on the original survey estimates and then combine it with a linking model on the transformed population means using a complete HB approach. A numerical study is conducted to compare the posterior means and posterior variances under the complete HB approach with those obtained under the customary linear mixed model approach.

Tuesday, June 6th, 16:00, Montpetit Hall 207

Design-Based Inference for Domain Totals under Imputation for Missing Data

D. Haziza & J.N.K.Rao, Carleton University

A population total, $\sum y_i$, can be estimated unbiasedly under random imputation for missing data and uniform response. But the imputed estimator of a domain total is generally biased because a domain total is of the form $\sum y_i x_i$, where x_i is the domain indicator variable. Following the approach of Skinner and Rao (1999), we obtain a bias-corrected estimator of a domain total under simple random sampling. We derive design-consistent variance estimators using jackknife as well as a method introduced by Fay (1991) in which the usual sample-response path is reversed. We extend these results to the case of stratified multistage sampling.

Tuesday, June 6th, 16:15, Montpetit Hall 207

Estimating Function Resampling Variance Estimators under Stratified Multistage Sampling

M. Tausi & J.N.K. Rao, Carleton University

Variance estimation for the poststratified estimator and the generalised regression estimator of a total under stratified multistage sampling is considered. Customary resampling methods (jackknife, balanced repeated replication and bootstrap) require the inversion of a $p \times p$ matrix for each resample, where p is the number of auxiliary variables. This could lead to illconditioned matrices for some of the resamples. We apply the estimating function resampling method of Hu and Kalbfleisch (1999) to obtain variance estimators. This method avoids repeated inverses, and we show that the resulting variance estimator is identical to a jackknife linearization variance estimator of Yung and Rao (1996). We extend the results to cover parameters defined as solutions of census estimating equations.

Tuesday, June 6th, 16:30, Montpetit Hall 207

Using Empirical Likelihood Methods to Obtain Positive Weights in Regression Estimation for Complex Surveys

Changbao Wu, University of Waterloo

The generalized regression estimator has been widely used in survey sampling when known population totals of auxiliary variables are available. One of the drawbacks of this method is that some of the regression weights can be negative for finite sample size. Some iterative procedures to adjust the weights to be positive exist (Fuller, 1968; Husian, 1969; Huang and Fuller, 1984; and Fuller, Lohghin and Baker, 1994). In this work, we present two general methods to obtain positive weights through the use of empirical likelihood and model-calibration methods. In both cases we construct estimators with positive weights and the algorithms are guaranteed to converge. Our second algorithm is particularly simple since it is not iterative, requires no

seeds, and guarantees that if a solution exits, it will be obtained. The problem of obtaining range restricted weights with these algorithms is also discussed.

Session 33: Applied Survey Methods - I

Wednesday, June 7th, 8:30, Montpetit Hall 207

Modified Regression Trees for Complex Survey Data

Anjela Tzontcheva, University of Toronto

Tree-based models come up as an alternative to linear additive models for regression problems and to linear logistic and additive logistic models for classification problems. Tree-based modeling includes growing classification and regression trees (CART). CART is a computationally intensive exploratory analysis tool, which attempts to describe the structure of the data in a tree-like fashion. The objective of modifying CART for complex survey data is to cope with the computationally infeasible task of growing binary trees, where a particular covariate may have thousands of observed values. In particular, the applications are to the Ontario Health Survey data of 1990, where there are 61,239 observations. As it is not possible to consider all possible splits in growing the tree, the data must be binned along covariates to reduce the size of the computation. Thus, trees may be grown in the usual way, where covariates are the midpoints of the bins and the response means. The binning is done in SAS and the trees are grown in S-Plus.

Wednesday, June 7th, 8:45, Montpetit Hall 207

Estimation in a Traffic Network Survey

A. Demnati, J. Lawson, B. Harbour, J. Young & G. Robbins, Statistics Canada

Roadside surveys such as the National Roadside Study conduct roadside observations and interviews to produce a profile of the volume and characteristics of the activity on a road network. The data collected at each site consists in one part of a random sample of interviews and observations of the trucks, and in another part of a series of count of trucks passing the site during the given time period. Data collected from different sites are then integrated into a single data set. The presentation will show the steps involved in this integration using a "toy" example. The "toy" example is also used as a motivation for the methodology. General formulae will also be given.

Wednesday, June 7th, 9:00, Montpetit Hall 207

Real-Time Editing: Some Technical Issues

Sanping Chen, Paul Hunsberger & Claude Poirier, Statistics Canada

Real-time editing within the data collection process allows early detection and follow-up of problematic records, improving the operational efficiency and timeliness. The challenge is the lack of the full knowledge of the current-cycle data, as required by almost all data editing and/or outlier detection regimes. Real-time editing tries to use proxies or estimates of individual current-cycle records, which are gradually replaced by the true records in the collection process. This can be viewed as some sort of sequential decision process. Technical issues regarding real-time editing are: The uncertainty level and its relationship with the collection sequence, dynamic boundaries based on collection time, retrospective decisions and their impact, and some group-sequential methods whenever editing is conducted at the end of a day or a week.

Wednesday, June 7th, 9:15, Montpetit Hall 207

The Effect of Political Expectations on Forecasting

Ante Rozga, University of Split, Croatia

Time series analysis has proved to be efficient tool in forecasting future economic activities. This is particularly true in predominantly market-oriented countries with stable economies, such as EU, USA or Canada, but not in transition countries with a great deal of uncertainty. Since these countries are not completely part of western (capitalist) economies, forecasting future economic activities is influenced by political changes, and the improvements of democratic standards. These changes affect economic activities because foreign investments and globalization of the economy depend very much on the risk involved. Good example is Croatia. At the beginning of the war Croatia lost 40% of GDP in only one year, followed by a slight recovery in the years after the liberation of its territory. There was an upward trend all the time, but in the year 1999 the GDP went down by 2%. This is particularly due to bad political decisions, keeping the country in some kind of isolation and blocking foreign investments by bad low regulation, and encouraging bureaucracy and corruption. After the death of president Tudman, there are strong signs of recovery, and great expectations in the stock exchange, as well among the western analysts. The other intervention came after general elections with the installation of social-democratic government who would like to get as close to western political and economic standards as possible. In this paper we will show great differences between forecasts using time series analysis and reality, particularly after turning points.

Session 34: Statistical Methods for Correlated Data

Wednesday, June 7th, 8:30, Montpetit Hall 203

Analysing Categorical Data with Multiple Responses Per Subject

D. Roland Thomas & Yves J. Decady, Carleton University

Several methods for analyzing multiple response data have been recently proposed (e.g., Loughin and Scherer, 1998, Biometrics; Agresti and Liu, 1999, Biometrics). However, these methods are computationally intensive and are based on techniques that are relatively unfamiliar to many practitioners. In the paper, the methods introduced by Rao and Scott (Journal of the American Statistical Association, 1981, 1984) for analyzing complex survey data are used to develop easily applied test procedures based on corrected chi-squared statistics. Tests of simple goodness of fit, equality of multiple response proportions, and tests of association in two-way tables (single by multiple and multiple by multiple) will be described, and evaluated using simulation.

Wednesday, June 7th, 9:00, Montpetit Hall 203

Applying Sample Survey Methods to Clinical Trials Data

Lisa M. Lavange, Quintiles, Inc., G.G. Koch, University of North Carolina, Chapel Hill & De Montfort University

This presentation explores the utility of statistical methods for sample surveys in analyzing clinical trial data. Sample survey statisticians face a variety of complex data analysis issues deriving from the use of multi-stage probability sampling from finite populations. One such issue is that of clustering of observations at the various stages of sampling. Survey data analysis approaches developed to accommodate clustering in the sample design have more general application to clinical studies in which repeated measures structures are encountered. Situations where these methods are of interest include multi-visit studies where responses are observed at two or more time points for each patient, dermatology studies where treatment is applied to two or more sites on each patient, multi-period crossover studies, and epidemiological studies for repeated occurrences of adverse events or illnesses. We describe statistical procedures for fitting multiple regression models to sample survey data that are more effective for repeated measures studies with complicated data structures than the more traditional approaches of multivariate repeated measures analysis. In this setting, one can specify a primary sampling unit within which repeated measures have intraclass correlation. This intraclass correlation is taken into account by sample survey regression methods through robust estimates of the standard errors of the regression coefficients. Regression estimates are obtained from model fitting estimation equations which ignore the correlation structure of the data (i.e., computing procedures which assume that all observational units are independent or are from simple random samples). The analytic

approach is straightforward to apply with logistic models for dichotomous data, proportional odds models for ordinal data, and linear models for continuously scaled data, and results are interpretable in terms of population average parameters. Through the features summarised here, the sample survey regression methods have many similarities to the broader family of methods based on generalised estimating equations (GEE). Sample survey methods for the analysis of time-to-event data in clinical trials are also described. Potential problems with the proposed methods and ways of addressing them are discussed, including the impact of small numbers of primary sampling units on the performance of the test statistics and the impact of missing observations on the resulting estimators. The capabilities of each of the methods discussed are illustrated with examples using the SUDAAN software system.

Wednesday, June 7th, 9:30, Montpetit Hall 203

On the Use of Spatial Analysis Methods in Assessing the Relationship Between Community Air Pollution and Mortality

Richard T. Burnett, Health Canada, Michael Jerrett, McMaster University, Daniel Krewski, University of Ottawa

The effects of long-term exposure to ambient air pollution on longevity have been examined by comparing age-sex adjusted mortality rates at the city or county level with corresponding outdoor concentrations of selected air pollutants, such as particulate matter, averaged over several years. Geographically specific mortality rates are assumed to be independent observations. These studies have been criticised on the grounds that only summary information on health outcomes and risk factors are used, potentially leading to the "ecological fallacy". In recent years, researchers have used cross-sectional surveys of individuals to obtain individual risk factor information on variables such as smoking habits, education, and diet. The cross-sectional survey is then turned into a longitudinal cohort study with follow-up on vital status. Exposure to ambient air pollution is still measured at the geographic level, with each individual residing in a geographic area being assigned the same exposure value. Observations on individuals from this hybrid design, sometimes referred to as "individual-ecologic", are assumed to be statistically independent. We present new methods to model dependency in the "individual-ecologic" design by using random effects models for survival analysis coupled with geographic modelling techniques to account for spatial clustering in health and risk factor data. Our approach is illustrated with data from the American Cancer Societies' Cancer Prevention II Survey and its link with ambient air pollution. The results from models that incorporated spatial autocorrelation differed significantly from those generated with traditional individual-ecological models.

Session 35: Nonparametric Ranking Methods

Wednesday, June 7th, 8:30, Montpetit Hall 201

A Maximum Entropy Approach to Recovering Information from Ranking Data

Philip L.H. Yu, Hong Kong University

In this paper, we investigate a maximum entropy (ME) approach to recover the conditional choice probabilities and hence recover the ranking probabilities. The beauty of this approach is that it is nonparametric in nature in the sense that no distributional assumption on the data is required. Moreover, we show that the ranking probabilities recovered by using the ME method satisfy the well known Luce model for ranking data. In particular, the ME estimates of the choice probabilities are the same as the corresponding maximum likelihood estimates under the Luce model. In addition, we propose a generalized ME (GME) approach which requires weaker moment constraints used in entropy maximization. Based on a simulation study, we find that the GME method produces estimates of smaller MSE and smaller bias as compared with the ME method. The proposed methods are applied to analyse the data given in Dansie (1986) in which 800 people were asked to indicate their preference on 4 motor cars.

Wednesday, June 7th, 9:15, Montpetit Hall 201

A Weighted Combination of Nonparametric Statistics for the Two-Sample Bivariate Problem

M. Mountassir, University of Ottawa & Y. Lepage, University of Montréal

A nonparametric rank test for the location and scale bivariate problem is proposed for symmetrically elliptic distribution. The asymptotic distribution of the statistic is derived under the null hypothesis and under a class of contiguous alternatives. The asymptotic relative efficiency of this test is computed and a simulation study is performed.

Session 36: MCMC session with Peter Green

Wednesday, June 7th, 8:30, Morisset Hall 218

Bayesian Analysis of Heterogeneity using Mixtures and Related Models

Peter Green, University of Bristol

The problem of finite mixture analysis dates back many years. The frequentist methodology presents many difficulties, but only recently has a really satisfactory Bayesian analysis become available, enabled, of course, by the use of novel MCMC methods. This talk will focus mainly on adaptations of the basic mixture model to various situations of more structured data, where the interest is in identifying and quantifying heterogeneity, and will include an approach to Bayesian ANOVA for factorial experiments, and a new analysis of geographical epidemiology data.

Session 37: Risk Estimation and Applications

Wednesday, June 7th, 8:45, Montpetit Hall 202

Risk Adjustment with Binary Outcomes when Covariate Information is Incomplete

P.D. Faris, R.F. Brant & W. Ghali, University of Calgary

In clinical outcomes research, risk adjusted measures such as the O/E ratio are used to assess the performance of treatment providers. Missing data was common in the initial year of an ongoing cardiovascular research initiative conducted in Alberta (the APPROACH initiative). Consequently, missing data methods were applied to the logistic regressions used to construct risk adjusted measures. The resulting measures were superior to measures based only on complete cases. Computer simulations verified that with modest confounding among covariates, risk adjusted measures based on missing data methods work well under a variety of missing data conditions.

Wednesday, June 7th, 9:00, Montpetit Hall 202

Migration of Elephant Seals

Jason Loeppky, Simon Fraser University

Elephant seals migrate over vast areas of ocean starting from southern California, to a distant point in the north Pacific where they forage for food. The seals travel a similar path every year. It is interesting to note that this path is free of visible land markings, indicating the seals may have some form of a global positioning system. Mathematically it would be interesting to know if the seals also travel the shortest path between the two points. By studying the motion of the seals as a sequence of orthogonal rotation matrices on the unit sphere, a least squares fit of the seals motion will be found.

Wednesday, June 7th, 9:15, Montpetit Hall 202

Carryover Effects in Bioequivalence (BE) Studies

Giuseppina D'Angelo, Diane Potvin, Jacques Turgeon, Phoenix International Life Sciences

The observed frequency of carryover effects and how they impact BE results were examined. A total of 96 6-sequence, 3-period, 3-treatment fed-fasted BE studies were analyzed for carryover effects and 324 2-sequence, 2-period, 2-treatment fasted BE studies were analyzed indirectly for carryover effects via sequence effects. Cmax and AUC0-t were log-transformed and modeled as responses in a standard ANOVA. Carryover effects were observed ($p < 0.05$) in less than 5% of the 96 studies analyzed and sequence effects ($p < 0.1$) in approximately 10% of the 324 studies analyzed. These correspond to the margins of error that are generally accepted. We thus propose that first-order carryover effects be removed from the model when analyzing BE studies.

Wednesday, June 7th, 9:30, Montpetit Hall 202

Estimating Baseline Distribution in Proportional Hazards Cure Models

Yingwei Peng, Memorial University of Newfoundland

Semiparametric proportional hazards mixture models have been proposed recently for survival data from some clinical trials where cure is a possibility. However, it is known that the estimated baseline distribution in the proposed methods will not be proper if there is no restriction on the tail of the distribution. In this work, several methods to estimate the baseline distribution are proposed, and compared with existing methods with respect to bias and mean square error. They are also examined in detail for a real data set from a clinical trial in breast cancer.

Wednesday, June 7th, 9:45, Montpetit Hall 202

Nonparametric Regression Analyses to Identify Prognostic Factors for Women with Early Breast Cancer

Jeff Bakal & Dongsheng Tu, Queen's University

Cox's proportional hazard regression model has been used extensively in medical studies to identify potential prognostic factors for diseases. The proportional hazard assumption behind this model, however, may not be valid for some survival data. A data set from a recently published phase III clinical trial conducted by NCIC Clinical Trials Group and comparing two chemotherapy regimens for women with early breast cancer provides us an interesting example. In this presentation, some nonparametric regression methods are developed or applied to analyse this dataset. The results of the nonparametric regression analyses are contrasted with that obtained from the Cox's regression model.

Session 38: Methods for Teaching Statistics

Wednesday, June 7th, 8:30, Morisset Hall 205

Methods for Teaching Statistics

Marc Duchesne, Via Systems

Teaching statistics to a group of people will always represent a challenge. However, for all students - whether in statistics or not - there is a very personal way of learning: approaches related to theory, visual perception, practical experience or illustrated by examples. The traditional method of lecturing does not enable us to reach all these learning types. This session will show various techniques aiming at a better "reception" of statistical concepts. The following elements will be covered: probability laws, normal distribution, hypothesis testing and design of experiments.

Session 39: Complex designs

Wednesday, June 7th, 10:30, Montpetit Hall 202

Undoing Complex Survey Data Structures: Some Theory of Inverse Sampling

Jon Rao, Carleton University & A.J. Scott, University of Auckland, New Zealand

Application of classical statistical methods to data from complex sample surveys without making allowance for the survey design features can lead to erroneous inferences. Methods have been developed that account for the survey design, but these methods require additional information such as survey weights, design effects or cluster identification for micro data. Inverse sampling (Hinkins et al.,1997) provides an alternative approach by undoing the complex survey data structures so that standard methods can be applied. Repeated subsamples with simple random sampling structure are drawn and each subsample is analysed by standard methods and combined to increase the efficiency. This method has the potential to preserve confidentiality of micro data, although computer-intensive. We present some theory of inverse sampling and explore the limitations and applications of the method.

Wednesday, June 7th, 11:00, Montpetit Hall 202

Composite Estimation for Complex Survey Data

John L. Eltinge, U.S. Bureau of Labor Statistics

This paper discusses composite estimation methods applicable to data obtained through a complex sample design, with special emphasis on the operating characteristics of the resulting point estimators and inference methods. Following standard approaches, we use a general linear model framework to develop composite estimators, and to evaluate some of their properties. We place special emphasis on decomposition of mean squared error into components attributable to sampling errors; nonsampling errors; errors in model approximations; and errors in model parameter estimators. We close with a discussion of specific composite estimators that combine information across time and across cross sectional units, respectively.

Wednesday, June 7th, 11:30, Montpetit Hall 202

The Generalised Exponential Model for Sampling Weight Calibration

Ralph Folsom & A.C. Singh, Research Triangle Institute

In survey practice, design weights are typically adjusted for extreme values via trimming or winsorization, for nonresponse by weighting class or inverse response propensity adjustments, and for post-stratification via raking-ratio adjustments. In this paper we present a generalised weight calibration model that is flexible enough to handle all three of these adjustment functions. This model, referred to as the generalised exponential model (GEM), is a derivative of Deville and Särndal's (1992) logit calibration model, Folsom and Witt's (1994) scaled constrained exponential model, and the deficiency raking approach of Singh, Wu and Boyer (1995). As is the case for Folsom and Witt's model as well as the deficiency-raking approach, the GEM can have a lower bound of 1 when it is used to model an inverse response propensity. Note that Deville and Särndal's logit calibration model has lower bound less than one. The important innovation introduced in the GEM model is its ability to incorporate unit-k specific upper and lower bounds (L_k, U_k) for a subset of the units. An important application of this feature is to identify an initial set of units with extreme weights and to use unit specific bounds to exercise control over their final adjusted weights preventing them from becoming too extreme after adjustment. If there is a high frequency of extreme values among the initial sample design weights, it is better to have a separate outlier treatment (OT) step. In this (OT) application of the GEM the initial design weights are first winsorized, which means that the weights identified as extreme are set to prescribed upper and lower boundaries like the stratum median $\pm$ 3 times the interquartile range. Now, the winsorized weights are held fixed at the boundaries, and the remaining weights are calibrated so that when added to the winsorized weight sums they reproduce the original weighted counts for various demographic groups. After specifying the GEM and illustrating how it is set up for the (OT), (NR), and post-stratification (POST) applications, we illustrate its use for these three weight adjustment tasks on the 1998 National Household Survey on Drug Abuse (NHSDA). In this application we observe that when an excessive number of control variables is employed in the GEM adjustment a notable increase in the CV of

the adjusted weights will typically occur indicating a possible loss in precision. A more direct measure of this precision loss (or gain) is obtained by computing for selected study variables the proper sandwich or linearisation variance (which accounts for variability in the GEM parameter estimates). Reference is made to Vaish and Singh's (1999) application of these sandwich variance calculations for the NHSDA.

Session 40: Graduate Student Presentations

Wednesday, June 7th, 10:30, Montpetit Hall 207

Equivariant Estimation of Multivariate Parameters from a Mixture of Normal Distributions with Certain Constraints

N'deye Rokhaya Gueye, Université de Montréal

Let X be a random vector whose distribution is given by a mixture of normal distributions, i.e. there exists a random variable Z such that the conditional distribution of X given Z is multinormal $N\left(\mu, Z^{-1}\Omega\right)$. Let $\Sigma = E^{-1}(Z)\,\Omega$. We deal with the problem of estimating the mean μ when the parameters satisfy the following constraints: (a) either $\mu^T\Sigma^{-1}\mu = \lambda^2$, with λ^2 known, or (b) $\Sigma = \mu^T\mu/c^2$, where c^2 is the known percentage standard deviation. We are interested in estimating the mean by a mixture of multinormal distributions. In this article, we develop the general form of the best equivariant rule and derive the risk function. We also study the class of James-Stein estimators, the class of linear estimators and the MLE for the case when the parameters satisfy (a). This study demonstrates that the optimal estimator in the class of James-Stein estimators is independent of λ^2. To study robustness of the best invariant estimator, we compare its risk with those of the estimators previously mentioned when the parametric information is wrong.

Wednesday, June 7th, 10:45, Montpetit Hall 207

Combining Information by using Relevance Weighted Likelihood Method

Steven X. Wang, University of British Columbia

The relevance weighted likelihood method was introduced by Hu and Zidek to formally embrace a variety of statistical procedures for trading bias for precision. Their approach combines all relevant information through a weighted version of the likelihood function. Their method will be described in this presentation. Moreover we propose an asymptotic paradigm different from theirs. The asymptotic properties of the estimator derived from our paradigm and some examples will be presented. These results complement those of Hu and Zidek.

Wednesday, June 7th, 11:00, Montpetit Hall 207

On Estimation of the Four-Parameter Kappa Distribution

Connie Winchester, Dalhousie University

The distribution of extremes are most often modeled by the generalized extreme-value distribution (GEVD) since theory has shown this distribution to be the limiting form of extreme value distributions. In reality, results are not always satisfactory when fitting the GEVD to finite samples. The four-parameter kappa distribution (KAPD), a generalization of the GEVD, is useful when the GEVD is not adequate. The current method of estimation of the KAPD (L-moment estimation) is disadvantageous in several respects. As an alternative, maximum likelihood estimation is proposed. A simulation study as well as an analysis of real life extreme wind data show these estimates to be beneficial.

Wednesday, June 7th, 11:15, Montpetit Hall 207

Powerful Goodness-of-Fit Tests

Jin Zhang, York University

Let X be a continuous random variable with distribution function $F(x)$, and $X_1, X_2, ..., X_n$ be a random sample from X. A new approach is established to construct statistics for testing the null hypothesis $H_0 : F(x) = F_0(x)$ for all x against the alternative $H_1 : F(x) \neq F_0(x)$ for some x, where $F_0(x)$ is a completely specified distribution function. The new method cannot only generate traditional goodness-of-fit test statistics (including Kolmogorov-Smirnov statistic, Cramér-von Mises statistic and Anderson-Darling statistic), but also produce a new type of statistics, which are much more powerful and are significant improvements or innovations on the old ones.

Session 41: Probability and Markov Chains

Wednesday, June 7th, 10:30, Morisset Hall 218

Bounds on Loglikelihood Derivatives in Hidden Markov Models and Various Applications

Peter Bickel, Ya'akov Ritov, & Tobias Ryden, University of California, Berkeley

Following up work in Bickel and Ritov (1993) and Bickel, Ritov and Ryden (1998), we discuss how to derive bounds on all derivatives of the log likelihood of Hidden Markov models which have the correct asymptotic structure under simple checkable conditions on the basic model. Using these bounds we show that standard higher order properties of the MLE hold under the usual conditions and that the parametric bootstrap is trustworthy. We also, using these bounds, obtain local expansions in the parameters of the model for Fisher information and Shannon entropy, generalizing work of Baum and Petrie (1966) and Petrie (1969). We finally indicate how useful qualitative conclusions may be drawn from these expansions.

Wednesday, June 7th, 11:00, Morisset Hall 218

Markov Chains and $1/f$ Noise

Priscilla Greenwood, University of British Columbia

In a great variety of contexts, Markov time series have a power spectrum of the form $1/f^a$ over a substantial range of f, where a is near 1. This phenomenon is called "$1/f$ noise". Examples include fluctuations in neurophysiological and psychophysical data. In this talk we explore the question: Under what model conditions does a Markov chain produce a $1/f$-type power spectrum?

Wednesday, June 7th, 11:30, Morisset Hall 218

Drift Conditions and Invariant Measures for Markov Chains without Irreducibility or Continuity Conditions

Richard Tweedie, University of Minnesota

We show that the classical Foster-Lyapunov conditions for the existence of a invariant measure for a Markov chain are still valid without any continuity or irreducibility assumptions, provided a weak uniform countable additivity condition is satisfied. The result relies on some deep structural theorems for Markov chains but the condition is much easier to check in many cases than are weak Feller or T-chain conditions, or irreducibility conditions. We also show that convergence to a finite mixture of invariant measures holds in this case, and give conditions under which the invariant measure is unique. Examples from time series are also given.

Session 42: Large Multi-Centre Clinical Trials

Wednesday, June 7th, 10:30, Montpetit Hall 203

Data Management Challenges for Large Multicenter Trials

Wayne Taylor, Clinical DataFax Systems

Today's "large simple trials" enroll thousands of patients from hundreds of clinical sites. Even more moderately sized trials present a considerable challenge to the trial coordinating center. Trial management staff must collect, review and clean hundreds of data items on each patient, ensure that patients are not lost to follow-up and that assessments are completed on time, monitor adherence to protocol by patients and clinical investigators, and provide timely and accurate study monitoring reports and interim statistical analyses. Failure in any of these tasks can mean the difference between ground breaking and uninterpretable clinical trial results. Further, if the data are to be submitted to the FDA and other regulatory agencies, trialists must adhere to federal regulations and guidances which dictate requirements for data accuracy, auditability and accountability. And recent regulations specify requirements for data management software. This presentation will review the authors experience with 3 different approaches to clinical trial data management: traditional collection and processing of paper case report forms (CRFs) (EC/IC trial 1977-1985), direct data entry at the clinical sites (NASCET 1989-1999), and the DataFax system in which CRFs are faxed to a central computer (ACE 1994-1998, and HOPE 1993-1999). In each case we will consider problems and solutions, costs and benefits and lessons learned. Thoughts will also be offered on the challenges that will be faced in attempts to implement internet based data management systems for clinical trials, and efforts required to meet the new FDA regulations.

Wednesday, June 7th, 11:00, Montpetit Hall 203

The CAPRIE Trial - A Case Study of an Academic/Pharmaceutical Industry Collaboration

Robin S. Roberts, McMaster University

The Clinical Trials Methodology Group (CTMG) at McMaster University is a well established research group with extensive experience in conducting randomized trials in collaboration with industry. This presentation will describe the anatomy, physiology, and sociology of collaborative trials. In other words, how large trials are organized, how scientific and management roles are determined, and how academic and commercial needs are balanced. The primary backdrop for the discussion of these issues will be CAPRIE, a large international study of the relative efficacy and safety of the investigational antiplatelet drug clopidogrel (compared to aspirin) in preventing major vascular events such as stroke and heart attack. An effective collaborative arrangement which offers advantages to both industry and academia is an achievable goal. To sustain the collaboration over the life of the project (and beyond) requires a certain amount of tenacity and resilience.

Wednesday, June 7th, 11:30, Montpetit Hall 203

Organization of Large Multi-centre Trials from the Perspective of a Pharmaceutical Company Statistician

David Henry, Bristol-Myers Squibb Company

Large multi-centre trials pose statistical and regulatory issues that are not seen in smaller trials. These issues will be discussed from the viewpoint of a pharmaceutical company statistician. Examples will be presented from the 15,000-subject InTIME-II trial for the treatment of acute myocardial infarction.

Session 43: The Bootstrap, Wavelets, Nonlinear Regression, and Dynamical Systems

Wednesday, June 7th, 10:30, Montpetit Hall 201

On Development of the Weighted Bootstrap

Mahmoud Zarepour, Université d'Ottawa

It is well known that the Efron bootstrap fails for the mean of heavy tail distributions. We demonstrate that this same type of failure also occurs in many types of weighted bootstraps, including the Bayesian bootstrap. Only through a careful selection for the weights can we produce a consistent weighted bootstrap procedure

that will work for the mean and other useful statistics. This is a joint work with H. Ishwaran and L.F. James.

Wednesday, June 7th, 10:45, Montpetit Hall 201

Balanced Bootstrap Samples: A New Method for Unbalanced Data

Patrick J. Farrell, Acadia University & T.W.F. Stroud, Queen's University

Variance components in factorial designs with balanced data are commonly estimated by equating mean squares to expected mean squares. For unbalanced data, the usual extensions of this approach are the Henderson methods, which require formulas that are rather involved. Alternatively, maximum likelihood estimation based on normality has been proposed. Although the algorithm for maximum likelihood is computationally complex, programs exist in some statistical packages. Nevertheless, an alternative to maximum likelihood is necessary since some investigators will avoid using these programs. We present a simpler method, that of creating a balanced data set by resampling from the original one. Using the two-way case as an example, comparisons are drawn between the performance of variance component estimators under the proposed approach and the analogous maximum likelihood and Henderson Method I estimators.

Wednesday, June 7th, 11:00, Montpetit Hall 201

Examining Pivotality using Bootstrap Recycling

Angelo J. Canty, Concordia University

The construction of accurate bootstrap confidence intervals relies on the concept of pivotality. When an exact pivot exists, exact confidence intervals are possible. In the nonparametric setting such exact pivots are rare but it is still of interest to know if a statistic is close to pivotal. For an estimator T of θ, we may want to examine how the distribution of $T - \theta$ varies with θ. Using the bootstrap requires examination of the distribution of $T^* - t$ as a function of t. Double bootstrapping can be used to do this but is very computationally intensive. In this talk I will describe a more efficient method based on the concept of bootstrap recycling and give examples of its use.

Wednesday, June 7th, 11:15, Montpetit Hall 201

Wavelet Methods and Designs for Curve Estimation

Alwell J. Oyet & Li Mei Sun, Memorial University of Newfoundland

We discuss wavelet versions of nonparametric methods of curve estimation and the method of weighted least squares (WLS). A modified version of the Gasser-Müller (GM) estimator is introduced. In terms of the integrated squared error (ISE), the modified version is shown to perform better than the usual GM estimator. A number of illustrative examples involving the Haar, Daubechies and Multiwavelet systems are provided. The simplicity of the weighted least squares approach when compared to the GM estimator is highlighted. In addition, the ISE is used to show that the WLS estimate performs better than the GM estimator. In what follows, we discuss integer-valued designs for estimating nonparametric curves based on the modified GM estimator and the simulated annealing algorithm. Some examples will be discussed.

Wednesday, June 7th, 11:30, Montpetit Hall 201

Efficiency of Aggregate Data in Non-linear Regression

J. Huh & K.C. Carri, University of Alberta

This work concerns estimating a regression function, which is not linear, using aggregate data. In much of the empirical research, data are aggregated for various reasons before statistical analysis. In a traditional parametric approach, a linear estimation of the non-linear function with aggregate data can result in unstable estimators of the parameters. More serious consequence is the bias in the estimation of the non linear

function. The approach we employ is the kernel regression smoothing. We describe the conditions when the aggregate data can be used to estimate the regression function efficiently. Numerical examples will illustrate our findings.

Wednesday, June 7th, 11:45, Montpetit Hall 201

Lyapunov Exponent Estimation for One-Dimensional Chaotic Dynamical Systems

Salim Lardjane, ENSAI, Campus de Ker Lann, France, lardjane@ensai.fr

It has been established that some 1-dimensional dynamical systems have trajectories which look perfectly random. Consider, for example, the system induced by the logistic map on the unit interval. This phenomenon can be quantified by the Lyapunov exponent, which measures the rate at which small perturbations are amplified by a dynamical system. The greater the exponent, the more chaotic the system. Our concern is with the estimation of the Lyapunov exponent of an unknown dynamical system from the observation of a finite number of successive states of the system. We propose a new estimator of the Lyapunov exponent, prove its consistency and illustrate its good computational properties by some simulations.

Session 44: Estimation, biostatistics and time series

Wednesday, June 7th, 10:30, Morisset Hall 205

Is There Anything Left to Say about Pearson's Correlation Coefficient?

Christian Genest, Université Laval

Despite not being recent, Pearson's concept of correlation has not been explored in a Bayesian framework. Nevertheless, this approach offers interesting implications for statistical inference. As we will demonstrate in this presentation, some of the most elementary properties of the classical linear correlation coefficient can sometimes have unexpected consequences.

Wednesday, June 7th, 10:45, Morisset Hall 205

Density Estimation on the Half-line

Serge B. Provost, University of Western Ontario, serge@uwo.ca

Various approaches for approximating densities whose support is the half-line will be discussed. Several parametric families of distributions such as the generalized gamma, the transformed beta and the generalized inverse Gaussian may be used to model unimodal distributions defined on the set of positive real numbers. The proposed method, which is based solely on the moments of the distribution to be estimated, also applies to multimodal distributions. First, the distribution is mapped onto a bounded interval wherein a least-squares polynomial approximation of the density function is obtained by making use of Legendre polynomials. An approximation to the density function whose support is the half-line is then obtained via the inverse transformation. Several examples will illustrate this approach and some applications will be proposed.

Wednesday, June 7th, 11:00, Morisset Hall 205

Interval Censoring in Survival Analysis - Application to an HIV Study

Meriem Saïd, Université Laval

Right-hand side censoring is a well-known phenomenon in survival analysis. There exists an important amount of statistical procedures that apply to this case. Interval censoring applies to follow-up analyses, in which each subject is examined at regular time intervals. The survival time (without right-hand side censoring) of a subject is then comprised between two times of examination. We will show an example

related to the study of HIV in which some statistical procedures allow to manipulate such data, in particular to estimate and compare survival functions.

Wednesday, June 7th, 11:15, Morisset Hall 205

A Poisson Approximation for Heterogeneous Epidemic Processes

M'hamed Mesfioui, Université Laval

This presentation will discuss the study of generalized epidemic processes within heterogeneous populations. We will first present L-I-E-type (likely-infected-eliminated) epidemic models, and then introduce another model based on a discrete Markov chain. We will also establish a Poisson approximation for the final size of the epidemic for the case with many infected types.

Wednesday, June 7th, 11:30, Morisset Hall 205

Weighted Local Adjustments of Polynomials to a Cumulative Distribution Function and its Partial Derivatives

Elmiloud Bensaïd, Université Laval

In this presentation, we will describe an approach to the problem of estimating a cumulative distribution function and its partial derivatives. The method developed is based on the theory of multivariate polynomials. We will also demonstrate the range of applications of the technique with a series of examples. The performance of those estimators will be verified using a theoretical study and numerical simulations.

Wednesday, June 7th, 11:45, Morisset Hall 205

Applications of a Serial Version of Kendall's Process

Christian Genest & Jean-François Quessy, Université Laval, & Bruno RÉMillard, Université du Québec à Trois-Rivières

Ferguson, Genest and Hallin (Canadian Statistical Review, 2000) demonstrated how Kendall's tau can be adapted to test for serial dependency in a univariate time series. Similarly, it is possible to define a serial generalization of Kendall's process whose asymptotic properties were studied by Barbe, Genest, Ghoudi and Rémillard (Journal of Multivariate Analysis, 1996). Using simulations, we will show that a characterization of the process' limit allows for the construction of test statistics which are often more powerful than the correlogram and the classical nonparametric measures for serial dependency.

Session 46: Applied Survey Methods — II

Wednesday, June 7th, 13:30, Montpetit Hall 201

Sampling for the Unified Enterprise Surveys (UES)

M.-N. Parent & M. Simard, Statistique Canada

The new Unified Enterprise Survey (UES) integrates many Statistics Canada's annual surveys. Some survey concepts and methods applied in the different sampling cycles remain sensibly the same from year to year. Those include the use of a single frame, a two-phase approach, the stratification algorithm, the use of permanent random numbers, the sample allocation and selection. Furthermore, a new method of determining coefficient of variation for multi-purpose surveys has been developed. Finally, strategies as to meet different industry specifications in term of precision have been prepared.

Wednesday, June 7th, 13:45, Montpetit Hall 201

Estimation for the Unified Enterprise Surveys (UES).

Caroline Cauchon, Alain Côté & Michelle Simard

The new Unified Enterprise Survey (UES) integrates many Statistics Canada's annual surveys. An estimation system with general applications and based on modules has been developed. The system main objective is to obtain final weights for each selected record. Among the possible steps, there are post-stratification technique, outlier detection and treatment, as well as modelling between survey and administrative data. Examples using reference year 1997 data presenting the impact on the estimates and on the variance based on the different outlier treatment and estimation strategies have been prepared.

Wednesday, June 7th, 14:00, Montpetit Hall 201

The Annual Wholesale Trade Survey: the Transition to the Unified Enterprise Survey

Jean-Sébastien Provençal & Hélène Bérard, Statistics Canada,
jean-sebastien_provencal@a.statcan.ca, helene.berard@a.statcan

Since 1993, the Annual Wholesale Trade Survey (AWTS) collects financial data on the economic activity in the wholesale trade. In 1998, the survey has been redesigned and integrated to the Unified Enterprise Survey (UES). The new AWTS collects information on origins and destinations of goods sold for different commodities as well as financial data. The arrival of an existing survey, such as the AWTS, within UES introduces new challenges. For example, some aspects of the former design were taken into consideration when selecting the sample and defining the precision required for the estimates. The introduction of the commodity component brought some major changes in the edit and imputation processes and will also necessitate the elaboration of a new estimation strategy. Indeed, UES uses survey data and administrative data to produce the final estimates. However, the administrative data sources only provide financial data and do not contain any information related to commodities. These aspects will be described in more detail during the presentation.

Wednesday, June 7th, 14:15, Montpetit Hall 201

Creation of the Dual Frame for the New Canadian Community Health Survey

Yves Béland, Suzanne Lessard & Marianna Morano, Statistics Canada

The Canadian Community Health Survey (CCHS) is a new health survey currently under development. In its first year, a sample of 130,000 respondents will be selected from a dual frame: an area frame (for face-to-face interviews) and a random digit dialing (RDD) frame (for telephone interviews). This paper will describe the process involved in the creation of the dual frame. This will include: the adaptation of the existing area frame (LFS structure) to the CCHS needs, the stratification of the RDD frame, the integration of the two, and challenges encountered with sample selection.

Session 47: Correlated Data

Wednesday, June 7th, 13:30, Montpetit Hall 207

Analysis of Multivariate Failure Time Data

Wenqing He, University of Waterloo

Multivariate failure time data arise frequently in scientific investigations, for example when there exists natural or artificial clustering of study subjects such that failure times within the same cluster are correlated. Methods for estimating the cumulative hazard functions, the associations among the failure times and covariate effects are presented based on Clayton models which incorporate piecewise-constant baseline hazard functions or $M(I)$-spline baseline hazard functions. The methods can accommodate the situation where the right censoring is dependent with the failure times. Simulation study is used to examine the effectiveness of the methods. An illustration with colon cancer data is provided.

Wednesday, June 7th, 13:45, Montpetit Hall 207

An Analysis of Familial Correlation in Expression of Neurofibromatosis

Yinshan Zhao, University of British Columbia

Neurofibromatosis Type 2 (NF2) is an autosomal dominant disorder characterized by development of multiple central and peripheral nervous system tumors. To quantify familial similarities will help with determining if the expression of NF2 is entirely attributable to the NF2 mutant allele. Two quantitative traits in NF2 are considered: (1) age at onset of hearing loss. Since right censoring is present, multivariate survival analysis using random effect model is considered. Point estimate and asymptotic confidence interval of intra-familial correlation are obtained from MLE. (2) number of cranial meningiomas. A Poisson Gamma mixture model is considered. In order to rule out the age effect, an age adjustment might also be needed in this case. MLE is used to estimate the parameters.

Wednesday, June 7th, 14:00, Montpetit Hall 207

Likelihood in QTL Analysis under a Mixed Model of Inheritance in Non-Experimental Populations

J.C. Loredo-osti, McGill University & Montreal General Hospital Research Institute, & B.R. Smith, Dalhousie University

This work addresses the problem of assessing segregation of a quantitative trait under a mixed model of inheritance in non-experimental populations, when the trait is linked to some monitored marker. One approach to the problem is through the unconditional trait-marker cosegregation analysis. Under mixed models of inheritance, in general, techniques for evaluating the "exact" likelihood are only feasible for nuclear families. Nevertheless, the nuclear family is an important case, because in many situations the sibships can be taken as the sampling unit in a Monte Carlo scheme. Here, it will be presented an extension to one of the techniques (the Morton-MacLean integral) to compute the "exact" likelihood whenever only a sibship at the bottom of the pedigree and, possibly, their parents, have quantitative trait records available, regardless the complexity of such pedigree. This result is used for proposing a Monte Carlo sampling scheme in blocks of sibships for the case of related families with records at the bottom of the pedigree.

Wednesday, June 7th, 14:15, Montpetit Hall 207

On Mixed Model Selection

J. Sunil Rao and Jiming Jiang, Case Western Reserve University

We consider the problem of selecting effects in a linear mixed effects model. We explore the selection of fixed effects from a set of candidate predictors when the random effects are assumed known, and also when they are not assumed to be known. We derive a form of the generalized information criterion (GIC) for these problems and prove consistency of the selection procedures. A series of simulations and tests on real datasets provide some insight as to the empirical performance of these procedures.

Wednesday, June 7th, 14:30, Montpetit Hall 207

An Optimal Linear Estimating Function For Latent Variable Models

Renjun Ma, University of Ottawa

Generalized linear models with latent variables such as stat-space, linear growth curve and random effects models are of wide applicability and practical importance. Due to the difficulty of evaluating likelihood, estimating functions are widely adopted to draw inference on latent variable models. Much of the effort in this area has concentrated on finding computationally feasible fitting schemes, whereas the optimality has largely been ignored. We study an optimal linear estimating function for latent variable models. The application of this estimating function to stat-space, linear growth curve and random effects models will also be illustrated.

Wednesday, June 7th, 14:45, Montpetit Hall 207

Distribution-Free Inferences about Latent Disease Progression and Survival

Mei-Ling Ting Lee, Harvard University & G.A. Whitmore, McGill University

In many medical contexts, diseases progress in subjects until a medical failure event (such as death) is triggered when the disease level first reaches a failure threshold. Time to the event is then described by a first passage or first hitting time. In this paper, we assume that the disease process is latent or unobservable. Hence, the process parameters must be inferred from survival data alone. The paper presents a model for the disease process that requires few distributional assumptions and, hence, is quite general in its potential application. Distribution-free methods for parametric and predictive inference are developed, as well as techniques for incorporating covariates. Computational aspects of the methods are examined. Case examples are presented that demonstrate the methodology and its practical use. The developments reported in the paper will provide medical scientists and clinicians with new and robust statistical tools for assessing treatment effects, a subject's current disease state and a prognosis of a subject's future survival prospects. The methodology also should help to deepen scientists' insights into the causes and nature of disease progression in many practical settings.

Session 48: Nonparametric Bayesian Inference

Wednesday, June 7th, 13:30, Montpetit Hall 203

Approximate Dirichlet Process Computing for Finite Normal Mixtures

Hemant Ishwaran, Cleveland Clinic Foundation

A rich nonparametric analysis of the finite normal mixture model is obtained by working with a precise truncation approximation of the Dirichlet process. Model fitting is carried out by a simple Gibbs sampling algorithm which mixes well, requires no tuning parameters, and directly samples the nonparametric posterior. Working directly with the nonparametric prior is conceptually appealing and among other things leads to a method for estimating the posterior distribution function for the mixing distribution.

Wednesday, June 7th, 14:00, Montpetit Hall 203

Analysis of Densities

George Tomlinson, University of Toronto & Mike Escobar, Harvard University

Consider a repeated measures model where the within subject distribution is very irregular. In this talk, a method is proposed which allows one to model in a general fashion the within subject and between subject distributions. First, the underlying densities from which the observations on individual subjects are drawn are estimated using Bayesian nonparametric methods. Next, Bayesian nonparametric methods allow one to consider these densities functions as samples from a distribution on the space of density functions. Methods of inference, such as the detection of outliner subjects, will also be discussed.

Wednesday, June 7th, 14:30, Montpetit Hall 203

Properties and Approximations of Dirichlet Processes

Jayaram Sethuraman, Florida State University

Dirichlet process priors were defined by Ferguson, Blackwell and MacQueen. A constructive definition for these processes was given by Sethuraman. In this talk, we will briefly review these and later results. We will show that any two Dirichlet processes are either absolutely continuous with respect to one another or singular with respect to one another. For computational purposes, approximations to Dirichlet priors and the resulting posteriors are useful. We will present different methods of approximations, and clarify the nature of the approximations.

Session 49: Probability, Design and Mathematical Statistics

Wednesday, June 7th, 13:30, Morisset Hall 218

The Secretary Problem with Correlated Rank and Order

A.W. Kemp, University of St. Andrews, Scotland

The paper gives the solution of Reeves and Flack's (1988) generalization of the secretary problem (with correlated rank and order) as a q-analogue of Lindley's (1961) solution for the classical uncorrelated case. When there is positive correlation, the length k of the learning run of applicants that gives the maximum probability, $P(k)$, of selecting the best applicant is somewhat less than that for no correlation; when there is negative correlation it is usually much greater. For the same number N of applicants and optimal values of k, $P(k)$ increases with the amount of correlation (both positive and negative).

Wednesday, June 7th, 13:45, Morisset Hall 218

Design Issues in a Blocked Fractional Factorial Split-Split-Plot Experiment – A Case Study from the Aerospace Sector

John F. Brewster & Robert G. McLeod, University of Manitoba

In industrial experiments fractional factorial designs are widely used for screening purposes. If some of the factors are hard-to-vary and others are easy-to-vary, restrictions on randomization may lead to the use of fractional factorial split-plot (FFSP) designs. Bingham and Sitter (1999) and others have considered the choice of optimal FFSP designs, using the minimum aberration criteria. Sitter, Chen and Feder (1997) have also considered the choice of optimal blocked fractional factorial (BFF) designs. In this talk, a plating experiment from the aerospace sector will be described, in which a blocked fractional factorial split-split plot (BFFSSP) design was used. It will be shown that blocking such an FFSSP (or FFSP) at the whole-plot level can introduce complications that affect the choice of design.

Wednesday, June 7th, 14:00, Morisset Hall 218

An Affine-Invariant Multivariate Sign Test for Cluster Correlated Data

Denis Larocque, Université du Québec à Trois-Rivières

A multivariate location model for cluster correlated observations is presented. An affine-invariant multivariate sign test for testing location is proposed. The test statistic is an adaptation of the one proposed by Randles (1999) and is asymptotically distributed as a chi-square random variable under the null hypothesis under very mild conditions. In particular, the test can be used for skewed populations. The values of its Pitman asymptotic efficiency relative to a test based on the overall average are obtained for a general multivariate normal model. These results show that there is an improvement in the relative performance of the new test as soon as intra-cluster correlation is present. Even for one-dimensional data, the new test (based on signs in that case) can be very competitive at the normal model. Furthermore, the statistic is easy to compute even for large dimensional data. A simulation study shows that the test performs well in comparison to the average based test. An example with a real data set is also given.

Wednesday, June 7th, 14:15, Morisset Hall 218

Assessment of Local Influence for Models with Incomplete Data

Hongtu Zhu, University of Victoria

We propose a method to assess local influence of minor perturbations of a statistical model with incomplete data. The essential idea is to utilize Cook's approach to the conditional expectation of the complete-data log-likelihood function in the EM algorithm. It is shown that the proposed method has some nice properties and it can be applied to a wide variety of complicated problems that cannot be handled by the existing

methods in local influence due to the intractable observed-data log-likelihood functions. Some illustrative artificial and real examples are presented.

Wednesday, June 7th, 14:30, Morisset Hall 218

Fitting Bivariate Extreme Value Models Robustly

D.J. Dupuis, Dalhousie University

The analysis of extreme values is often required from short series which are biasedly sampled or contain outliers. Univariate extreme value methods are poorly equipped to quantify the extreme values for these data. If related observations are available, bivariate extreme value methods may be used to analyse jointly the short series with these observations. Barao and Tawn (Applied Statistics, 1999) show that bivariate methods provide substantial benefits. In this talk, we illustrate some of these benefits and show how these models may be fitted robustly to further enhance inference.

Wednesday, June 7th, 14:45, Morisset Hall 218

Approximating the Distribution of Ratios of Quadratic Forms

Young-Ho Cheong, University of Western Ontario

A simple recursive formula for evaluating the moments of a quadratic form is obtained. Using this result in conjunction with orthogonal polynomials, approximations to the distribution of ratios of quadratic forms are derived. Various applications including the Durbin-Watson test statistic will be discussed.

Session 50: Survey Methodology

Wednesday, June 7th, 13:30, Morisset Hall 205

Exploring the Path of an Origin-Destination Survey

Mario Pimparé & Johanne Thiffault, Ministère des transports du Québec

The purpose of our session is to present all planning aspects and activities related to the realisation of the origin-destination survey which took place in the Montreal region between September and December 1998. We will examine in detail the sampling plan, the data collection technique, as well as the sample management and quality control procedures, using the software that was used for these activities. We will also show how the collected data are used in our transportation modelling environment.

Origin-destination surveys represent an essential source of information used to evaluate the population travel needs. They allow us to acquire a valid and reliable knowledge of travel behaviour. This source of information is crucial to support transportation infrastructure planning and development, as well as to guide the evaluation of transportation politics.

First, the survey data allows the updating of our transportation demand prediction model. The data also becomes the input for simulation and microsimulation on the road and public transit networks (EMME/2, MADITUC, AIMSUN2). These models, which are used to evaluate the impact on the networks of various transportation projects and politics planned by the Ministry and its partners, will be demonstrated during our session.

Index of Participants