RUTGERS
School of Arts and Sciences

Department of Statistics

Conference on Advances in Bayesian & Frequentist Statistics

With a Celebration of the 80th Birthday of Professor William E. Strawderman

April 1–2, 2022

Fiber Optic Materials Research Building, Auditorium EHA

101 Bevier Road, Piscataway, NJ

Funding from NSF is gratefully acknowledged for the participation support.

The Department of Statistics of Rutgers University welcomes you to the Conference on Advances in Bayesian and Frequentist Statistics, with a Celebration of the 80th Birthday of Professor William E. Strawderman.

Professor Strawderman, currently Distinguished Professor of Statistics at Rutgers University, is a world renowned leading expert in Decision Theory, Bayesian Statistics and Multivariate Analysis. He received a BS (1963) in Engineering from University of Rhode Island, an MS (1965) in Mathematics from Cornell University, and an MS (1967) and PhD (1969) in Statistics from Rutgers University. Prior to joining Rutgers, he was a Member of Technical Staff at Bell Labs. He served as Department Chair of Statistics at Rutgers for many years and played a decisive role in building up the department. He also held visiting or adjunct positions at many universities and research centers worldwide, including Cornell, Princeton, Stanford, Universite de Paris (France), Université de Rouen (France), University of British Columbia (Canada), Université de Sherbrooke (Canada), University of Rome (Italy), Academica Sinica (Taiwan), ETS, and NIST. He has published 2 books and more than 220 papers. He has been invited as keynote speaker in numerous major international conferences, including an IMS Medallion Lecture. He is an elected fellow of the Institute of Mathematical Statistics and of the American Statistical Association.

Program Committee: Rong Chen (co-chair), Dominique Fourdrinier, Ed George, John Kolassa, Zhiqiang Tan (co-chair), Martin Wells, Cun-Hui Zhang

Poster Award Review Committee: Yaqing Chen, John Kolassa (chair), Subrata Kundu, Nilanjan Laha, Xialu Liu, Yaakov Malinovsky, Eric Marchand, Takeru Matsuda, Emma Zhang

Local Committee: Javier Cabrera, Zijian Guo, Qiyang Han, Ying Hung, John Kolassa, Zhiqiang Tan (chair), Guanyang Wang, Han Xiao, Min Xu, Linjun Zhang

Conference Coordinator: Eileen Sharkey, `esharkey@stat.rutgers.edu`

RUTGERS
School of Arts and Sciences
Department of Statistics

Conference on Advances in Bayesian and Frequentist Statistics
with a Celebration of the 80th Birthday of Professor William E. Strawderman

Program

Friday, April 1

Breakfast and Registration (8:00am – 8:45am)

Opening Remarks (8:45am – 9:00am)

Thu D. Nguyen; Dean of Mathematical and Physical Sciences, School of Arts and Sciences, Rutgers University

Rong Chen; Chair of Department of Statistics, Rutgers University

Session I (9:00am – 10:40am)
Chair: **Cun-Hui Zhang**; Rutgers University

9:00am	**Jim Berger**; Duke University *The Many Facets of the Strawderman Prior*
9:40am	**Iain Johnstone**; Stanford University *Expectation Propagation in Mixed Models*
10:20am	**Linjun Zhang**; Rutgers University *Discussion*

Coffee Break (10:40am – 11:00am)

Session II (11:00am – 12:40pm)
Chair: **Javier Cabrera**; Rutgers University

11:00pm	**Genevera Allen**; Rice University *Graph Learning for Functional Neuronal Connectivity*

11:40am	**Martin Wells**; Cornell University *On Graphical Models and Convex Geometry*
12:20pm	**Yaqing Chen and Min Xu**; Rutgers University *Discussion*

Lunch (12:40pm – 1:40pm) Lunch box will be provided on-site.

Session III (1:40pm – 3:20pm)
Chair: **Regina Liu**; Rutgers University

1:40pm	**Dominique Fourdrinier**; Université de Rouen, France *Data Based Loss Estimation of the Mean of a Spherical Distribution with a Residual Vector*
2:20pm	**Éric Marchand**; Université de Sherbrooke, Canada *The Search for Efficient Predictive Density Estimators*
3:00pm	**Pierre Bellec**; Rutgers University *Discussion*

Coffee Break (3:20pm – 3:40pm)

Session IV (3:40pm – 5:20pm)
Chair: **David Tyler**; Rutgers University

3:40pm	**Ying Hung**; Rutgers University *Functional-Input Gaussian Processes with Applications to Inverse Scattering Problems*
4:20pm	**Andrew L. Rukhin**; National Institute of Standards and Technology, USA *Heterogeneous Data and Objective Priors*
5:00pm	**Tirthankar Dasgupta**; Rutgers University *Discussion*

Poster Session (5:20pm – 6:10pm)

Banquet (6:20pm – 9:00pm)
Busch Student Center, Multi-Purpose Room (next to the Gerlanda's Pizza)

Saturday, April 2

Breakfast and Registration (8:30am – 9:00am)

Session V (9:00am – 10:40am)
Chair: **Minge Xie**; Rutgers University

9:00am	**Ed George**; University of Pennsylvania *From Minimax Shrinkage Estimation to Minimax Shrinkage Prediction*
9:40am	**Christian P. Robert**; Universitè Paris Dauphine PSL, France *Bayesian Model Choice in Finite and Infinite Mixtures*
10:20am	**Ruobin Gong**; Rutgers University *Discussion*

Coffee Break (10:40am – 11:00am)

Session VI (11:00am – 12:40pm)
Chair: **Michael Stein**; Rutgers University

11:00am	**Jianqing Fan**; Princeton University *How Do Noise Tails Impact on Deep ReLU Networks?*
11:40am	**Emma Zhang**; University of Miami *Network Community Detection: New Algorithms and Goodness-of-Fit Tests*
12:20pm	**Koulik Khamaru and Guanyang Wang**; Rutgers University *Discussion*

Lunch (12:40pm – 1:40pm) Lunch box will be provided on-site.

Session VII (1:40pm – 3:20pm)
Chair: **Steve Buyske**; Rutgers University

1:40pm	**Takeru Matsuda**; RIKEN Center for Brain Science, Japan *Matrix Estimation by Singular Value Shrinkage*
2:20pm	**Fatiha Mezoued**; Ecole Nationale Supérieure de Statistique et d'Économie Appliquée, Algeria *Estimation of the Inverse Scatter Matrix for a Scale Mixture of Wishart Matrices under Efron-Morris Type Losses*

3:00pm	**Qiyang Han**; Rutgers University *Discussion*

Coffee Break (3:20pm – 3:40pm)

Session VIII (3:40pm – 5:20pm)
Chair: **John Kolassa**; Rutgers University

3:40pm	**Larry Wasserman**; Carnegie Mellon University *Causal Inference in the Time of Covid-19*
4:20pm	**Robert Strawderman**; University of Rochester *Robust Q-learning*
5:00pm	**Nicole Pashley**; Rutgers University *Discussion*

Poster Award Session and Closing Remarks (5:20pm – 5:30pm)
Rong Chen; Chair of Department of Statistics, Rutgers University

Posters

1. **Nilanjana Chakraborty**, University of Florida.
   A Bayesian framework for sparse estimation in high-dimensional mixed frequency Vector Autoregressive models.

2. **Emily Diana**, University of Pennsylvania.
   Multiaccurate Proxies for Downstream Fairness.

3. **Yajie Duan**, Rutgers University.
   Particle Count Estimation in Dilution Series Experiments.

4. **Christopher Geoga**, Rutgers University.
   Half-Spectral Covariance Functions for Nonstationary Space-Time Gaussian Processes.

5. **Wei Jin**, Johns Hopkins University.
   A Bayesian Decision Framework for Optimizing Sequential Combination Antiretroviral Therapy in People with HIV.

6. **Mitchell Krock**, Rutgers University.
   Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails.

7. **Yu-Chun Kao**, Rutgers University.
   Rate optimal and adaptive Estimation of the Location of a Symmetric Distribution.

8. **Gen Li**, University of Pennsylvania.
   Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model.

9. **Xiang Li**, George Washington University.
   Heterogenous Block Covariance Model for Community Detection.

10. **Zebang Li**, Rutgers University.
    Autoregressive Models for Tensor-Valued Time Series.

11. **Chun Pang Lin**, Rutgers University.
    Automated Spot Counting in Microbiology.

12. **Ryumei Nakada**, Rutgers University.
    The Power of Contrast for Feature Learning: A Theoretical Analysis.

13. **Prabrisha Rakshit**, Rutgers University.
Inference for Case Probability in High-dimensional Logistic Regression.

14. **Arun Ravichandran**, Rutgers University.
Optimal Allocation of Sample Size for Randomization-Based Inference from 2^K Factorial Designs.

15. **Samrat Roy**, University of Pennsylvania.
Regularized High Dimension Low Tubal-Rank Tensor Regression.

16. **Kai Tan**, Rutgers University.
Noise Covariance Estimation in Multi-Task High-dimensional Linear Models.

17. **Ye Tian**, Columbia University.
Transfer Learning under High-dimensional Generalized Linear Models.

18. **Ziyue Wang**, Rutgers University.
Tractable and Near-Optimal Adversarial Algorithms for Robust Estimation.

19. **Yuling Yan**, Princeton University.
Inference for Heteroskedastic PCA with Missing Data.

20. **Yachong Yang**, University of Pennsylvania.
Doubly Robust Calibration of Prediction Sets under Covariate Shift.

21. **Maxine Yu**, Princeton University.
Understanding Implicit Regularization in Over-Parameterized Single Index Model.

22. **Ruofan Yu**, Rutgers University.
Monte Carlo with Funnels: Optimal Path Finding.

23. **Pei Zhang**, University of Maryland.
Polygenic Risk Scores for Longitudinal Data.

24. **Zhe Zhang**, Rutgers University.
High-Dimensional Differentially-Private EM Algorithm: Methods and Optimality.

25. **Lili Zheng**, Rice University.
Gaussian Process Parameter Estimation with Mini-batch Stochastic Gradient Descent: Convergence Guarantees and Empirical Benefits.

26. **Zheshi Zheng**, Rutgers University.
Robust and Individualized Conformal Predictive Inference in Survival Analysis.

Abstracts

Graph Learning for Functional Neuronal Connectivity

April 01
11:00am

Genevera Allen

Associate Professor and Founder and Faculty Director of the Rice D2K Lab, Rice University

Understanding how large populations of neurons communicate and jointly fire in the brain is a fundamental open question in neuroscience. Many approach this by estimating the intrinsic functional neuronal connectivity using probabilistic graphical models. But there remain major statistical and computational hurdles to estimating graphical models from new large-scale calcium imaging technologies and from huge projects which image up to one hundred thousand neurons in the active brain. In this talk, I will highlight a number of new graph learning strategies my group has developed to address many critical unsolved challenges arising with large-scale neuroscience data. Specifically, we will focus on Graph Quilting, in which we derive a method and theoretical guarantees for graph learning from non-simultaneously recorded and pairwise missing variables. We will also highlight theory and methods for graph learning with latent variables via thresholding, graph learning for spikey data via extreme graphical models, and computational approaches for graph learning with huge data via minipatch learning. Finally, we will demonstrate the utility of all approaches on synthetic data as well as real calcium imaging data for the task of estimating functional neuronal connectivity.

The Many Facets of the Strawderman Prior

April 01
9:00am

Jim Berger

Arts and Sciences Distinguished Professor Emeritus of Statistics, Duke University

In 1971, Bill Strawderman published an article solving an interesting outstanding problem in mathematical statistics: do proper prior Bayes estimators for a multivariate normal mean exist, which are also minimax in a frequentist sense. He showed that they do exist. To carry out the proof, he invented a prior for a normal mean that had a closed form posterior mean. The minimax properties of his prior are no longer viewed as unusual, but the fact that the prior yields closed form estimates (and closed form posterior distributions) has had a profound effect on many areas of statistics, and other sciences such as astrophysics. These other facets of the Strawderman prior are explored in this talk.

How Do Noise Tails Impact on Deep ReLU Networks?

April 02
11:00am

Jianqing Fan

Professor of Statistics, Frederick L. Moore '18 Professor of Finance, Princeton University

This paper investigates the stability of deep ReLU neural networks for nonparametric regression under the assumption that the noise has only a finite p-th moment. We unveil how the optimal rate of convergence depends on p, the degree of smoothness, and the intrinsic dimension in a class of nonparametric regression functions with hierarchical composition structure when both the adaptive Huber loss and deep ReLU neural networks are used. This optimal rate of convergence cannot be obtained by the ordinary least squares but can be achieved by the Huber loss with a properly chosen parameter that adapts to the sample size, smoothness and moment parameters. A concentration inequality for the adaptive Huber ReLU neural network estimators with allowable optimization errors is also derived. To establish a matching lower bound within the class of neural network estimators using the Huber loss, we employ a different strategy from the traditional route: constructing a deep ReLU network estimator that has a better empirical loss than the true function and the difference between these two functions furnishes a low bound. This step is related to the Huberization bias, yet more critically to the approximability of deep ReLU networks. As a result, we also contribute some new results on the approximation theory of deep ReLU neural networks.
(Joint with Yihong Gu and Wen-Xin Zhou)

Data Based Loss Estimation of the Mean of a Spherical Distribution with a Residual Vector

April 01
1:40pm

Dominique Fourdrinier

Professor, Université de Rouen, France

In the canonical setting of the general linear model, we are concerned with estimating the loss of a point estimator when sampling from a spherically symmetric distribution. More precisely, from an observable (X, U) in $\mathbb{R}^p \times \mathbb{R}^k$ having a density of the form $(1/\sigma^n) f(\|x - \theta\|^2 + \|u\|^2/\sigma^2)$ where θ and σ are both unknown, we consider general estimators $\varphi(X, \|U\|^2)$ of θ under two losses: the usual quadratic loss $\|\varphi(X, \|U\|^2) - \theta\|^2$ and the data-based loss $\|\varphi(X, \|U\|^2) - \theta\|^2/\|U\|^2$. Then, for each loss, we compare, through a squared error risk, their unbiased loss estimator $\delta_0(X, \|U\|^2)$ with a general alternative loss estimator $\delta(X, \|U\|^2)$. Thanks to the new Stein type identity in Fourdrinier and Strawderman [2015], we provide an unbiased estimator of the risk difference between $\delta(X, \|U\|^2)$ and $\delta_0(X, \|U\|^2)$, which gives rise to a sufficient domination condition of $\delta(X, \|U\|^2)$ over $\delta_0(X, \|U\|^2)$. Minimax estimators of Baranchik form illustrate the theory. It is found that, as the class of improved Baranchik type estimators is larger under the data-based loss than under the usual quadratic loss, so is the class of improved loss estimators. Besides, the distributional assumptions and the dimensional conditions on the residual vector U are weaker when the data-based loss is used.

From Minimax Shrinkage Estimation to Minimax Shrinkage Prediction

April 02
9:00am

Ed George

Universal Furniture Professor Emeritus of Statistics and Data Science, University of Pennsylvania

In a remarkable series of papers beginning in the 1950's, Larry Brown, Charles Stein and Bill Strawderman among others, set the stage for the future development of minimax shrinkage estimation of a multivariate normal mean under quadratic risk. More recently, parallel developments have seen the emergence of minimax shrinkage estimation of multivariate normal predictive densities under Kullback–Leibler risk. In this talk, we reflect on these parallels emphasizing the central role played by the marginal distributions of Bayes procedures in both settings, determining the nature of the shrinkage, the superharmonic conditions for minimaxity, generalizations for minimax multiple shrinkage, and the establishing of complete class theorems.

Functional-Input Gaussian Processes with Applications to Inverse Scattering Problems

April 01
3:40pm

Ying Hung

Professor, Rutgers University

Surrogate modeling based on Gaussian processes (GPs) has received increasing attention in the analysis of complex problems in science and engineering. Despite extensive studies on GP modeling, the developments for functional inputs are scarce. Motivated by an inverse scattering problem in which functional inputs representing the support and material properties of the scatterer are involved in the partial differential equations, a new class of kernel functions for functional inputs is introduced for GPs. Based on the proposed GP models, the asymptotic convergence properties of the resulting mean squared prediction errors are derived and the finite sample performance is demonstrated by numerical examples. In the application to inverse scattering, a surrogate model is constructed with functional inputs, which is crucial to recover the reflective index of an inhomogeneous isotropic scattering region of interest for a given far-field pattern.

Expectation Propagation in Mixed Models

April 01
9:40am

Iain Johnstone

Marjorie Mhoon Fair Professor in Quantitative Science, Professor of Statistics, Professor of Biomedical Data Science, Stanford University

Matt Wand and colleagues have recently shown that the machine learning technique of expectation propagation (EP) yields state of the art methods for estimating parameters in generalized linear mixed models. We review this work to set the stage for an open question: is it possible to show asymptotic efficiency for these new estimators? The problem becomes one of defining an appropriate objective function that captures the EP iteration and approximates maximum likelihood well enough to inherit its efficiency. We indicate how a landscape approach and a new definition of "EP free energy" can get the job done. Along the way there are even implications for the accuracy of the classical Laplace approximation. This is joint work with the late Peter Hall, Alan Huang, Song Mei, John Ormerod and Matt Wand.

The Search for Efficient Predictive Density Estimators

April 01
2:20pm

Éric Marchand

Professor, Université de Sherbrooke, Canada

This talk will address the estimation of predictive densities, Bayesian or otherwise, and their efficiency as measured by frequentist risk. For Kullback-Leibler, $\alpha-$divergence and integrated L_1 losses, we review several recent findings that bring into play improvements by scale expansion, as well as duality relationships with point estimation and point prediction problems. A range of models are studied and include multivariate normal with both known and unknown covariance structure, scale mixture of normals, mean mixture of normals including skew-normal distributions, Gamma, as well as models with restrictions on the parameter space.

Matrix Estimation by Singular Value Shrinkage

April 02
1:40pm

Takeru Matsuda

Statistical Mathematics Unit Leader, RIKEN Center for Brain Science, Japan

In the estimation of a normal mean vector under the quadratic loss, the maximum likelihood estimator (MLE) is inadmissible and dominated by shrinkage estimators (e.g. James–Stein estimator) when the dimension is greater than or equal to three (Stein's paradox). In particular, generalized Bayes estimators with respect to superharmonic priors (e.g. Stein's prior) are minimax and dominate MLE. Note that a function is said to be superharmonic if its average value on a supersphere is always not greater than its value at the center.

In this talk, I will introduce recent studies on generalizations of the above results to matrices. First, we develop a superharmonic prior for matrices that shrinks singular values, which can be viewed as a natural generalization of Stein's prior. This prior is motivated from

the Efron–Morris estimator, which is an extension of the James–Stein estimator to matrices. The generalized Bayes estimator with respect to this prior is minimax and dominates MLE under the Frobenius loss. In particular, since it shrinks to the space of low-rank matrices, it attains large risk reduction when the unknown matrix is close to low-rank (e.g. reduced-rank regression). Next, we construct a theory of shrinkage estimation under the "matrix quadratic loss", which is a matrix-valued loss function suitable for matrix estimation. A notion of "matrix superharmonicity" for matrix-variate functions is introduced and the generalized Bayes estimator with respect to a matrix superharmonic prior is shown to be minimax under the matrix quadratic loss. The matrix-variate improper t-priors are matrix superharmonic and this class includes the above generalization of Stein's prior. Applications include matrix completion and nonparametric estimation.

Estimation of the Inverse Scatter Matrix for a Scale Mixture of Wishart Matrices under Efron-Morris Type Losses

April 02
2:20pm

Fatiha Mezoued

Professor, Ecole Nationale Supérieure de Statistique et d'Économie Appliquée, Algeria

We consider estimation of the inverse scatter matrix Σ^{-1} for a scale mixture of Wishart matrices under various Efron-Morris type losses, $\mathrm{tr}[\{\hat{\Sigma}^{-1} - \Sigma^{-1}\}^2 S^k]$ for $k = 0, 1, 2, \ldots$, where S is the sample covariance matrix. We improve on the standard estimators aS^+, where S^+ denotes the Moore-Penrose inverse of S and a is a positive constant, through an unbiased estimator of the risk difference between the new estimators and aS^+. Thus we demonstrate that improvements over the standard estimators under a Wishart distribution can be extended under mixing. We give a unified treatment of the two cases where S is invertible ($S^+ = S^{-1}$) and where S is singular.

Bayesian Model Choice in Finite and Infinite Mixtures

April 02
9:40am

Christian P. Robert

Professor, Universitè Paris Dauphine PSL, France

Testing for the number of components in a finite mixture model or against the fit of a finite mixture model for a given dataset has long been and still is an issue of much interest, albeit yet missing a fully satisfactory resolution. Using a Bayes factor to find the right number of components K in a finite mixture model is known to provide a consistent procedure. We furthermore establish the consistence of the Bayes factor when comparing a parametric family of finite mixtures against the nonparametric location Dirichlet Process Mixture (DPM) model. In practice, estimating the model evidence (1) is unfortunately a notoriously difficult task for finite and infinte mixture models and we reexamine here different Monte Carlo techniques advocated in the recent literature, as well as novel approaches based on Geyer's (1994) reverse logistic technique and sequential Monte Carlo (SMC).

Heterogeneous Data and Objective Priors

April 01
4:20pm

Andrew L. Rukhin

National Institute of Standards and Technology, USA

Challlenging statistical issues arise in collaborative studies when within-study uncertainties are unreliable or not reported. A Bayesian model with a non-informative prior for a common mean of several independent,h but not identically distributed observations is considered. The discrete posterior distribution which also appears in approximation theory, is shown to converge to a normal law but at a rate slower than square root of the number of observations. In the random effects setting, when the variance estimates cannot be trusted, it is suggested to use them only as the lower bounds for true uncertainties under non-informative prior. The model with a number of different clusters each formed by the cases with the same heterogeneity variance is introduced; the selection of such a model is discussed.

Robust Q-learning

April 02
4:20pm

Robert Strawderman

Donald M. Foster, MD Distinguished Professor of Biostatistics , University of Rochester

Q-learning is a regression-based approach that is widely used to formalize the development of an optimal dynamic treatment strategy. Finite dimensional working models are typically used to estimate certain nuisance parameters, and misspecification of these working models can result in residual confounding and/or significant efficiency loss. Using Robinson's transformation, we propose a robust Q-learning approach that allows key nuisance parameters to be estimated using data-adaptive techniques. Methodology, asymptotics and simulations will be summarized and highlight the utility of the proposed methods in practice. Time permitting, data from the "Extending Treatment Effectiveness of Naltrexone" multistage randomized trial will be used to illustrate the proposed methods. This is joint work with Ashkan Ertefaie.

Causal Inference in the Time of Covid-19

April 02
3:40pm

Larry Wasserman

UPMC Professor of Statistics and Data Science, Carnegie Mellon University

I will discuss our recent work on developing causal models for estimating the effect of social mobility on deaths from Covid-19. We propose a semiparametric marginal structural model motivated by an epidemic model. We estimate the counterfactual time series of deaths under interventions on mobility. We conduct several types of sensitivity analyses. We find that the data support the idea that reduced mobility causes reduced deaths, but the conclusion comes with caveats. There is evidence of sensitivity to model misspecification and unmeasured confounding which implies that the size of the causal effect needs to be interpreted with caution. While there is little doubt the effect is real, our work highlights the challenges in drawing causal inferences from pandemic data. This is joint work with Matteo Bonvini, Edward Kennedy and Valerie Ventura.

On Graphical Models and Convex Geometry

April 01
11:40am

Martin Wells

Department Chair and Charles A. Alexander Professor of Statistical Sciences, Cornell University

We introduce a mixture-model of beta distributions to identify significant correlations among P predictors when P is large. The method relies on theorems in convex geometry, which we use to show how to control the error rate of edge detection in graphical models. Our 'betaMix' method does not require any assumptions about the network structure, nor does it assume that the network is sparse. The results hold for a wide class of data generating distributions that include light-tailed and heavy-tailed spherically symmetric distributions.

Network Community Detection: New Algorithms and Goodness-of-Fit Tests

April 02
11:40am

Emma Zhang

Associate Professor, University of Miami

One of the fundamental problems in network data analysis is community detection that aims to partition nodes into cohesive communities. The stochastic block model, along with its variants, is one of the most studied statistical models for this purpose. Directly fitting the stochastic block model likelihood function on large-scale networks is known to be challenging. In this talk, I will discuss a pseudo likelihood approach that uses a new idea of "label decoupling" that permits an alternating maximization and can efficiently handle up to millions of nodes. The proposed method has provable convergence guarantee and enjoys strong consistency. I will also briefly discuss my work on testing for the number of communities in a stochastic block model and finally illustrate the usefulness of our methods through an analysis of international trade data.

Author Index