Statistics at the Frontier of Science
June 24 to 29, 2006

MEALS

Breakfast (Continental): 7:00 – 9:00 am, 2nd floor lounge, Corbett Hall, Sunday – Thursday
*Lunch (Buffet): 11:30 am – 1:30 pm, Donald Cameron Hall, Sunday – Thursday
*Dinner (Buffet): 5:30 – 7:30 pm, Donald Cameron Hall, Saturday – Wednesday
Coffee Breaks: As per daily schedule, 2nd floor lounge, Corbett Hall
*Please remember to scan your meal card at the host/hostess station in the dining room for each lunch and dinner.

MEETING ROOMS

All lectures will be held in Max Bell 159 (Max Bell Building accessible by bridge on 2nd floor of Corbett Hall). Hours: 6 am – 12 midnight. LCD projector, overhead projectors and blackboards are available for presentations. *Please note that the meeting space designated for BIRS is the lower level of Max Bell, Rooms 155-159. Please respect that all other space has been contracted to other Banff Centre guests, including any Food and Beverage in those areas.*

SCHEDULE

Saturday June 24

16:00 — Check-in begins (Front Desk – Professional Development Centre - open 24 hours)
Lecture rooms available after 18:00 (if desired)

17:30-19:30 Buffet Dinner, Donald Cameron Hall

20:00 — Informal gathering in 2nd floor lounge, Corbett Hall (if desired)
Beverages and small assortment of snacks available on a cash honour-system basis.

Sunday June 25

7:00-8:35 Breakfast
8:45-9:00 Introduction and Welcome to BIRS by BIRS Station Manager, Max Bell 159
9:00-9:40 **A Gaussian Calculus for Inference from High Frequency Data**
Per Mykland, University of Chicago
9:40-10:20 **Heterogeneous Autoregressive Realized Volatility Model**
Yazhen Wang, University of Connecticut
10:20-10:50 Coffee Break, 2nd floor lounge, Corbett Hall
10:50-11:30 **Statistical Approaches to Option Pricing and Portfolio Management**
Jianqing Fan, Princeton University
11:30-12:10 **Efficent Likelihood Estimation in State Space Models**
Cheng-Der Fuh, Academia Sinica

12:10-13:20 Lunch
13:30-14:30 Guided Tour of The Banff Centre; meet in the 2nd floor lounge, Corbett Hall
14:30-15:25 Coffee Break, 2nd floor lounge, Corbett Hall

15:30-17:00 **Discussion - Financial and Risk Analysis**

17:30-19:30 Dinner

Monday   June 26

7:00-8:50   Breakfast
9:00-9:30   **Structural Break Detection in Time Series Models**
              Richard A. Davis, Colorado State University
9:30 –10:00 **Mixed Signal Processing for Regional and Teleseismic Arrays**
              R.H. Shumway, University of California, Davis
10:00-10:30 Coffee Break, 2nd floor lounge, Corbett Hall
10:30-11:00 **DNA Splice Site Detection with Group-penalty Methods for Categorical Predictors**
              Peter Buhlmann, ETH Zurich
11:00-11:30 **My Current Research in Time Series**
              Ian McLeod, University of Western Ontario
11:30-12:00 **Least Absolute Deviation Estimation for Fractionally Integrated Autoregressive
              Moving Average Time Series Models with Conditional Heteroscedasticity**
              Wai Keung Li, The University of Hong Kong

12:00-13:10 Lunch

13:20   **Group Photo**: Meet on the front steps of Corbett Hall at **13:20 sharp**.

13:40-14:30 **Functional Data Analysis: Where it's been and where it might be going?**
              Jim Ramsay, McGill University
14:30-15:20 **Functional Data Analysis: Tools and issues**
              Hans-Georg Mueller, University of California at Davis
15:20-16:00 Coffee Break, 2nd floor lounge, Corbett Hall
16:00-18:00 **Combined Discussion - Time Series & Functional Data Analysis**

18:00-19:30 Dinner

Tuesday   June 27

7:00-8:50   Breakfast
9:00-9:30   **Deformation Based Morphometry, Roy's Maximum Root and Recent Advances
              in Random Fields**
              Jonathan Taylor, Stanford University
9:30-10:00 **Three Statistical Imaging Problems**
              Jiayang Sun, Case Western Reserve University
10:00-10:30 Coffee Break, 2nd floor lounge, Corbett Hall
10:30-11:00 **Rice and Geometry**
              Robert J. Adler, Technion - Israel Institute of Technology
11:00-11:30 **Granger Causality on Spatial Manifolds: Applications to Neuroimaging**
              Pedro A. Valdés-Sosa, Cuban Neuroscience Center
11:30-12:00 **Large Scale Kalman Filtering Solutions to the Electrophysiological Source
              Localization Problem An MEG Case Study**
              Emery N. Brown, Massachusetts Institute of Technology

12:00-13:30 Lunch

              Free Afternoon

17:30-19:30 Dinner

Wednesday   June 28

7:00-8:10	Breakfast
8:20-8:50	**Bayesian Methods for Inferring Epistasis**
	Jun Liu, Harvard University
8:50-9:20	**Using Genetic Linkage to Inform Positional Cloning**
	Mary Sara McPeek, University of Chicago
9:20-9:50	**Protein Interaction Predictions Through Integrating High-throughput Data From Diverse Organisms**
	Hongyu Zhao, Yale University
9:50-10:20	**Stochastic Modeling and Inference in Nano-scale Biophysics**
	Samuel Kou, Harvard University
10.20-10:40	Coffee Break, 2nd floor lounge, Corbett Hall
10:40-11:10	**ChIP-chip: Data, Model, and Analysis**
	Ying Nian Wu, University of California at Los Angeles
11:10-11:40	**Model-Based Analysis of Keratin Filament Networks in Scanning Electron Microscopy Images of Cancer Cells**
	Volker Schmidt, Universität Ulm
11:40-12:10	**Semiparametric Regression Models with Staggered Entries and Progressive Multi-Stage Censoring**
	Murray D. Burke, University of Calgary

12:10-13:20 Lunch

13:30-15:00 **Discussion - Random Field and Image Analysis**
15:00-16:00 Coffee Break, 2nd floor lounge, Corbett Hall
16:00-17:30 **Discussion - Bio-medical**

17:30-19:30 Dinner

Thursday June 29

7:00-9:00   Breakfast
9:00-10:30 **Discussion for all five groups**
10:30-11:00 Coffee Break, 2nd floor lounge, Corbett Hall

11:30-13:30 Lunch

Checkout by 12 noon.

** 5-day workshops are welcome to use the BIRS facilities (2nd Floor Lounge, Max Bell Meeting Rooms, Reading Room) until 4 pm on Thursday, although participants are still required to checkout of the guest rooms by 12 noon. **

Abstracts

Financial and Risk Analysis

1. A Gaussian Calculus for Inference from High Frequency Data

Per Mykland, University of Chicago
mykland@galton.uchicago.edu

In the econometric literature of high frequency data, it is often assumed that one can carry out inference conditionally on the underlying volatility processes. In other words, conditionally Gaussian systems are

considered. This is often referred to as the assumption of ``no leverage effect''. This is often a reasonable thing to do, as generalestimators and results can often be conjectured from considering the conditionally Gaussian case. The purpose of this paper is to try to give some more structure to the things one can do with the Gaussian assumption. We shall argue in the following that there is a whole treasure chest of tools that can be brought to bear on high frequency data problems in this case. We shall in particular consider approximations involving locally constant volatility processes, and develop a general theory for this approximation. As applications of the theory, we propose an improved estimator of quarticity, an ANOVA for processes with multiple regressors, and an estimator for error bars on the Hayashi-Yoshida estimator of quadratic covariation.

2. Heterogeneous Autoregressive Realized Volatility Model

Yazhen Wang, University of Connecticut
yzwang@stat.uconn.edu

Volatilities of asset returns are pivotal for many issues in financial economics. The availability of high frequency intraday data should allow us to estimate volatility more accurately. Realized volatility is often used to estimate integrated volatility. To obtain better volatility estimation and forecast, someautoregressive structure of realized volatility is proposed in the literature. This talk will present my recent work on heterogeneous autoregressive models of realized volatility.

3. Statistical Approaches to Option Pricing and Portfolio Management

Jianqing Fan, Princeton University
jqfan@princeton.edu

Financial mathematical models provide useful tools for option pricing. These physical models give us a good first order approximation to the underlying dynamics in the financial market. Their power in option pricing can be significantly enhanced when they are combined with statistical approaches, which empirically learn and correct pricing errors through estimating the state price densities. In this talk, two new semiparametric techniques are proposed for estimating state price densities and pricing financial derivatives. Our empirical studies based on the options of SP500 index show our methods outperform the ad hoc Black-Scholes method and significantly so when the latter method has largepricing errors, based on over 100,000 tests.

The second part of the talk will focus on estimation high-dimensional covariance matrix for portfolio allocation and risk management. Motivated by the Capital Asset Pricing Model, we propose to use a factor model to reduce the dimensionality and to estimate the covariance matrix. The performance is compared with the sample covariance matrix. We demonstrate and identify the situations under which the factor approach can gain substantially the performance and the cases where the gains are only marginal. Furthermore, the impacts of the covariance matrix estimation on portfolio allocation and risk management are studied. The theoretical results are convincingly supported by a thorough simulation study.

4. Effient Likelihood Estimation in State Space Models

Cheng-Der Fuh, Academia Sinica
stcheng@stat.sinica.edu.tw

Motivated by studying asymptotic properties of the maximum likelihood estimator (MLE) in stochastic volatility (SV) models, in this talk, we investigate likelihood estimation in state space models. We first prove, under some regularity conditions, there is a consistent sequence of roots of the likelihood

equation that is asymptotically normal with the inverse of the Fisher information as its variance. With an extra assumption that the likelihood equation has a unique root for each n, then there is a consistent sequence of estimators of the unknown parameters. If, in addition, the supremum of the log likelihood function is integrable, the MLE exists and is strong consistent. Edgeworth expansion of the approximate solution of likelihood equation is also established. Several examples, including Markov switching models, ARMA models, (G)ARCH models and stochastic volatility (SV) models, are given for illustration.

Time Series and Stochastic Processes

1. Structural Break Detection in Time Series Models

Richard A. Davis, Colorado State University
rdavis@stat.colostate.edu

Much of the recent interest in time series modeling has focused on data from financial markets, from communications channels, from speech recognition and from engineering applications, where the need for non-Gaussian, non-linear, and nonstationary models is clear. With faster computation and new estimation algorithms, it is now possible make significant in-roads on modeling more complex phenomena. In this talk, we will develop estimation procedures for a class of models that can be used for analyzing a wide range of time series data that exhibit structural breaks. The novelty of the approach taken here is to combine the use of genetic algorithms with the principle of minimum description length (MDL), an idea developed by Rissanen in the 1980s, to find "optimal" models over a potentially large class of models.

This methodology will be demonstrated in a number of applications. In addition to fitting piece-wise autoregressive models, which works well even for local stationary models that are smooth, we will also consider extensions to piece-wise nonlinear models including stochastic volatility and GARCH models.

This is joint work with Thomas Lee and Gabriel Rodriguez-Yam

2. Mixed Signal Processing for Regional and Teleseismic Arrays

R.H. Shumway, University of California, Davis
rhshumway@ucdavis.edu

Successful monitoring of a proposed Comprehensive Nuclear Test-Ban Treaty (CTBT) ultimately rests on interpretations of time series that are produced on seismic and infrasound arrays as well as onauxiliary information from other sources such as satellites and radionuclide sampling. Underground events such as earthquakes and explosions generate plane waves propagating across arrays ofseismometers and proper use of this information is critical to the successful detection, location and identification of the source phenomenon.

When simultaneous events occur or when propagating noises are present at an array, mistakes can be made in locating an event as well as in reading the magnitude-related variables that are criticalfor discriminating between classes of events. We examine the performance of conventional high-resolution estimators such as MUSIC for two typical mixtures of signals and develop an alternateapproach using a combination of nonlinear stepwise regression and model selection techniques. The method yields the correct number of signals on two typical mixtures and allows deconvolution of the component signals.

This work is joint with Eugene Smart and Dean A. Clauter of the Air Force Technical Applications Center, Patrick AFB, Florida.

3. DNA Splice Site Detection with Group-penalty Methods for Categorical Predictors

Peter Buhlmann, ETH Zurich
`buhlmann@stat.math.ethz.ch`

DNA splice sites detection has been pursued, among other approaches, by non-Markovian time series models (cf. Zhao, Huang and Speed (2004)). As an alternative, we use logistic regression with (short) DNA sequence as categorical predictors. When including higher-order interaction terms, such models become very high-dimensional (e.g. 1'000 - 16'000 predictors). We propose to use the group-penalty (Yuan and Lin, 2006) and new modifications thereof for hierarchical model fitting. We present efficient algorithms which are particularly suited for high-dimensional problems, and we show that the methods are statistically consistent for sparse but high-dimensional problems where the number of predictor variables may be much larger than sample size. Despite the generality of our approach, it performs surprisingly well for the DNA splice site detection problems which we have analyzed.

4. My Current Research in Time Series

Ian McLeod, University of Western Ontario
`aim@stats.uwo.ca`

An overview of several topics that I am working on is present.

(i) recent work on statistical algorithms for time series: subset AR, faster ARMA maximum likelihood, automatic Brillinger monotonic trend test
(ii) improved portmanteau diagnostic checks for univariate and vector ARMA time series
(iii) applications of time series in bioinformatics
(iv) a new diagnostic check for lack of statistical independence applicable to a wide variety of statistical models including regression and generalized linear models

5. Least Absolute Deviation Estimation for Fractionally Integrated Autoregressive Moving Average Time Series Models with Conditional Heteroscedasticity

Wai Keung Li, The University of Hong Kong
`hrntlwk@hku.hk`

In order to model time series exhibiting the features of long memory, conditional heteroscedasticity and heavy tails, a least absolute deviation approach is considered to estimate fractionally autoregressive integrated moving average models with conditional heteroscedasticity. The time series generated by this model is short memory or long memory, stationary or nonstationary, depending on whether the fractional differencing parameter $d \in (-1/2, 0)$ or $(0, \infty)$, $(-1/2, 1/2)$ or $(1/2, \infty)$ respectively. Using a unified approach, the asymptotic properties of the least absolute deviation estimation are established. This article also derives the large sample distribution of residual autocorrelations and absolute residual autocorrelations and these results lead to two useful diagnostic tools for checking the adequacy of the ̄tted models. Some Monte Carlo experiments were conducted to examine the performance of the theoretical results in finite sample cases. As an illustration, the process of modeling the absolute return of the daily closing Dow Jones Industrial Average Index (1995-2004) is also reported.

This is joint work with Guodong Li.

Functional Data Analysis

1. Functional Data Analysis: Where it's been and where it might be going?

Jim Ramsay, McGill University
ramsay@psych.mcgill.ca

This talk will review functional data analysis in terms of where the field tends to fit in with respect to other methods looking a distributed data, what the more important accomplishments have been over the last decade, current work on modeling dynamic systems, and what the future might hold. An interesting spin-off is the idea of parameter hierarchies or cascades, an extension of the notion of structural versus nuisance parameters. The presentation will be entirely non technical, and illustrated throughout with results from applications.

2. Functional Data Analysis: Tools and issues

Hans-Georg Mueller, University of California at Davis
Mueller@wald.ucdavis.edu

In contrast to both multivariate and time series data, the domain of functional data is considered to be an interval (or its multidimensional analogue). Accordingly, smoothness and time-warping play a central role for functional data. Some of the consequences are the need for a stochastic framework for warping and for dimension reduction. The concept of functional principal components is one useful tool for dimension reduction, and it can be applied to develop methodology for functional linear (and generalized linear) models. Modeling functional data poses both theoretical and applied challenges, making this an attractive area for research. Extensions to longitudinal data, nonlinear models, and applications in the life sciences and genomics have found increasing interest. New types of applications, for example the analysis of spatial or financial data, will likely drive future methodological developments.

Random Field and Image Analysis

1. Deformation Based Morphometry, Roy's Maximum Root and Recent Advances in Random Fields

Jonathan Taylor, Stanford University
Mueller@wald.ucdavis.edu

The starting point of our talk is a study of anatomical differences between controls and patients who have suffered non-missile trauma. We use a multivariate linear model at each location in space, using Hotelling's T^2 to detect differences between cases and controls. If we include further covariates in the model, Roy's maximum root is a natural generalization of Hotelling's T^2. This leads to the Roy's maximum root random field, which includes many special types of random fields: Hotelling's T^2, T, and F, so, in effect the Roy's maximum root random field "unifies" many different random fields. We discuss the geometric interpretations of this "unified" random field theory.

2. Three Statistical Imaging Problems

Jiayang Sun, Case Western Reserve University
jiayang@sun.STAT.cwru.edu

In this talk, we first describe two neuron-imaging problems that challenge the status quo, statistical models assumed for typical neuronal data. They offer opportunities for new modeling and statistical

inference, including multiple comparisons arising from a negative binomial random field, which is generally applicable for analyzing data from over-dispersed Poisson regression models. If the time permits, we shall describe the third imaging problem that led to our research on identifying the pixels that most likely correspond to the false discoveries from a FDR procedure. This is joint work with Xiaofeng Wang, Kenneth Gustafson and Guang Yue.

3. Rice and Geometry

Robert J. Adler, Technion - Israel Institute of Technology
`robert@ieadler.technion.ac.il`

The classic Rice formula for the expected number of upcrossings of a smooth stationary Gaussian process on the real line is one of the oldest and most important results in the theory of smooth Gaussian processes. It has a multitude of applications, and has been generalised over the years to non-stationary and non-Gaussian processes, both over the real and over more complex parameter spaces, and to vector valued rather than real valued processes.

Over the last few years, a new Rice "super formula" has been developed, which incorporates effectively all the (constant variance) special cases known until now. More interesting, however, is that it shows that these formulae all have a
deep geometric interpretation, giving a version of the Kinematic Fundamental Formula of Integral and Differential Geometry for Gauss space.

This talk will discuss the "big picture" and some of the later ones in the session will discuss new results and applications.

4. Granger Causality on Spatial Manifolds: Applications to Neuroimaging

Pedro A. Valdés-Sosa, Cuban Neuroscience Center, Ciudad Habana, Cuba
`peter@cneuro.edu.cu`

The (discrete time) vector Multivariate Autoregressive (MAR) model is generalized as a stochastic process defined over a continuous spatial manifold. The underlying motivation is the study of brain connectivity via the application of Granger Causality measures to functional Neuroimages. Discretization of the spatial MAR (sMAR) leads to a densely sampled MAR for which the number of time series p is much larger than the length of the time series N. In this situation usual time series models work badly or fail. Previous approaches, reviewed here, involve the reduction of the dimensionality of the MAR, either by the selection of arbitrary regions of interest or by latent variable analysis. An example of the latter is given using a multi-linear reduction of the multichannel EEG spectrum into atoms with spatial, temporal and frequency signatures. Influence measures are applied to the temporal signatures giving an interpretation of the interaction between brain rhythms. However the approach introduced here is that of extending usual influence measures for Granger Causality to sMAR by defining "influence fields", that is the set of influence measures from one site (voxel) to the whole manifold. Estimation is made possible by imposing Bayesian priors for sparsity, smoothness, or both on the influence fields. In fact, a prior is introduced that generalizes most common priors studied to date in the literature for variable selection and penalization in regression. This prior is specified by defining penalties paired with a priori covariance matrices. Simple pairs of penalties/covariances include as particular cases the LASSO, Data Fusion and Ridge Regression. Double pairs encompass the recently introduced Elastic Net and Fussed Lasso. Quadruples of penalty/covariance combinations are also possible and used here for the first time. Estimation is carried out via the MM algorithm, a new technique that generalized the EM algorithm and allows efficient estimation even for massive time series dimensionalities. The proposed technique performs adequately for a simulated "small world" cortical network with linear dynamics, validating the use of the more complex penalties. Application of this

model to fMRI data validate previous conceptual models for the brain circuits involved in the generation of the EEG alpha rhythm.

5. Large Scale Kalman Filtering Solutions to the Electrophysiological Source Localization Problem–An MEG Case Study

Emery N. Brown, Massachusetts Institute of Technology
brown@neurostat.mgh.harvard.edu

Computational solutions to the high-dimensional Kalman Filtering problem are described in the setting of the MEG inverse problem. The overall objective of the described work is to localize and estimate dynamic brain activity from observed extraneous magnetic fields recorded at an array of sensor positions on the scalp and to do so in a manner that takes advantage of the true underlying statistical continuity in the current sources. To this end, we outline inverse mapping procedures that combine models of current dipoles with dynamic state-space estimation algorithms. While these algorithms are eminently well-suited to this class of dynamic inverse problems, they possess computational limitations that need to be addressed either by approximation or through the use of high performance computational resources. In this work we describe such a High Performance Computing (HPC) solution to the Kalman filter and demonstrate its applicability to the Magnetoencephalography (MEG) inverse problem.

Bio-Medical Research

1. Bayesian Methods for Inferring Epistasis

Jun Liu, Harvard University
jliu@stat.harvard.edu

I will discuss a Bayesian approach in detecting multi-locus interactions (Epistasis) for case-control association studies. Existing methods are either of low power or computationally infeasible when facing of a large number of markers. Using MCMC sampling techniques, the method can efficiently detect interactions among thousands of markers. I will also discuss the issue of statistical significance and how to adjust multiple comparisons in this situation (much of these are conjectures, though).

Based on joint work with Yu Zhang.

2. Using Genetic Linkage to Inform Positional Cloning

Mary Sara McPeek, University of Chicago
mcpeek@galton.uchicago.edu

The first step in mapping a gene for a trait often involves using linkage analysis to identify a region on a chromosome where a gene of interest may lie. Linkage disequilibrium mapping may sometimes be used to further refine the region. At this point, one may be able to identify one or several genes within the region. Even if only a single gene lies in the region, it may contain a large number of polymorphic sites (base pairs of DNA that vary across individuals), and a question of interest is to determine which site or combination of sites influence the trait. Ultimately, only well-designed biological studies can establish that particular variation influences susceptibility. However, we can address the question of whether a particular set of polymorphisms can fully "explain", in the statistical sense, the observed linkage to the region. We suppose that many tightly-linked SNPs have been identified and genotyped in affected

relatives in a region showing strong linkage with a binary trait. We note that if a particular set of SNPs contains all the sites in the region that influence the trait, then conditional on the genotypes at those SNPs, there should be no excess sharing in the region among affected individuals. We use this idea to develop a statistical test of the null hypothesis that a particular SNP or pair of SNPs can explain all the evidence for linkage and to develop a confidence set of individual SNPs and pairs of SNPs that has appropriate coverage probability of the causal SNP or pair of SNPs assuming that there are no more than 2 in that region. We allow for arbitrary genetic model (including epistasis with other unlinked susceptibility loci) and linkage disequilibrium and develop appropriate methods to take into account the uncertainty in haplotype frequencies. We also discuss approaches to the problem of adjusting for the selection of the region based on the linkage results in the same sample of individuals.

This is joint work with Maoxia Zheng, Lei Sun, Jian Zhang, and Nancy Cox.

3. Protein Interaction Predictions Through Integrating High-throughput Data From Diverse Organisms

Hongyu Zhao, Yale University
hz27@email.med.yale.edu

Predicting protein-protein interactions is critical for understanding cellular processes. Because protein domains represent binding modules and are responsible for the interactions between proteins, several computational approaches have been proposed to predict protein interactions at the domain level. The fact that protein domains are likely evolutionarily conserved allows us to pool information from data across multiple organisms for the inference of domain-domain and protein-protein interactions. In this talk, we present our results on estimating domain-domain interaction probabilities through integrating large-scale protein interaction data from three organisms, yeast, worms, and fruit flies. The estimated domain-domain interaction probabilities can be then used to predict protein-protein interactions in a given organism. Based on a thorough comparison of sensitivity and specificity, and other analyses, the proposed approaches have better performance due to their ability to borrow information from multiple species, and the estimated domain-domain interaction probabilities can also be informative in predicting protein-protein interaction in other organisms.

4. Stochastic Modeling and Inference in Nano-scale Biophysics

Samuel Kou, Harvard University
kou@stat.harvard.edu

Recent advances in nanotechnology allow scientists to follow a biological process on an unprecedented single molecule scale. These advances also raise many challenging stochastic modeling and statistical inference problems. First, by zooming in on single molecules, recent nano-scale experiments reveal that some classical stochastic models derived from oversimplified assumptions are no longer valid. Second, the stochastic nature of the experimental data and the presence of latent processes much complicate the statistical inference. In this talk we will use the modeling of subdiffusion phenomenon in enzymatic conformational fluctuation and the inference of DNA hairpin kinetics to illustrate the statistical and probabilistic challenges in single-molecule biophysics.

5. ChIP-chip: Data, Model, and Analysis

Ying Nian Wu, University of California at Los Angeles
ywu@stat.ucla.edu

ChIP-chip (or ChIP-on-chip) is a technology for isolation and identification of genomic sites occupied by specific DNA binding proteins in living cells. The ChIP-chip data can be obtained over the whole genome by tiling arrays, where a peak in the signal is generally observed at a protein binding site. In this talk, I will present a probability model for ChIP-chip data. Then I will propose a model-based computational method for locating and testing peaks for the purpose of identifying potential protein binding sites. I will also present a non-parametric method for identifying and representing peaks in multiple resolutions.

Joint with Ming Zheng, Leah Barrera, and Bing Ren.

6. Model-Based Analysis of Keratin Filament Networks in Scanning Electron Microscopy Images of Cancer Cells

Volker Schmidt, Universität Ulm
volker.schmidt@uni-ulm.de

The keratin filament network is an important part of the cytoskeleton in epithelial cells. It is involved in the regulation of shape and viscoelasticity of the cells. In-vitro studies indicated that geometrical network characteristics, such as filament cross-link density, determine the biophysical properties of the filament network.

Scanning electron microscopy images of filaments were processed by a skeletonisation algorithm based on morphological operators to obtain a graph structure which represents individual filaments as well as their connections. This method was applied to investigate the effects of the so-called transforming growth factor alpha (TGF-alpha) on the morphology of keratin networks in pancreatic cancer cells. By estimating geometrical network characteristics, like the length and orientation distributions of the keratin filaments, and by fitting random tessellation models, a significant alteration of keratin network morphology could be detected in response to TGF-alpha.

7. Semiparametric Regression Models with Staggered Entries and Progressive Multi-Stage Censoring

Murray D. Burke, University of Calgary
burke@math.ucalgary.ca

A semiparametric regression model, when subjects enter the study in a staggered fashion, is studied. A strong martingale approach is used to model the two-time parameter counting processes. It is shown that well-known univariate results such as weak convergence and martingale inequalities can be extended to this two-dimensional model. Strong martingale theory is also used to prove weak convergence of a general weighted goodness-of-fit process and its weighted bootstrap counterpart.

If there is time, I will discuss three progressive multi-stage censoring schemes whereby the experimenter purposely censors a given number of individuals under study at fixed time points and indicate how this can be incorporated into the above model.