samsi

NSF·Duke·NCSU·UNC

MUMS Program

Sixth Bayesian, Fiducial, and Frequentist (BFF6) Conference on Model Uncertainty

April 28-May 1, 2019

SPEAKER TITLES/ABSTRACTS

Pierre Barbillon
UMR MIA-Paris, AgroParisTech, INRA

"Variable Selection in the Discrepancy Function Associated with a Simulator"

A statistical model which links field experiments with a simulator usually embeds a discrepancy function. The discrepancy function models the systematic gap between the simulator and the real system. Analyzing the discrepancy should help to understand to what extent the simulator is reliable. In particular, determining that some variables are active or inert in the discrepancy function is of major interest since it indicates which variables are correctly modeled or not by the simulator. Therefore, this could give some leads to improve the simulator and help to determine if extrapolation is safe or not with respect to a specific input.

The discrepancy function is modeled as a Gaussian process which is parametrized as in Linkletter et al. (2006). This parametrization provides a simple distinction between active and inert variables. The variable selection is performed through a model selection where the models in competition differ on the prior distribution considered for the parameter associated with the variables in the Gaussian process.

We resort to computations of Bayes factors by using a bridge sampling to perform the model selection. Contrasted synthetic examples are considered to support the proposed technique.

Co-authors: Rui Paulo, Anabel Forte

David Bickel
University of Toronto

"Blends of Bayesian and Frequentist Inference"

Blended inference draws on strengths of the frequentist and Bayesian theories of statistics. In one recipe, Bayesian inference is used when the prior distribution is known, frequentist inference when nothing is known about the prior, and a blend of both when there is some knowledge about the prior.

If the confidence distribution that corresponds to frequentist inference is practically consistent with the set of posterior distributions derived from the set of priors, then the confidence distribution alone is enough for inference. If not, it exerts a strong influence on inferences drawn on the basis of the set of posterior distributions.

That blend is illustrated for the hypothesis testing case by a simple combination of a confidence distribution with a set of plausible posterior probabilities.

Alisa Bokulich
Boston University

"Models as Tools not Mirrors: Crossover Themes from the Philosophy of Science"

In this talk I draw out some of the philosophical implications of understanding models as targeted tools, rather than mirrors of nature. In particular, I want to highlight a number of potential crossover themes from the philosophy of science, relevant to thinking about model uncertainty. The first theme relates to model evaluation: understanding models as tools suggests that models should not be evaluated as "true" but rather as "adequate for particular purposes." The second theme directs our attention to the importance of the underlying conceptual model, and the representational choices made when conceptualizing the phenomenon in the world to be investigated or explained. The third theme examines reduced complexity modeling and argues that more is not always better. Finally, a fourth theme defends model pluralism and a division of cognitive labor among models. Although these points will be briefly illustrated with examples from across the sciences, my focus will be on the proverbial forest, and the hope is that these crossover themes will facilitate building further bridges between the philosophy of science and MUMS literatures.

Philip Dawid
Cambridge University

"Can a Fiducial Phoenix Rise from the Ashes?"

Fiducial inference has a long history, but was not taken seriously by most statisticians until a recent resurgence of interest. Can new life be breathed into its corpse? I will reprise some ancient analyses aiming to discover just how far fiducial logic can be taken before it collapses under its own weight.

Anabel Forte
University of Valencia

"Model Uncertainty: a review"

Nowadays the number of models trying to explain reality is huge. These models can be as "simple" as a linear regression or as complex as a mathematical model which take hours or days to run and give a single output. But from statistical to mathematical models they all have a common feature (quoting Box): All models are wrong, some of them are useful.

Statistics plays a very important role in this scenario trying to fill the gap between the models and reality. On the one hand, it is important to correctly measure the uncertainty so the unknown inputs (or parameters) of the model can be properly calibrated. On the other hand, it is crucial, in the presence of several models, to understand which one is closer to the truth.

During this tutorial I will introduce the problem pointing out to the main issues as well as showing some of the approaches that have been developed to solve them.

Gonzalo Garcia-donato
Universidad de Castilla-La Mancha

"Including Factors in Bayesian Variable Selection Problems"

Factors are categorical variables. The sums of the values of these variables are called levels. In this talk, we consider the variable selection problem where the set of potential predictors contains both factors and numerical variables. Formally, this problem is a particular case of the standard variable selection problema, where factors are coded using dummy variables. As such, the Bayesian solution would be straightforward and, possibly because of this, the problem. Despite its importance, this issue has not received much attention in the literature. Nevertheless, we show that this perception is illusory and that in fact several inputs, like the assignment of prior probabilities over the model space or the parameterization adopted for factors may have a large (and difficult to anticipate) impact on the results. We provide a solution to these issues that extends the proposals in the standard variable selection problem and does not depend on how the factors are coded using dummy variables. Our approach is illustrated with a real example concerning a childhood obesity study in Spain.

Authors: Gonzalo Garcia-donato and Rui Paulo

Edward George
University of Pennsylvania

"Multidimensional Monotonicity Discovery with MBART"

For the discovery of a regression relationship between y and x, a vector of p potential predictors, the flexible nonparametric nature of BART (Bayesian Additive Regression Trees) allows for a much richer set of possibilities than restrictive parametric approaches. To exploit the potential monotonicity of the predictors, we introduce mBART, a constrained version of BART that incorporates monotonicity with a multivariate basis of monotone trees, thereby avoiding the further confines of a full parametric form. Using mBART to estimate such effects yields (i) function estimates that are smoother and more interpretable, (ii) better out-of-sample predictive performance and (iii) less post-data uncertainty. By using mBART to simultaneously estimate both the increasing and the decreasing regions of a predictor, mBART opens up a new approach to the discovery and estimation of the decomposition of a function into its monotone components. (This is joint work with H. Chipman, R. McCulloch and T. Shively).

Subhashis Ghosal
North Carolina State University

"Coverage of Credible Intervals for Monotone Regression"

Shape restrictions such as monotonicity often naturally arise. In this talk, we consider a Bayesian approach to monotone nonparametric regression with a normal error. We assign a prior through piecewise constant functions and impose a conjugate normal prior on the coefficient. Since the resulting functions need not be monotone, we project samples from the posterior on the allowed parameter space to construct a "projection posterior". We obtain the limit posterior distribution of a suitably centered and scaled posterior distribution for the function value at a point. The limit distribution has some interesting similarity and difference with the corresponding limit distribution for the maximum likelihood estimator. By comparing the quantiles of these two distributions, we observe an interesting new phenomenon that coverage of a credible interval can be more than the credibility level, the exact opposite of a phenomenon observed by Cox for smooth regression. We describe a recalibration strategy to modify the credible interval to meet the correct level of coverage.

This talk is based on joint work with Moumita Chakraborty, a doctoral student at North Carolina State University.

Mengyang Gu
Johns Hopkins University

“Generalized Probabilistic Principal Component Analysis of Correlated Data”

Abstract: Principal component analysis (PCA) is a well-established tool in machine learning and data processing. \cite{tipping1999probabilistic} proposed a probabilistic formulation of PCA (PPCA) by showing that the principal axes in PCA are equivalent to the maximum marginal likelihood estimator of the factor loading matrix in a latent factor model for the observed data, assuming that the latent factors are independently distributed as standard normal distributions. However, the independence assumption may be unrealistic for many scenarios such as modeling multiple time series, spatial processes, and functional data, where the output variables are correlated.

In this paper, we introduce the generalized probabilistic principal component analysis (GPPCA) to study the latent factor model of multiple correlated outcomes, where each factor is modeled by a Gaussian process. The proposed method provides a probabilistic solution of the latent factor model with the scalable computation. In particular, we derive the maximum marginal likelihood estimator of the factor loading matrix and the predictive distribution of the output. Based on the explicit expression of the precision matrix in the marginal likelihood, the number of the computational operations is linear to the number of output variables. Moreover, with the use of the Mat{Ã©}rn covariance function, the number of the computational operations is also linear to the number of time points for modeling the multiple time series without any approximation to the likelihood function. We discuss the connection of the GPPCA with other approaches such as the PCA and PPCA, and highlight the advantage of GPPCA in terms of the practical relevance, estimation accuracy and computational convenience. Numerical studies confirm the excellent finite-sample performance of the proposed approach.

Jan Hannig
University of North Carolina

“Are Reported Likelihood Ratios Well Calibrated?

Many computer programs and software systems used in the interpretation of forensic evidence have as their output Bayes factors also commonly referred to as likelihood ratios. For example, it is not unusual to see it reported that the DNA recovered at the crime scene is a million times more likely under the assumption that the defendant is a contributor to the crime stain than under the assumption that the defendant is not a contributor. In this talk we summarize existing approaches for examining the validity of likelihood ratio systems and discuss a new statistical methodology, based on generalized fiducial inference, for empirically examining the validity of such likelihood ratio assessments. We illustrate our approach by examining LR values calculated with one or more widely available data sets.

Joint work with Hari Iyer at National Institute of Standards and Technology

Leah Henderson
University of Groningen

"Unifying Ockham's Razor"

In science, there is a preference for simpler theories. Simplicity emerges as a criterion for theory choice as a result of the attempt to determine which of the theories is getting the basics right. Model selection methods provide a formal explication of this procedure. More detailed fitting of the theory to the data generally follows. There has been an unfortunate tendency among philosophers to pit different model selection methods against one another and find fault with some as opposed to others. I argue for a more ecumenical approach, which sees the different methods as delivering a more unified explication of features which philosophy of science has described in qualitative terms. According to this point of view, the different methods are better seen as as complementary, rather than competing.

Michael I. Jordan
University of California, Berkeley

"Inference Meets Computation: Dynamical, Stochastic and Economic Perspectives"

Statisticians have generally been content to leave considerations of computation out of discussions of the foundations of inference. Although the roots of the two dominant paradigms of statistical computation---optimization and MCMC sampling---were in the mix in foundational discussions in the 1950s, the mathematical study of these topics diverged and accordingly failed to provide a forcing function to bring warring statisticians to the same table. Indeed, in the ensuing years, when implementing their principles computationally, frequentists have mostly been able to focus on optimization and Bayesians on MCMC, and the resulting implementations do not hint at any deeper inferential unity. I wish to complicate matters by developing some new mathematical links between optimization and sampling. These links appear most readily when one works in continuous time, as variational characterizations of certain classes of ODEs and SDEs, and they are most striking in the setting of nonconvex geometry.

Kevin Kelly
Carnegie Mellon University

Konstantin Genin
University of Toronto

"A New, Truth-directed Explanation of Ockham's Razor in Model Inference"

Science aims at true models and theories that stand up in counterfactual situations not yet sampled. All parties agree that science should prefer simpler theories compatible with experience to complex ones. But why? Bayesians explain Ockham's razor with a prior probabilistic bias against complex possibilities that look simple, but that does not begin to explain how such a bias conduces to finding the true model. Frequentists punt on the question entirely, because there is no bound on chance of error for model inference. We propose a new frequentist explanation of how, and in what sense, Ockham's razor helps one find the true model in science. At best, a statistical method can converge to the true model in chance (pointwise consistency). But convergence in chance is not even remotely monotonic---the chance of producing a model may drop precipitously with sample size. When the drop in chance is at least alpha, say that the model has been alpha-retracted. Paying more for a larger sample that eliminates a false conclusion sounds like progress, but paying more to reject the true conclusion is retrograde. Say that a method is alpha-progressive if it is pointwise consistent and never retracts the true hypothesis by more than alpha. A method is alpha-Ockham iff its chance of producing a model more complex than the truth (at an arbitrary parameter) is at most alpha. Think of that as a refined way to converge to the truth. Our main result is that only alpha-Ockham methods are alpha-progressive. As alpha is tuned downward, the method will continue to favor simple theories at more complex parameter settings, which provides a new interpretation of the statistical tradeoff between simplicity and fit. Also, when statistical tests at nominal significance alpha are used to fish for models, as in the discovery of causal DAGs, the significance level can be interpreted rigorously as a bound on alpha progressiveness. Finally, there is the prospect of a new, frequentist foundation for objective simplicity biases in Bayesian prior probabilities of the sort recommended by Harold Jeffreys---what is the least prior bias toward simple models that guarantees alpha progressiveness of Bayesian inference?

Regine Liu
Rutgers University

Min-ge Xie
Rutgers University

“Fusion Learning and BFF Approaches”

Inferences from different data sources can often be fused together, referred to as `fusion learning', to yield more powerful findings than those derived from individual data sources alone. Effective fusion learning approaches are in growing demand as increasing number of data sources have become easily available in this big data era. Fusion learning broadens the scope of traditional meta-analysis, whose goal is also combining findings from different studies (often in the analysis of clinical trials), to a wider range of problem settings and with greater emphases on both the efficiency and computational feasibility of the methods used. This short course is designed to: 1) review existing fusion learning approaches derived from BFF concepts, especially the approach derived from the so-called confidence distribution (CD); and 2) show that BFF can help develop versatile tools for fusion learning in many challenging problem settings, including but not limited to: i) fusion learning from independent studies, ii) fusion learning for zero (or near zero) counts in discrete data, iii) robust and nonparametric fusion learning methods, iv) fusion learning in split-conquer-combine approach to efficient and feasible analyses of big data, v) individualized fusion (iFusion) learning to enhancing inference for individual studies by borrowing useful information from similar studies. The goal is to show that BFF developments can help provide efficient and broadly applicable approaches for fusion learning.

Peter McCullagh
University of Chicago

“Statistical Sparsity”

The talk is concerned with a particular definition of statistical sparsity as a stochastic limit. The limit definition is satisfied by every example that has been proposed in the literature on sparse signal detection, so, in that sense it is uncontroversial. Nonetheless, the definition has implications for sparse signal detection. For example, it puts very specific limits on the types of inferential questions (integrals or conditional expectations) that we can hope to address. It also implies that certain pairs of sparse models are first-order equivalent, or effectively equivalent.

(Joint work with Nick Polson)

Rui Paulo
Universidade de Lisboa

"Model Selection in the Context of Computer Models"

We approach the screening problem - i.e. detecting which inputs of a computer model significantly impact the output - from a formal Bayesian model selection point of view. That is, we place a Gaussian process prior on the computer model and consider the 2^p models that result from assuming that each of the subsets of the p inputs affect the response. The goal is to obtain the posterior probabilities of each of these models. In this talk, we focus on the specification of objective priors on the model-specific parameters and on convenient ways to compute the associated marginal likelihoods. These two problems that normally are seen as unrelated, have challenging connections since the priors proposed in the literature are specifically designed to have posterior modes in the boundary of the parameter space, hence precluding the application of approximate integration techniques based on e.g. Laplace approximations. We explore several ways of circumventing this difficulty, comparing different methodologies with synthetic examples taken from the literature.

Authors: Gonzalo Garcia-Donato (Universidad de Castilla-La Mancha) and Rui Paulo (Universidade de Lisboa)

Veronika Rockova
University of Chicago

"Multiscale Analysis of BART"

The widespread popularity of Bayesian tree-structured regression methods has raised considerable interest in theoretical understanding of their empirical success. However, theoretical literature on methods such as Bayesian CART and BART is still in its infancy. This paper affords new insights about Bayesian CART in the context of structured wavelet shrinkage under the white noise model. We exhibit precise connections between tree-shaped sparsity priors and unstructured spike-and-slab priors, which are regarded as ideal but are rather theoretical in nature. We show that the more practical Bayesian CART priors lead to adaptive rate-minimax posterior concentration in the l_∞ sense, performing nearly as well as the theoretical ideal (up to a log term). To further explore the benefits of structured shrinkage, we propose the g-prior for trees, which departs from the typical wavelet product priors by harnessing correlation structure induced by the tree topology. While the majority of wavelet type theoretical results for CART focus on dyadic trees, here we do not require that splits are at dyadic locations. We introduce the library of weakly balanced Haar wavelets and show that Bayesian CART is equivalent to Bayesian basis selection from this library. To illustrate that l_∞ adaptation is an intricate phenomenon, where internal sparsity plays a key role, we show that dense trees are incapable of adaptation. While one of the major appeals of BART is uncertainty quantification via credible sets, asymptotic normality justifications have thus far been unavailable. Building on the l_∞ adaptation property, we provide new fully non-parametric and adaptive Bernstein-von Mises statements for Bayesian CART using multiscale techniques. (Joint work with Ismael Castillo)

Chiara Sabatti
Stanford University

"Selecting Important Features in Presence of Correlation—a story from Genetics"

A research goal in genetics is to identify which DNA variants influence traits of medical interest. In Genome Wide Association Studies (GWAS), hundreds of thousand polymorphisms are genotyped and statistical approaches are used to identify places in the genome where DNA variation is associated to phenotypic variation, as well as to identify those variants that are more likely to have causal effects. While the ultimate scientific goal is well defined, its translation in the language of statistical inference has been somewhat challenging. On the one hand, in order to rely on available statistical methodology, researchers have often tested hypotheses that are only "approximation" of those of interest. On the other hand, the presence of strong dependence among groups of the studied polymorphisms poses specific challenges in describing interpretable hypotheses that can be investigated with substantial power.

In work with Matteo Sesia, Eugene Katsevich and Emmanuel Candes we tackle this problem aiming to make interpretable and precise findings, using the Knockoffs framework to control the rate of false discoveries. The lessons we learn have applications outside GWAS.

John Snyder
Bayer Crop Science

"Objective Bayesian Analysis for a 2 x 2 Contingency Table"

In categorical data analysis, the odds ratio is an important approach to quantify the strength of association between two variables in a contingency table. Here, we present a novel Bayesian approach to analyze an unrestricted 2x2 table along with several constructed nuisance parameters using objective Bayesian methods. The prior for the odds ratio has many desirable properties such as propriety, symmetry and finite moments on log scale, and others. Simulation results indicate that the proposed approach to this problem is far superior to the straightforward and widely used frequentist approaches that dominate this area as well as other objective candidates. Real data examples also typically yield more sensible results, especially for small sample sizes or for tables that contain zeros.

Dongchu Sun
University of Missouri

"Bayesian Analysis for Misaligned Regions and Applications in Cancer Mortality"

For many datasets, multiple variables measured on (possibly differing) areal units are available. We wish to simultaneously model both the spatial relations within each variable and the relations between variables. Unlike other approaches to constructing multivariate generalizations of the CAR model, we require that each marginal distribution is an ordinary CAR model. The method is based on transforming each variable to a variable that is marginally standard normal, but with a cross-covariance matrix between each pair proportional to a fixed region ``overlap matrix" analogous to an adjacency matrix. This method allows the possibility that the geographies of each variable differ from each other, but at the disadvantage of a set of conditions to ensure positive definiteness that quickly become cumbersome as the number of variables increases. The methods are illustrated to fit Missouri cancer mortality datasets.

Stéphanie van der Pas
Leiden University

"Uncertainty Quantification for Bayesian Survival Analysis"

The Bayesian framework offers an intuitive approach towards uncertainty quantification. We present general Bernstein-von-Mises theorems for several survival objects, and discuss how these lead to relevant uncertainty quantification in practice.

Marina Vannucci
Rice University

"Spatially Informed Variable Selection Priors and Applications to Large-scale Data"

There is now a huge literature on Bayesian methods for variable selection that use spike-and-slab priors. Such methods, in particular, have been quite successful for applications in a variety of different fields. High-throughput genomics and neuroimaging are two of such examples. There, novel methodological questions are being generated, requiring the integration of different concepts, methods, tools and data types. These have in particular motivated the development of variable selection priors that go beyond the independence assumptions of a simple Bernoulli prior on the variable inclusion indicators. In this talk I will describe various prior constructions that incorporate information about structural dependencies among the variables. I will also address extensions of the models to the analysis of count data. I will motivate the development of the models using specific applications from neuroimaging and from studies that use microbiome data.

Vladimir Vovk
Royal Holloway, University of London

"Calibration of Probability Forecasts"

A popular view of probability forecasting is that its aim is to maximize the sharpness of predictive distributions subject to their calibration (Gneiting et al., 2003+). Informally, calibration is the agreement between the predictive distributions and the observations, and its most popular formalization is calibration in probability. Sharpness refers to the concentration of the predictive distributions and does not depend on the observations. In this talk I will focus on conformal prediction, which is a method for producing provably calibrated predictive distributions, in the sense of calibration in probability and under the assumption, standard in machine learning, that the observations are produced independently from the same distribution (the IID assumption). While calibration is automatic under the IID assumption, achieving sharpness requires careful design of conformal predictors. My plan is to state asymptotic and small-sample results about their sharpness, with and without the IID assumption.

Jingshen Wang
University of Michigan

"Inference on Treatment Effects after Model Selection"

Inferring cause-effect relationships between variables is of primary importance in many sciences. In this talk, I will discuss two approaches for making valid inference on treatment effects when a large number of covariates are present. The first approach is to perform model selection and then to deliver inference based on the selected model. If the inference is made ignoring the randomness of the model selection process, then there could be severe biases in estimating the parameters of interest. While the estimation bias in an under-fitted model is well understood, I will address a lesser known bias that arises from an over-fitted model. The over-fitting bias can be eliminated through data splitting at the cost of statistical efficiency, and I will propose a repeated data splitting approach to mitigate the efficiency loss. The second approach concerns the existing methods for debiased inference. I will show that the debiasing approach is an extension of OLS to highÂ dimensions,Â and that a careful bias analysis leads to an improvement to further control the bias. The comparison between these two approaches provides insights into their intrinsic bias-variance trade-off, and I will show that the debiasing approach may lose efficiency in observational studies.

Jonathan Williams
University of North Carolina

"The EAS Approach for Graphical Selection Consistency in Vector Autoregression Models"

As evidenced by various recent and significant papers within the frequentist literature, along with numerous applications in macroeconomics, genomics, and neuroscience, there has been substantial interest to understand the theoretical estimation properties of high-dimensional vector autoregression (VAR) models. To date, however, while Bayesian VAR models have been developed and explored empirically (primarily in the econometrics literature) there exist very few theoretical investigations of the repeated sampling properties for Bayesian VAR models in the literature. Despite this fact, Bayesian methodology can surely offer important contributions to the high-dimensional VAR model literature, beyond what could be developed in a frequentist framework. One notable such contribution is the construction of posterior distributions over the set of all relative model probabilities. This framework of posterior inference has been widely exploited over the last decade in the high-dimensional linear regression literature, and we anticipate it will see comparable success for high-dimensional VAR models in the near future. In this direction, we construct methodology via the ε-$admissible$ subsets (EAS) approach for posterior-like inference of relative model probabilities over all sets of active/inactive components of the VAR transition matrix. We provide a mathematical proof of $strong$ graphical selection consistency for the EAS approach for stable VAR(1) models, and demonstrate numerically that it is an effective strategy in high-dimensional settings. The EAS methodology is an entirely new perspective on model selection which was originally developed to effectively account for linear dependencies among subsets of covariates in the high-dimensional linear regression setting.