Abstracts

Infinite arms bandit: optimality via confidence bounds

Hock Peng Chan

National University of Singapore, Singapore

ABSTRACT

The infinite arms bandit problem was initiated by Berry et al. (1997). They derived a regret lower bound of all solutions for Bernoulli rewards with uniform priors, and proposed bandit strategies based on success runs, but which do not achieve this bound. Bonald and Proutiere (2013) showed that the lower bound was achieved by their two-target algorithm, and extended optimality to Bernoulli rewards with general priors. We propose here a confidence bound target (CBT) algorithm that achieves optimality for general, unknown reward distributions.

For each arm we require the mean and standard deviation of its rewards to compute a confidence bound. We play the arm with the smallest bound provided it is smaller than a target mean. If the bounds are all larger, then we play a new arm. We show how for a given prior, the target mean can be computed to achieve optimality. In the absence of information on the prior, the target mean can be determined empirically, and we show that the regret achieved is still comparable to the regret lower bound. Numerical studies show that CBT is versatile and outperforms its competitors.

Reversing the Stein effect

Sanjay Chaudhuri[a] and Michael D. Perlman[b]

[a] *National University of Singapore, Singapore*
[b] *University of Washington, USA*

ABSTRACT

The Reverse Stein Effect is identified and illustrated: A statistician who shrinks his/her data toward a point chosen without reliable knowledge about the underlying value of the parameter to be estimated but based instead upon the observed data will not be protected by the minimax property of shrinkage estimators such as that of James and Stein, but instead will likely incur a greater error than if shrinkage were not used.

Stein's concentration inequality for proving the Berry-Esseen theorem

Louis H. Y. Chen

National University of Singapore, Singapore

ABSTRACT

When I was a graduate student at Stanford University, my first encounter with Stein's method was reading a set of lecture notes given to me by Richard Shorrock, who was a fellow graduate student and a good friend of mine. In these notes, I found a proof of the Berry-Esseen theorem for sums of i.i.d. random variables using what we now called Stein's method. What struck me was Stein's proof and use of an inequality, which he called concentration inequality, to circumvent the discontinuity of the derivative of the solution of the Stein equation. I will describe how this inequality has influenced my research in the years that followed and discuss its extensions in different settings.

Joint subtree distributions under two evolutionary tree models

Kwok Pui Choi

National University of Singapore, Singapore

ABSTRACT

Phylogenetic trees are widely used in ecology and evolutionary biology to describe the relatedness among a group of genes or organisms. In addition to providing the basis for comparing and organizing evolutionary relationships, the shapes of the phylogenetic trees inform us about genealogical events such as speciation and extinction. One way to apply this information is to compare empirical tree shape indices with those predicted by the neutral models. Two commonly used models are the Yule-Harding-Kingman (YHK) model and the Proportional to Distinguishable Arrangements (PDA) model. Motivated by our interest to discriminate the YHK model from the PDA model, we discuss efficient computation of the joint distribution of the numbers of subtrees of various sizes under these two models. This talk is based on joint work with Taoyang Wu.

Divergence loss and the Stein phenomemon

Malay Ghosh

University of Florida, USA

ABSTRACT

Stein in his seminal 1956 paper showed the inadmissibility of the sample mean in three or higher dimensions under squared error loss. The paper led to a string of interesting further developments and multiple generalizations. Recent work shows that the Stein phenomenon continues to hold under a general divergence loss for both estimation as well as prediction. This talk will primarily be a review of these recent results. I will point out also the duality between estimation and prediction.

Single index models and non-linear compressed sensing in the non-Gaussian case via Stein type discrepancies

Larry Goldstein and Xiaohan Wei

University of Southern California, USA

ABSTRACT

Gaussian sensing is a ubiquitous assumption for single index models and one bit compressed sensing. In the non-Gaussian case, performance guarantees for estimates of unknowns in nonlinear single index and compressed non-Gaussian sensing models can be achieved through the use of distributional characteristics, or discrepancies, arising in Stein's method that are tailored to be sensitive to distances to normality. These discrepancies return the value of zero under Gaussian, or linear, sensing thus recovering bounds for the standard models. The use of these discrepancies improves previous results by relaxing conditions and tightening performance bounds. These discrepancies continue to be tractable to compute when Gaussian sensing is corrupted by either additive errors or mixing.

Space-time data, intrinsic stationarity and functional models

Tailen Hsing

University of Michigan, USA

ABSTRACT

The topic of functional time series has received some attention recently. This is timely as many applications involving space-time data can benefit from the functional-data perspective. In this talk, I will start off with the Argo data, which have fascinating features and are highly relevant for climate research. I will then turn to some extensions of stationarity in the context of functional data. The first is to adapt the notion of intrinsic random functions in spatial statistics, due to Matheron, to functional data. Such processes are stationary after suitable differencing, where the resulting stationary covariance is referred to as generalized covariance. A Bochner-type representation of the generalized covariance as well as preliminary results on inference will be presented. The second extension considers intrinsic stationarity in a local sense, viewed from the perspective of so-called tangent processes. Motivations of this work can be found from studying the multifractional Brownian motion.

From Stein to minimax predictive density estimation: the sparse normal means case

Iain Johnstone

Stanford University, USA

ABSTRACT

Starting with the 1956 paper, Stein's breakthroughs on high dimensional estimation inspired–and continue to influence–decades of work on shrinkage and sparsity. This talk illustrates the theme by first reviewing the striking parallels between Stein's canonical setting of estimation of a multivariate normal mean under quadratic loss and the problem of predictive density estimation under Kullback-Leibler loss.

In joint work with Gourab Mukherjee, we then focus on ℓ_0-sparse Gaussian sequence models and propose proper Bayes predictive density estimates that achieve asymptotic minimaxity in sparse models. One surprise is the existence of a phase transition in the future-to-past variance ratio r. For $r < r_0 = (\sqrt{5} - 1)/4$, the natural discrete prior ceases to be asymptotically optimal. Instead, for subcritical r, a 'bi-grid' prior with a central region of reduced grid spacing recovers asymptotic minimaxity.

A class of tractable predictive densities for curved exponential families

Fumiyasu Komaki

The University of Tokyo, Japan

ABSTRACT

We consider extended plugin densities for curved exponential families. Although Bayesian predictive densities are optimal under the Bayes risk with the Kullback-Leibler loss, It is often difficult to obtain their explicit forms. Extended plugin densities are tractable to obtain the explicit forms and better than plugin densities in many examples. Asymptotic theory of extended plugin densities based on shrinkage priors is investigated.

This talk is based on a joint work with Michiko Okudo.

Recent applications of Stein's method in machine learning

Qiang Liu

The University of Texas at Austin, USA

ABSTRACT

As a fundamental technique for approximating and bounding distances between probability measures, Stein's method has caught the attention in the machine learning community recently; some of the key ideas in Stein's method have been leveraged and extended for developing practical and efficient computational methods for learning and using large scale, intractable probabilistic models. We will give an overview of Stein's method in machine learning, focusing on a computable, kernel-based Stein discrepancy for approximating and evaluating (via goodness-of-fit test) distributions with intractable normalization constants, and a Stein variational gradient descent (SVGD) for finding particle-based approximation to intractable distributions that combines the advantages of Markov chain Monte Carlo (MCMC), variational inference and numerical quadrature methods.

Charles Stein, covariance matrix estimation and some memories from one of his students

Wei-Liem Loh

National University of Singapore, Singapore

ABSTRACT

The first part of my talk will be on Charles Stein and his work, focusing (but not exclusively) on covariance matrix estimation. The second part of my talk will be a brief (and biased) survey on the current state of covariance matrix estimation.

Ensemble minimaxity of James-Stein estimators

Yuzo Maruyama

The University of Tokyo, Japan

ABSTRACT

We consider estimation of a multivariate normal mean. Under heteroscedasticity, estimators shrinking more on the coordinates with larger variances, seem desirable. However, they are not necessarily ordinary minimax.

We show that such James-Stein type estimators can be ensemble minimax, minimax with respect to the ensemble risk, related to empirical Bayes perspective of Efron and Morris. We also consider estimation of a multivariate normal mean under homoscedasticity. A geometric-mean-based James-Stein estimator is shown to be ensemble minimax.

This is a joint work with Larry Brown and Ed George.

Singular value shrinkage prior: a matrix version of Stein's prior

Takeru Matsuda

The University of Tokyo, Japan

ABSTRACT

We develop singular value shrinkage priors for the mean matrix parameters in the matrix-variate normal model with known covariance matrices. Our priors are superharmonic and put more weight on matrices with smaller singular values. They are a natural generalization of Stein's prior. Bayes estimators and Bayesian predictive densities based on our priors are minimax and dominate those based on the uniform prior in finite samples. In particular, our priors work well when the true value of the parameter has low rank. We also develop an empirical Bayes algorithm for the matrix completion problem, which is motivated from the singular value shrinkage estimator by Efron and Morris.

Charles Stein: A personal view

David Siegmund

Stanford University, USA

ABSTRACT

During a long career Charles Stein published a small number of papers, all containing very original ideas. Some are systematic and thorough, while others contain the germ of ideas that stimulated future research, and some suggested a still unrealized program for research. In this talk I will discuss a few favorites (but avoiding those for which Stein is best known). Included will be (i) his 1949 abstract on fixed width confidence intervals for a normal mean, (ii) his 1956 paper posing the problem of adaptive estimation and testing,(iii) the Efron-Stein inequality, and his beautiful short criticism of the likelihood principle.