NAOMI ALTMAN
Penn State University

Inferring Isoform Expression Using RNA-seq with Restriction Enzyme Tags

Genes in complex organisms may encode multiple mRNAs, called splice variants or isoforms which result in different proteins. Isoform expression is often more relevant than gene expression. The advent of ultra-high throughput sequencing technology has made it possible to directly sequence large numbers of RNA fragments, obtaining a direct measure RNA abundance. Restriction enzyme fragmentation reliably produces a single sequenced fragment per RNA, providing a direct measure of gene expression from tag counts. However, since each gene can have multiple isoforms, the counts must be partitioned to derive accurate isoform expression. The problem reduces to a problem of

DAVID R. BRILLINGER
UC Berkeley

TRandom trajectories in the plane: some theory and applications

Trajectories, paths of moving objects, are basic to classical mechanics. Models developed in mathematical and physical fields may be adapted to explore and model empirical trajectory data that are arising commonly these days. The methods to be discussed will include: spectrum analysis, ordinary and functional stochastic differential equations and potential functions. Data from marine biology (elephant seals, monk seals, whale sharks), animal biology (elk and deer) and soccer will be analyzed. Markov and nonMarkov processes will be considered as will the inclusion of explanatory variables in the models.

A. P. DEMPSTER
Harvard University

Nonparametric Protective Weakening: an Illustration of Dempster-Shafer Inference

Assume observations $X_1, X_2, ..., X_n$ from a univariate population having unknown continuous cdf F(x). What can be said about F(b) - F(a), for prescribed a and b that may or may not depend on the data? Standard sampling theory provides nonparametric tolerance regions with predictable long run properties. Associated long run frequencies may be reinterpreted as Dempster-Shafer (DS) posterior probabilities, including familiar $p \geq 0$ and $q \geq 0$ for and against the truth of an assertion about F(b) - F(a), but possibly allowing a third probability $r \geq 0$ of dont know, so that now $p + q + r = 1$, generalizing the more familiar $p + q = 1$. Such fiducial-like probabilities are not typically protected as long

run betting probabilities, when the counterparty is allowed to choose the side of each bet. DS theory provides a menu of weakened models whose resulting DS inferences are more conservative, and hence more protected. In my talk I will define and begin to explore one such weakened DS analysis.

DON FRASER
University of Toronto

TBA

STEPHEN E. FIENBERG
Carnegie Mellon University

Longitudinal Mixed-Membership Models for Survey Data on Disability

Co-author: Daniel Manrique-Vallier

We describe a new family of models to analyze longitudinal data by combining features from a version of the cross-sectional Grade of Membership Model and from the longitudinal Multivariate Latent Trajectory Model. These models assume the existence of a small number of typical or extreme classes of individuals and model their evolution over time. We regard individuals as belonging to all of these classes in different degree, by considering them as convex weighted combinations of the extreme classes. In this way, we are able to describe distinct general tendencies (the extreme cases) while accounting for the individual variability. We propose a full Bayesian specification and estimation methods based on Markov Chain Monte Carlo sampling. We apply our method to data the National Long Term Care Survey (NLTCS), a longitudinal survey with six completed waves aimed to assess the state and characteristics of disability among U.S. citizens age 65 and above. A simple extension of our methods allows us to answer some relevant questions about the changes in disability across generations.

JERRY LAWLESS
University of Waterloo

Profile Likelihood Methods with Semi-parametric Models

Semi-parametric models are widely used in modern statistical settings. Familiar examples include the Cox model and semi-parametric transformation models in survival analysis, semi-parametric models for dealing with missing covariates in regression analysis, and partially linear models for longitudinal data. This talk will review some recent applications of semi-parametric likelihood, emphasizing the role of profile likelihood for inference about finite-dimensional parameters. Illustrations involving incomplete covariate data and multivariate survival analysis will be discussed.

RICHARD LOCKHART
Simon Fraser University

Goodness of Fit and the Large Hadron Collider

I will discuss the relationship between ideas of Bayes assisted goodness of fit and the problem of detecting the Higgs particle or other previously undiscovered, but expected, phenomena. The goal is a partially Bayesian analysis which will lead to convincing (to people other than the analyst) evidence of the existence of the phenomenon.

PETER MCCULLAGH
University of Chicago

Improper mixtures and Bayes's theorem

Although Bayes's theorem demands a prior that is a probability distribution on the parameter space, formal application of Bayes calculus sometimes generates sensible procedures from improper priors, Pitman's estimator being a good example. However, the formal application of improper priors may also lead to procedures that are paradoxical or otherwise unsatisfactory prompting some authors to insist that all priors be proper. This talk begins with the observation that an improper measure on Θ satisfying a certain countability condition is in fact a probability distribution on the power set. We show how to extend a model in such a way that the parameter space is extended to the power set. The conditions for Bayes's theorem are satisfied under an additional finiteness condition, which is needed for the existence of a sampling region. Lack of interference is a key property of the extension, ensuring that the posterior distribution in the extended space is compatible with the original parameter space. Provided that the key finiteness condition is satisfied, this probabilistic analysis of the extended model may be interpreted as a vindication of improper Bayes procedures generated from the original model.

Joint work with Han Han.

KAI NG
Hong Kong University

How Good Is the Likelihood Asymptotic Inference?

TBA

ANA-MARIA STAICU
North Carolina State University

Generalized Multilevel Functional Regression

We introduce Generalized Multilevel Functional Linear Models (GMFLM), a novel statistical framework motivated by and applied to the Sleep Heart Health Study (SHHS), the largest community cohort study of sleep. The primary goal of SHHS is to study the association between sleep disrupted breathing (SDB) and adverse health effects. An exposure of primary interest is the sleep electroencephalogram (EEG), which was observed for thousands of individuals at two visits, roughly 5 years apart. This unique study design led to the development of models where the outcome, e.g. hypertension, is in an exponential family and the exposure, e.g. sleep EEG, is multilevel functional data. We show that GMFLMs are, in fact, generalized multilevel mixed effect models. Two consequences of this result are that: 1) the mixed effects inferential machinery can be used for GMFLM and 2) functional regression models can be extended naturally to include, for example, additional covariates, random effects and nonparametric components. We propose and compare two inferential methods basedon the parsimonious decomposition of the functional space. This is joint work with C. M. Crainiceanu and C.Di.

LARRY WASSERMAN
Carnegie Mellon University

Minimax Estimation of Filaments and Manifolds

We develop a nonparametric method for estimating filamentary structure and manifolds in point processes. We show that, under weak conditions, the manifold is the medial axis of the support of the distribution. Then we estimate the boundary of the support of the distribution and we convert the boundary estimator into an estimate of the medial axis. We then derive the rate of convergence of the estimator. We also find the minimax lower bound for this problem and we derive a modified estimator that achieves the minimax rate.

CHANGBAO WU
University of Waterloo

Empirical Likelihood Methods for Survey Data

Non-parametric likelihood methods, based on a scale-load approach were first introduced by Hartley and Rao (1968) in the context of simple random or stratified simple random sampling without replacement. Their focus was on point estimation. Twenty years later Owen (1988) introduced empirical likelihood (EL) which is equivalent to the Hartley-Rao

likelihood in the IID case. Owen developed a unified theory including confidence interval estimation. In this talk I will review some work on empirical likelihood in the context of survey data and present pseudo-EL methods to handle complex survey data, including Bayesian pseudo-EL intervals that are asymptotically valid under the design-based set up. Applications to sampling populations containing many zero item values, adaptive sampling for rare and clustered populations, and sampling from multiple incomplete frames will be presented.