Asymptotic Thoery, Robust Statistics, and Quantile Regression

Event Program
October 16, 2021
9:00 AM - 5:00 PM

A Workshop Celebrating the Contributions of Stephen Portnoy

ILLINOIS

Meetings Conduct Policy

The Department of Statistics is committed to providing an atmosphere that encourages the free expression and exchange of ideas. Consistent with this commitment, it is the policy of the department that all participants in University of Illinois activities will enjoy a welcoming environment free from unlawful discrimination, harassment, and retaliation. All participants in university activities also agree to comply with all rules and conditions of the activities, which are subject to change without notice.

Please read the complete Code of Conduct Policy: https://www.ethics.uillinois.edu/compliance/university_code_of_conduct.

Asymptotic Theory, Robust Statistics, and Quantile Regression: A Workshop Celebrating the Contributions of Stephen Portnoy

October 16, 2021
Hybrid (Online & In-Person)
I-Hotel and Conference Center
1900 S. 1st St, Champaign, IL 61820

Workshop Committee:
Chair: Xuming He (University of Michigan)
Co-Chair: Xiaofeng Shao (University of Illinois)

Event Schedule

Opening Statements

8:50-9:00	Xiaofeng Shao (Committee Co-chair) Bo Li (Department Chair)

Session 1: Chair, Xiaofeng Shao

9:00-9:40	Roger Koenker, University College London
9:40-10:10	Elvezio Ronchetti, University of Geneva
10:10-10:40	Break

Session 2: Chair, Sabyasachi Chatterjee

10:40-11:05	Tereza Neocleous, University of Glasgow
11:05-11:30	Jana Jureckova, Charles University, Prague
11.30-12:00	Steve Stigler, University of Chicago
12:00-1:20	Lunch at I-Hotel

Session 3: Chair, Daniel Eck

1:20-1:50	Ying Wei, Columbia University
1:50-2:20	Quanshui Zhao, Citi
2:20-2:50	Xiaohui Chen, University of Illinois
2:50-3:20	Break

Session 4: Chair, Dave Zhao

3:20-3:50	Stanislav Volgushev, University of Toronto
3:50-4:20	Sandy Zabell, Northwestern University
4:20-5:00	Peter Bickel, University of California, Berkeley
5:00	Steve Portnoy, University of Illinois
6:00-8:00	Dinner at I-Hotel
Dinner Speakers	Douglas Simpson, University of Illinois Peter Imrey, Cleveland Clinic And others

Talk Abstracts

Session 1: Chair, Xiaofeng Shao

9:00-9:40 Roger Koenker, University College London

Title: Ranking and Selection from Pairwise Comparisons: Empirical Bayes Methods for Citation Analysis

Abstract: We study Stigler's (1994) model of citation flows among journals adapting the pairwise comparison model of Bradley and Terry to do ranking and selection of journal influence based on nonparametric empirical Bayes procedures. Comparisons are made with a variety of other ranking procedures.

9:40-10:10 Elvezio Ronchetti, University of Geneva

Title: Optimal transportation through saddlepoints

Abstract: We showcase some unexplored connections between saddlepoint approximations, measure transportation, and some key topics in information theory by reviewing selectively some fundamental results available in the literature. We start with the link between Esscher's tilting (which is a result rooted in information theory and that lies at the heart of saddlepoint approximations) and the solution of the dual Kantorovich problem (which lies at the heart of measure transportation theory) via the Legendre transform of the cumulant generating function. We then investigate these links in the framework of M-estimators and quantile regression. The unveiled connections offer the possibility to view saddlepoint approximations from different angles, putting under the spotlight the links to e.g. convex analysis (via the notion of duality) or differential geometry (via the notion of geodesic).

Joint work with Davide La Vecchia and Andrej Ilievski.

Session 2: Chair, Sabyasachi Chatterjee

10:40-11:05 Tereza Neocleous, University of Glasgow

Title: A Bayesian approach to quantile regression for counts

Abstract: The classical quantile regression approach for count data relies on “jittering”, ie adding uniform random noise to the dependent variable, and thus overcoming the problem that the conditional quantile function is not a continuous function of the parameters of interest. Although jittering enables estimation of the conditional quantiles, it has the drawback that, for small values of the dependent variable, the added noise can have a relatively large influence on the estimated quantiles. The talk will outline an alternative Bayesian approach to quantile regression for count data, which is based on an adaptive Dirichlet Process mixture of COM-Poisson regression models. The quantiles are determined by estimating the density of the data using an MCMC algorithm that can be applied to distributions on which the likelihood can only be computed up to a normalising constant.

Session 2: Chair, Sabyasachi Chatterjee

11:05-11:30 Jana Jureckova, Charles University, Prague

Title: Rank Test in Linear Model with Autoregressive Errors

Abstract: In the linear regression model with possibly autoregressive errors, we propose a family of nonparametric tests for regression under a nuisance autoregression of error terms. The tests avoid the estimation of nuisance parameters, in contrast to the tests proposed in the literature. We consider the linear regression model of order *s*, whose model errors follow a p-th -order stationary autoregressive process.

$$y_t = \beta_0 + \mathbf{x_t}^\top \boldsymbol{\beta}^* + \varepsilon_\mathbf{t} = \beta_\mathbf{0} + \mathbf{x_{t1}}\beta_\mathbf{1} + ... + \mathbf{x_{tp}}\beta_\mathbf{p} + \varepsilon_\mathbf{t},$$

$$\varepsilon_t = \varphi_0 + \varphi_1\varepsilon_{t-1} + u_t + \ldots + \varphi_p\varepsilon_{t-p},\ t = 1, 2, ..., n,$$

where $\varphi_0, \varphi_1, \ldots, \varphi_p$ unknown autoregression parameters. The innovations ut are assumed being independently and identically distributed (*i.i.d.*) with a continuous distribution function *F* and density *f* exponentially tailed, satisfying $E(u_t) = 0, \quad Var(u_t) = \sigma^2 < \infty$ otherwise generally unknown. We construct the rank tests of the hypothesis:

$$\mathbf{H_0}: \ \boldsymbol{\beta}^* = \mathbf{0}, \ \text{ with } \ \beta_\mathbf{0}, \ (\varphi_\mathbf{0}, \varphi_\mathbf{1}, \ldots, \varphi_\mathbf{p})' \neq \mathbf{0} \ \text{ unspecified}$$

The tests of $\mathbf{H}_0$ are based on the autoregression rank scores and on the quadratic forms of linear autoregression rank statistics for the hypothetical model. The score-generating functions should satisfy the Cherno-Savage condition, what covers the Wilcoxon, van der Waerden and median scores. The tests are asymptotically equivalent to the rank tests of $\mathbf{H}_0$ in the situation without nuisance autoregression, under $\mathbf{H}_0$ as well as under local alternatives.

The idea of similar tests originates from the joint work with Steve Portnoy, Roger Koenker and Cornelius Gutenbrunner, made once in Urbana-Champaign.

Reference:
C. Gutenbrunner, J. Jureckova, R. Koenker, S. Portnoy (1993):
Tests of linear hypotheses based on regression rank scores,
Journal of Nonparametric Statistics, 2:4, 307-331.

Session 2: Chair, Sabyasachi Chatterjee

11.30-12:00 Steve Stigler, University of Chicago

Title: The Law of the Maturity of Chances

Abstract: Narrow minded probabilists believe this Law is a delusion, a fallacy. To the contrary, I will present proof, both mathematical and empirical, drawn from work of Laplace and others, from two hundred years ago.

Session 3: Chair, Daniel Eck

1:20-1:50 Ying Wei, Columbia University

Title: Integrated quantile rank test (IQRAT) for gene-level associations

Abstract: Gene-based testing is a commonly employed strategy in many genetic association studies. Gene-trait associations can be complex due to underlying population heterogeneity, gene-environment interactions, and various other reasons. Existing gene-based tests, such as Burden and Sequence Kernel Association Tests (SKAT), are mean-based tests, and may miss or underestimate higher-order associations that could be scientifically interesting. In this paper we propose a new family of gene-level association tests that integrate quantile rank score process to better accommodate complex associations. The resulting test statistics have multiple advantages: (1) they are almost as efficient as the best existing tests when the associations are homogeneous across quantile levels and have improved efficiency for complex and heterogeneous associations, (2) they provide useful insights into risk stratification, (3) the test statistics are distribution-free and could hence accommodate a wide range of underlying distributions, and (4) they are computationally efficient. We established the asymptotic properties of the proposed tests under the null and alternative hypotheses and conducted large-scale simulation studies to investigate their finite sample performance. The performance of the proposed approach is compared with that of conventional mean-based tests, i.e., the Burden and SKAT tests, through simulation studies and applications to a Metabochip dataset on lipid traits, and to the genotype-tissue expression data in GTEx to identify eGenes, i.e., genes whose expression levels are associated with cis-eQTLs.

Session 3: Chair, Daniel Eck

1:50-2:20 Quanshui Zhao, Citi

Title: Quantile Regression of Binary Data

Abstract: Regression quantiles for linear models are widely used in analyzing data of continuous distribution. With binary response data, regression quantiles are not uniquely defined. In this study, we heuristically define a quantile regression estimator of binary data by maximizing the unevenly powered likelihood function

$$\hat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmax}} \prod_{i=1}^{n} p^{\lambda}(\mathbf{x}_i, \beta) q^{1-\lambda}(\mathbf{x}_i, \beta)$$

where 0 < λ < 1 and p(xi, β) = P (yi = 1) = 1 − q(xi, β) is a link function for response variable *yi* corresponding to regressor x_i. Suppose $y_i = I\{Y_i > 0\}$ is an indicator function of a latent variable Y_i defined by a linear model

$$Y_i = \beta \mathbf{x}_i + u_i,$$

where $\{u_i\}$ are i.i.d. logistically distributed errors and $p(x_i, \beta) = 1/(1 + e^{-\beta xi})$ is the logistic function. The estimator β̂(λ) converges to the regression quantile of the above linear model. Test of heteroscedasticity can be established.

For application of this model, we consider a financial product of collateralized debt obli- gation (CDO), where the credit risk level of a CDO tranche bond may be rated based on the predicted one-year default probability. This probability can be explicitly derived from analytical models or estimated through logistic regression using historical data. With the quantile regression technique, one can extend the analysis of CDO default probability from expectation to quantiles.

Session 3: Chair, Daniel Eck

2:20-2:50 Xiaohui Chen, University of Illinois

Title: Maximum likelihood estimation of potential energy in interacting particle systems from singletrajectory data

Abstract: We consider the parameter estimation problem for the quadratic potential energy in interacting particle systems from continuous-time and single-trajectory data. Even though such dynamical systems are high-dimensional, we show that the vanilla maximum likelihood estimator (without regularization) is able to estimate the interaction potential parameter with optimal rate of convergence simultaneously in mean-field limit and in long-time dynamics. This to some extend avoids the curse-of-dimensionality for estimating large dynamical systems under symmetry of the particle interaction.

Session 4: Chair, Dave Zhao

3:20-3:50 Stanislav Volgushev, University of Toronto

Title: Structure learning for Extremes

Abstract: Extremal graphical models are sparse statistical models for multivariate extreme events. The underlying graph encodes conditional independencies and enables a visual interpretation of the complex extremal dependence structure. For the important case of tree models, we provide a data-driven methodology for learning the graphical structure. We show that sample versions of the extremal correlation and a new summary statistic, which we call the extremal variogram, can be used as weights for a minimum spanning tree to consistently recover the true underlying tree. Remarkably, this implies that extremal tree models can be learned in a completely non-parametric fashion by using simple summary statistics and without the need to assume discrete distributions, existence of densities, or parametric models for marginal or bivariate distributions. Extensions to more general graphs are also discussed.

Session 4: Chair, Dave Zhao

3:50-4:20 Sandy Zabell, Northwestern University

Title: The Bowl of Ariantas

Abstract: Steve Portnoy has exhibited a longstanding interest in statistics and history. In this talk I discuss a curious passage in Herodotus that, taken literally, might permit us to estimate the population of ancient Scythia. Whether or not this is actually the case, and the connection of the census Herodotus describes with others said to have occurred in later times, is the subject of this talk.

4:20-5:00 Peter Bickel, University of California, Berkeley

Title: Independence and functional dependence

Abstract: Chatterjee(2019) (see also Dette et al(2013)),introduced a novel rank based measure of dependence between X and Y real,which was 0 iff X and Y were independent ,and 1 iff Y=h(X) for some h..Subsequent work by Cao and Bickel(2020) and Shi,Drton and Han(2020) pointed to poor local power properties for testing independence of this statistic as compared to classical procedures ,such as those of Spearman,Blum,Kiefer and Rosenblatt and others.We show that the statistics such as Chatteejee's capture dependence in a way complementary to that of classical tests and can in fact be combined with them to yield better results for both types of procedures.Measuring functional dependence can be combined with these properties to yield simple measures which have somewhat better independence testing behaviour than Chatterjees and point to functional dependence with the same reliability.

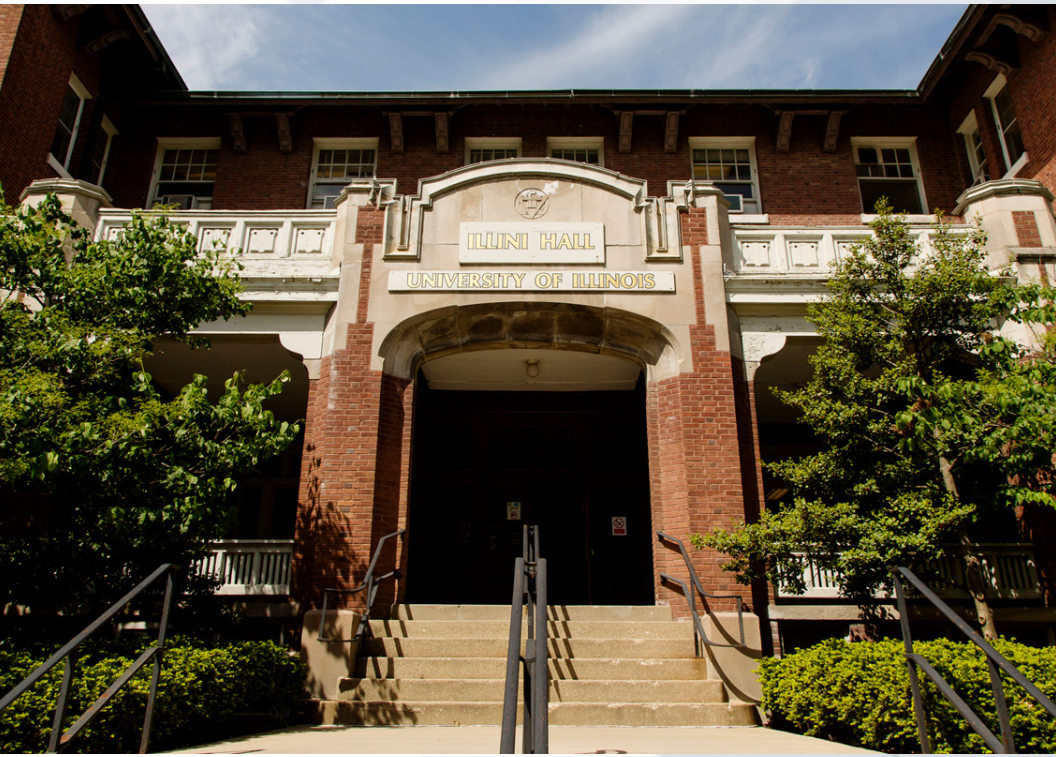
5:00-5:15 Steve Portnoy, University of Illinois

Title: What I have learned in 60+ years

About the Department of Statistics

The Department of Statistics at the University of Illinois at Urbana-Champaign was established in 1985 with the approval of the Board of Trustees on April 18, 1985 and the Illinois Board of Higher Education on September 4, 1985. In its approval letter, the Board of Trustees wrote:

The purpose of the Department of Statistics, now a division within the Department of Mathematics, is to provide a clear focus and visibility for statistics as a separate discipline. The need for this focus is recognized by most major academic institutions. With resources now in the Division of Statistics and those committed by the College, the Department would provide a more vigorous program by augmenting graduate enrollment in statistics and developing cross-disciplinary programs and research.



The original proposal from the Division of Statistics developed under the leadership of Jerome Sacks, Division Head and Founding Department Head, described the objectives of establishing a freestanding department as follows:

The chief motivation for creating a statistics department is to provide a focus and visibility for statistics as a discipline in order to further statistical research, to train statisticians to fill the serious shortages of qualified statisticians at all levels, and to promote statistical literacy in a general educational context. In order to reach these objectives a number of specific goals are to be met:

- Augmentation and widening of undergraduate offerings to bolster the undergraduate degree programs. Expansion of introductory courses at elementary levels to enhance the spread of statistical literacy.
- Increase in the availability of fundamental and specialized courses in applied and theoretical statistics at the graduate level for statistics majors and for other “user” departments.
- Expansion of statistical consulting services for researchers on the campus, for state agencies and for industry.
- Increase in the enrollment of graduate students in statistics.
- Development of cross-disciplinary programs and research.
- Expansion of the statistics faculty.

The drive toward accomplishing these goals encapsulates much of the subsequent development of the Department of Statistics, its programs, its research profile and its interactions with other programs, and its contributions to the campus, students and alumni. Furthermore, with the continued growth, indeed explosion, of data intensive research, analysis and decision making in all aspects of modern society, the need for research and education in Statistics is even stronger today than 30 years ago.

Along with Robert Bohrer, Dennis Jennings, Kumar Joag-dev, John Marden, Adam Martinsek, William Stout, Jerome Sacks, Robert Wijsman, and Stanley Wasserman, Stephen Portnoy was a founding faculty member of the new Department of Statistics at the University of Illinois Urbana-Champaign in 1985.

We count on the generosity of alumni and friends to support students as they embark on earning a world-class education and to fund faculty members as they conduct world changing research and train students. Your investment makes a difference.

Donor support allows us to:

- Offer scholarships to academically strong undergraduate and graduate students, as well as to those with financial need
- Recruit and retain the best faculty members
- Provide state-of-the-art facilities that house important research and teaching
- Provide learning and professional development opportunities for our students

To make a gift, please visit: stat.illinois.edu/giving

Gift opportunities include:

Stephen L. Portnoy Fellowship in Statistics
To support fellowships in the Department of Statistics, in the College of Liberal Arts and Sciences at the University of Illinois at Urbana-Champaign.

Statistics Opportunities Scholarship Fund
A fund for our corporate and professional partners to provide scholarships to students in the Department of Statistics at the University of Illinois Urbana-Champaign.

David H. Blackwell Diversity Fund
David H. Blackwell's illustrious accomplishments have inspired the Department of Statistics to initiate several important programmatic efforts in his name. With support from our friends and alumni, we will be able to continue our graduate scholarship program for diversity, establish an undergraduate summer research program aimed to create opportunities for underrepresented minority populations in statistics, and support a distinguished lecture series.

LAS Annual Fund for Statistics
To provide unrestricted support for the Department of Statistics in the College of Liberal Arts and Sciences at the University of Illinois at Urbana-Champaign.

Department of Statistics
725 S. Wright St.,MC374
Champaign, IL 61820
stat-office@illinois.edu
stat.illinois.edu

ILLINOIS