
Mathematical Statistics
Statistiques mathématiques
(Org: **Jiahua Chen** (UBC) and/et **Chi Song Wong** (Windsor))

FUQI CHEN, University of Windsor, 401 Sunset Avenue, Windsor, ON N9B 3P4
Equivariance method and generalized inference in two-sample location-scale families

Recently, generalized inference has become an efficient and useful tool which gives more accurate intervals for a variety of intractable complex problems such as the Behrens–Fisher problem. In this talk, we will present a generalized inference solution of typical Behrens–Fisher problem in general location-scale families. The proposed solution is based on the minimum risk equivariant estimators and thus, the underlying approach is an extension of the methods based on maximum likelihood estimators and conditional inference, which have been so far, applied to some specific distributions. Finally, we will present some simulation results as well as analysis results of two real data sets.

JIAHUA CHEN, University of British Columbia
Adjusted Empirical Likelihood with High-Order Precision

Empirical likelihood is a popular nonparametric or semi-parametric statistical method with many nice statistical properties. Yet when the sample size is small, or the dimension of the accompanying estimating function is high, the application of the empirical likelihood method can be hindered by low precision of the chisquare approximation and by non-existence of solutions to the estimating equations. In this paper, we show that the adjusted empirical likelihood is effective at addressing both problems. With a specific level of adjustment, the adjusted empirical likelihood achieves the high-order precision of the Bartlett correction, in addition to the advantage of a guaranteed solution to the estimating equations. Simulation results indicate that the confidence regions constructed by the adjusted empirical likelihood have coverage probabilities comparable to or substantially more accurate than the original empirical likelihood enhanced by the Bartlett correction.

ABBAS KHALILI, McGill University, Dept. of Mathematics and Statistics
New estimation and variable selection method in mixture-of-experts models

We study estimation and variable selection problems in mixture-of-experts (MOE) models. A new modified maximum likelihood estimation (MMLE) method is proposed. It is shown that the MMLE is root- n consistent, and simulations indicate its better finite sample behavior compared to the ordinary MLE. For variable selection, we apply two penalty functions to the modified likelihood. The method is computationally efficient, and theoretically it is shown to be consistent in variable selection. Two Bayesian information criteria are suggested for data adaptive choice of tuning parameters. A modified EM-Newton–Raphson algorithm is developed for numerical computations. The performance of the method is also studied through simulations. A real data analysis is presented.

REG KULPERGER, University of Western Ontario
A Corporate Exit Model, Smooth Baseline Hazards, and Biostatistics Tools in Finance

There is a large amount of publicly available financial information on publicly traded corporations, usually on a quarterly year time period. These same corporations also undergo bankruptcy or acquisition through merger. It is natural to model these in a discrete time framework due to the nature of the data. We consider a bivariate discrete time hazard model. The framework is similar to that in classical biostatistics modeling, where one treats the two forms of exit from the system, namely bankruptcy and merger/acquisition, but with additional information on the type of exit. In biostatistics the cause of exit (usually death) is not known explicitly.

Such models are constructed and fit to a data base of some 12,000 publicly traded US corporations. With a large number of covariates some data reduction is needed. Both in and out of sample prediction is considered. A constant baseline hazard model does not fit well, so a smooth baseline hazard model is considered. This later model seems to give a reasonable fit in terms of prediction, and has a nice robustness property. Some tools for model assessment are developed. One useful tool for this is a limit theorem on rare multinomials which is originally due to McDonald (1980).

This is joint work with Dr Taehan Bae.

HELENE MASSAM, York University, 4700 Keele Street, Toronto, ON, M3J 1P3

A conjugate prior for discrete hierarchical loglinear models

In Bayesian analysis of multi-way contingency tables, the selection of a prior distribution for either the loglinear parameters or the cell probabilities parameters is a major challenge. In this talk we define a flexible family of conjugate priors for the wide class of discrete hierarchical loglinear models which includes the class of graphical models. These priors are defined as the Diaconis–Ylvisaker conjugate priors on the loglinear parameters subject to “baseline constraints” under multinomial sampling. We also derive the induced prior on the cell probabilities and show that the induced prior is a generalization of the hyper Dirichlet prior. We show that this prior has several desirable properties and illustrate its usefulness by identifying the most probable decomposable, graphical and hierarchical loglinear models for a six-way contingency table.

This work has been done in cooperation with Jinnan Liu and Adrian Dobra.

JOHANNA NESLEHOVA, McGill University, 805 Sherbrooke Street West, Montreal, Quebec H3A 2K6

Goodness-of-fit tests for bivariate extreme-value dependence

It is often reasonable to assume that the dependence structure of a bivariate continuous distribution belongs to the class of extreme-value copulas. The latter are characterized by their Pickands dependence function. The talk is concerned with a procedure for testing whether this function belongs to a given parametric family. The test is based on a Cramér–von Mises statistic measuring the distance between an estimate of the parametric Pickands dependence function and either one of two nonparametric estimators thereof studied by Genest and Segers (2009). As the limiting distribution of the test statistic depends on unknown parameters, it must be estimated via a parametric bootstrap procedure, whose validity is established. Monte Carlo simulations are used to assess the power of the test, and an extension to dependence structures that are left-tail decreasing in both variables is considered.

WEI NING, Bowling Green State University, Department of Mathematics and Statistics

A Generalized Lambda Distribution (GLD) Change Point Model For the Detection of DNA Copy Number Variations in Array CGH Data

In this talk, we study the detection of the multiple change points of parameters of generalized lambda distributions (GLD). The advantage of studying the GLD is that the GLD family is broad and flexible, compared to the other distributions. There are fewer restrictions on the distribution while fitting the data. We combine the binary segmentation procedure together with Schwarz information criterion (SIC) to search all the possible change points in the data. The method is applied on fibroblast cancer cell line data which is publicly available, and change points are successfully located.

JUNFENG SHANG, Bowling Green State University, Bowling Green, OH 43403

A Multiple Comparison Procedure Based on a Variant of the Schwarz Information Criterion in Mixed Models

Repeated measurements are collected in a variety of situations and are generally characterized by a mixed model where the correlation within the subject is specified by the random effects. In such a mixed model, we propose a multiple comparison procedure based on a variant of the Schwarz information criterion (SIC, Schwarz, 1978). The derivation of SIC indicates that SIC serves as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. Therefore, an approximated posterior probability for a candidate model can be calculated based upon SIC. We suggest a variant

of SIC which includes the terms which are asymptotically negligible in the derivation of SIC. The variant improves upon the performance of SIC in small and moderate sample-size applications. Based upon the proposed variant, the corresponding posterior probability is calculated for each candidate model. A hypothesis testing for multiple comparisons involves one or more models in the candidate class, the posterior probability of the hypothesis testing is therefore evaluated as the sum of the posterior probabilities for the models associated with the testing. The approximated posterior probability based on the variant accommodates the effect of the prior on each model in the candidate class, and therefore is more effectively approximated than that based on SIC for conducting multiple comparisons. We derive the computational formula of the approximated posterior probability based on the variant in the mixed model. The applications demonstrate that the proposed procedure based on the SIC variant can perform effectively in multiple comparisons.

CHRISTOPHER SMALL, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, N2L 3G1
Analyzing the UK 2001 foot-and-mouth disease outbreak using stochastic graph models

In recent decades stochastic graphs have been used in many fields to explain the evolution of a set of random objects (vertices), along with a relationship structure (edges). We consider statistical inference in a dynamic random graph, in the absence of edge information. It is shown that the dynamic behavior of the graph, accompanied with the vertex information is a useful in making inference about the edges. The problem is motivated by the foot-and-mouth disease (FMD) outbreak in the UK in 2001. A stochastic Euclidean graph model with Markov property, is introduced to model this epidemic. In addition, it is shown that the existing information, is sufficient to draw inference about the model and hence the missing edges.

DAVID STEPHENS, McGill University
Bayesian Nonparametric Hypothesis Testing in Two Sample Problems

In this talk I will discuss Bayesian hypothesis testing in the two sample problem. I will introduce some procedures based on a Bayesian nonparameteric formulation, and examine their performance in comparison to classical nonparametric procedures. This is joint work with Chris Holmes (Oxford), François Caron (Bordeaux) and Jim Griffin (Kent).

DAVID WOLFSON, Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, QC, H3A 2K6
The Statistical Analysis of Survival Data from Prevalent Cohort Studies with Follow-up

Estimation of the incidence rate of a disease generally entails the follow-up of a disease-free cohort until a sufficient number of incident cases of the disease have been observed. Sometimes it is possible, however, to avoid the time and cost of carrying out an incidence study by following prevalent cases with the disease forward for a relatively short time period. That is, we may identify prevalent cases through a cross-sectional survey and follow them forward as part of what is known as a prevalent cohort study with follow-up. In this presentation we show how one may find the maximum likelihood estimator of the age-specific constant incidence rate from a prevalent cohort study with follow-up. Our key expression is related to the well-known epidemiological relationship between incidence, prevalence and disease duration. We apply our results to estimate the incidence rate of dementia in Canada.

Joint work with Victor Addona (Macalester College, St. Paul, MN) and Masoud Asgharian (McGill University).

CHI SONG WONG, University of Windsor
Robustness of optimal designs for correlated random variables

Suppose that $Y = (Y_i)$ is a normal random matrix with mean Xb and covariance $\sigma^2 I_n$, where b is a p -dimensional vector (b_j) , $X = (X_{ij})$ is an $n \times p$ matrix with $X_{ij} \in \{-1, 1\}$; this corresponds to a factorial design with $-1, 1$ representing low or high level respectively, or corresponds to a weighing design with $-1, 1$ representing an object j with weight b_j placed on the left and right of a chemical balance respectively. E -optimal designs Z are chosen that are robust in the sense that they remain

E -optimal when the covariance of $Y_i, Y_{i'}$ is $\rho > 0$ for $i \neq i'$. Within a smaller class of designs similar results are obtained with respect to a general class of optimality criteria which include the A - and D -criteria.

The talk is based on my three joint papers with Joe Masaro published in 2008 in JSPI and LAA.

YUEHUA WU, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3

A note on the convergence rate of the kernel density estimator of the mode

In this talk, the mode estimator based on the Parzen–Rosenblatt kernel estimator is considered (Parzen, 1962). In light of Shi et al. (2009), under mild conditions, we establish the relationship between the convergence rate of the mode estimator and the window width. In this way, we obtain a better convergence rate of the mode estimator.

This is joint work with X. Shi and B. Miao.

GRACE YI, University of Waterloo

Some thoughts on composite likelihood

The composite likelihood method has been proposed and systematically discussed by Besag (1974), Lindsay (1988), and Cox and Reid (2004). The approach based on using the composite likelihood, especially the pairwise likelihood, has received increasing attention in recent years due to the simplicity in defining the objective function and computational advantages when dealing with data with complex structures. In this talk, I will discuss some modeling issues concerning the composite likelihood formulation.

This is joint work with Nancy Reid.

RONG ZHU, McMaster University, Hamilton, Ontario, Canada

The geometric down-weighting method and its applications

The geometric down-weighting method can be applied to enlarge an existing discrete distribution family. The enlarged family has one more parameter which regulates the decreasing rate of probability mass function, thus, yielding new moment features. Applying the geometric down-weighting method to a family with infinite mean, we can obtain an enlarged family which can have both finite and infinite means. Such an enlarged family can accommodate for the heavy-tailed count data, because it allows various tail heaviness. Particularly, when applying this method to the two-parameter discrete stable family which has infinite mean, we obtain a three-parameter discrete distribution family called the generalized Poisson-inverse Gaussian (GPIG). Apart from the extremely heavy-tailed discrete stable, the GPIG family extends the over-dispersed Poisson-inverse Gaussian (PIG) and also includes the equally-dispersed Poisson. Therefore, the GPIG family is flexible in handling less or more over-dispersed count data. We illustrate the GPIG family by the application of the citation counts of published articles in 1990 in JASA and JSPI respectively.