

Cincinnati Symposium on Probability Theory and Applications

March 20-23, 2009

Abstracts of talks

Printed: March 14, 2009

Rick Bradley

Indiana University

A strictly stationary, N -tuplewise independent counterexample to the central limit theorem (joint work with Alexander Pruss)

This is joint work with Alexander Pruss. For any given integer $N \geq 2$, there exists a strictly stationary, ergodic sequence $(X_k, k \in \mathbf{Z})$ of random variables with the following properties: (i) the random variable X_0 is uniformly distributed on the interval $[-3^{1/2}, 3^{1/2}]$ (and hence $EX_0 = 0$ and $EX_0^2 = 1$); (ii) for every choice of N distinct integers $k(1), k(2), \dots, k(N)$, the random variables $X_{k(1)}, X_{k(2)}, \dots, X_{k(N)}$ are independent; (iii) the random variables $|X_k|, k \in \mathbf{Z}$ are independent; and (iv) for every infinite set $Q \subset \mathbf{N}$, there exist an infinite set $T \subset Q$ and a nondegenerate, non-normal probability measure μ on \mathbf{R} such that $(X_1 + X_2 + \dots + X_n)/n^{1/2}$ converges to μ in distribution as $n \rightarrow \infty, n \in T$. This example is a modification of a somewhat similar, non-stationary, N -tuplewise independent, identically distributed counterexample in Pruss [PTRF 111 (1998) 323-332]. It complements the strictly stationary, pairwise independent counterexamples in Janson [Stochastics 23 (1988) 439-448], and the strictly stationary, triplewise independent counterexamples developed in Bradley [PTRF 81 (1989) 1-10 and Rocky Mountain J. Math. 37 (2007) 25-44].

Bruce Budowle

FBI Laboratory

Statistical inferences on Genetic Data from Challenged Samples: From Low Copy Number Human Identification to Attribution in Microbial Forensics

Genetic typing is a powerful means to establish identity in criminal cases where biological evidence is found at crime scenes, for paternity testing and inheritance matters, identification of victims in mass disasters, and identification of missing persons from human remains, as well as providing investigative leads in cases of bioterrorism and biocrime. Conveying the significance of a failure to exclude is an important part of the interpretation of the results of an analysis. The statistical approaches for human single source, mixture, and kinship analyses are based on principles of population genetics and sampling statistics and include issues such as Hardy-Weinberg

and linkage equilibrium expectations, correction for effects of population substructure, minimum allele frequencies, lineage-based counting methods, and population data sets. While the statistical approaches for inferences of the weight of evidence are well-established, new situations have arisen that require additional thought regarding placing significance on a result of failure to exclude. Two areas are Low Copy Number (LCN) typing and Microbial Forensics and the statistical issues regarding their applications will be raised and discussed.

LCN typing refers to the analysis of any sample that contains less than 200 pg of template DNA (~ 33 human diploid cells). Exaggerated stochastic sampling effects are observed when amplifying a small number of starting templates during the PCR. The result is that several phenomena can occur: a substantial imbalance of two alleles at a given heterozygous locus, allelic dropout, and/or increased stutter. With increased sensitivity of detection there also is a concomitant increased risk of contamination. The approach most widely used for the designation of an allele in a LCN sample requires the division of the sample into two or more aliquots and reporting only the alleles that are common in at least two replicates. However, the suppositions and degree of confidence surrounding such practices are not well-defined. Moreover, the approaches used may have some application for single-source samples but are far more complex for mixtures. Therefore, until better developed it may be better to use LCN typing for identification of missing persons and for investigative leads.

Microbial forensics is an evolving sub-discipline of forensic science for analysis of evidence from a bioterrorism act, biocrime, hoax or an inadvertent release for attribution purposes. Some practices will be similar to that applied to human DNA analyses, such as Quality Assurance, generation of databases, and conveying significance of a nucleic acid analysis result. However, most of the approaches for developing statistical inferences will be very different for microbial forensics because of the lineage-based markers (more like Y chromosome and mitochondrial DNA markers) and additional genetic inheritance characteristics, such horizontal gene transfer. The databases generated will be very different. Moreover, the power of discrimination (or ability to individualize) will almost always not be able to reach that enjoyed for human identity testing, instead a most recent common ancestor approach will be considered. Additional major limitations will be the unknown history of samples and host-pathogen interactions.

Sourav Chatterjee

UC Berkeley

A rigorous theory of chaos in disordered systems

Disordered systems are an important class of models in statistical mechanics, having the defining characteristic that the energy landscape is a fixed realization of a random field. Examples include various models of glasses and polymers. They also arise in other subjects, like fitness models in evolutionary biology. The ground state of a disordered system is the state with minimum energy. The system is said to be chaotic if a small perturbation of the energy landscape causes a drastic shift of the ground state. In this talk I will present a rigorous theory of chaos in disordered systems that confirms long-standing physics intuition about connections between chaos, anomalous fluctuations of the ground state energy, and the existence of multiple valleys in the energy landscape. Combining these results with mathematical tools like hypercontractivity, I will present a proof of the existence of chaos in directed polymers. This is the first rigorous proof of chaos in any nontrivial disordered system. Applications to other models like spin glasses, fitness models, and general Gaussian fields will also be discussed.

Ranajit Chakraborty

U. Cincinnati

Statistical and Probabilistic Issues in DNA Forensics – Current Paradigms

(no abstract available)

Arthur J. Eisenberg

U. of North Texas

New technologies of DNA Forensics and attendant statistical issues

(no abstract available)

Michail Gordin

Multiparameter Martingale Approximation and Baker Sequences

For a class of random Fields (which are Markov in some sense) we show how the solvability of a higher analogue of the Poisson equation implies a kind of martingale/coboundary representation. The latter can be used, like in the one-parameter setting, to reduce the central limit problem to the case of (multiparameter) martingale differences. While in general the choice of the definition of multiparameter martingale differences can be discussed, here we are led to a unique version of it compatible with the situation we are interested in. As an application (developed in collaboration with Michel Weber, IRMA, Strasbourg) we consider an alternative approach to a problem originated in metric number theory and treated before by means of different methods.

Abdelhamid Hassairi

Sfax University, Tunisia.

On the Dirichlet distribution on symmetric matrices

We extend to the Dirichlet distribution on symmetric matrices two remarkable characterizations of the real Dirichlet distribution. We then introduce a generalization of the matrix Dirichlet distribution which represents the multivariate version of the Connor and Mosimann generalized real Dirichlet distribution, and we give some properties of this generalized distribution. We also introduce the Riesz-Dirichlet distribution as an extension of the Wishart-Dirichlet distribution, and we show that some of its projections related to the Pierce decomposition are also Dirichlet.

Gerard Letac

Universite Paul Sabatier, Toulouse, France

Mean length of some random plane convex sets generated by the Brownian motion.

Denote by $C(1)$ the convex hull of the Brownian motion in the complex plane watched from time 0 to time 1. Let w be a primitive n th root of the unity. Denote by $C(n)$ the convex hull of the union of all $w^k C(1)$ for $k = 1, 2, \dots, n$. For instance, $C(2)$ is the symmetric convex hull of the above Brownian curve. The lecture will explain how to compute the mean perimeter of $C(n)$ for $n = 1, 2, 3, 4, 6$, and will consider also three other convex sets built with $C(1)$.

Elizabeth Meckes

Case Western Reserve

Quantitative asymptotics of graphical projection pursuit.

There is a large class of results which say that, for parametrized collection of random variables, a random variable from the collection behaves a certain way for “most” values of the parameter. A nice example of such a result is a theorem of Persi Diaconis and David Freedman, which roughly says that if you have a large collection of high-dimensional data points, most one-dimensional projections of the data will look Gaussian even if the data have no particular structure. In this talk, I’ll discuss a quantitative version of this result and its proof, which uses Stein’s method, the concentration of measure phenomenon, and the geometry of certain function spaces.

Stewart Ethier

Univ. Utah

Limit theorems for Parrondo’s paradox

That one can combine two losing games to form a winning game is known as Parrondo’s paradox. We establish a strong law of large numbers and a central limit theorem for the Parrondo player’s cumulative profit, both in a one-parameter family of profit-dependent games and in a two-parameter family of history-dependent games, with the winning game being either a random mixture or a nonrandom pattern of the two losing games. We give formulas for the mean and variance parameters of the central limit theorem in nearly all such scenarios; formulas for the mean permit an analysis of when the Parrondo effect is present. (Joint work with Jiyeon Lee of Yeungnam University, South Korea.)

Qi-Man Shao

Hong Kong University of Science and Technology

Normal and Non-normal Approximations via Stein’s Method

In this talk we shall first give a brief survey on the Stein method for normal approximation and then provide a general approach for normal and non-normal approximations. An application to Curie-Weiss model will be discussed.

Kavita Ramanan

Carnegie Mellon U.

A Unified Framework for Processes with Discontinuous Dynamics

Many deterministic and stochastic processes that arise naturally show discontinuities in their transition rates. These discontinuities may arise due to the fact that the processes are constrained to live within a domain, as in the case of reflected diffusions, or as a result of an abrupt change in the transition rates in the interior of the domain, a characteristic exhibited by many optimally controlled processes. We describe a general unified framework for the analysis of these processes, and illustrate the application of this theory to several problems of interest.

Emmanuel Rio

Versailles, France

A Bernstein inequality and moderate deviations for weakly dependent sequences

(joint work with Magda Peligrad and Florence Merlevède)

In this talk we present a tail inequality for the maximum of partial sums of a weakly dependent sequence of random variables that is not necessarily bounded. The class considered includes geometrically and subgeometrically strongly mixing sequences. The result is then used to derive asymptotic moderate deviation results. Applications include classes of Markov chains, functions of linear processes with absolutely regular innovations and ARCH models.

Sunder Sethuraman

Iowa State

Large deviations for a tagged particle in one dimensional symmetric simple exclusion

The exclusion process follows a collection of random walks on a lattice which interact in that jumps to already occupied vertices are suppressed. Of interest is the behavior of a distinguished, or tagged particle in this system.

Recently, starting from certain non-equilibrium initial distributions, a law of large numbers and a central limit theorem, has been proved for a tagged particle in the exclusion model on \mathbb{Z} with symmetric nearest-neighbor jump probabilities (Jara-Landim '06). In this talk, we discuss a corresponding large deviation principle, and evaluate the rate function in certain regimes, showing a phase transition. Joint work with SRS Varadhan.

Stan Szarek

Case Western Reserve University

Concentration for non-commutative polynomials in Wigner ensembles

The last dozen or so years dramatically improved our understanding of concentration of various functionals of random matrices. Still, a number of unresolved questions remains. One difficult issue is generalizing facts which are known for Gaussian matrices to general subgaussian ensembles or to Wigner ensembles. We sketch an argument, in the spirit of Talagrand-Guionnet-Zeitouni, showing exponential concentration for the trace of (non-commutative) polynomials in Wigner-type random matrices.

Srinivasa Varadhan

Courant Institute

Large Deviations in a random environment

We will consider random walks on Z^d governed by transition probabilities that are themselves randomly chosen. They may remain fixed or may evolve over time in some prescribed fashion. The typical and atypical behavior of the random walk in this medium is of interest. The atypical behavior or "large deviations" can be due to many reasons. One of them is passing more often through advantageous environments and avoiding ones that are not so great. Understanding this behavior requires us to investigate the conditioned walk.

Tsewei Wang

U. Tennessee Knoxville

Pedigree-based Inference of Missing Person Identification from DNA Data

Using forensic DNA tools to identify human remains in mass disaster or missing person situations have shown to be feasible. When personal reference samples are not available, pedigree family members' DNA become the means by which identification can possibly be made. The more family members available for DNA typing, and the closer the relationship to the missing person, the more accurate the identification can be made. In this presentation, we present a pedigree based likelihood ratio approach in the identification of missing persons. We present simulation results of distributions of Pedigree-Likelihood-Ratios for pedigrees with various mix of family members to assess the relative power of discrimination in identifying the right remain to be the missing person. A ranked list of pedigree member mix is given, ranked with respect to their relative discrimination power in identifying the right missing person. In addition, we present some new results in probability study of total allele and shared allele count between pairs of relatives. Results of which can be used as an upfront tool to filter out the unlikely candidates for the missing person.

Michael Woodroffe
Coauthors: Dalibor Volny and Ou Zhao

Univ Michigan

The Conditional Central Limit Question For Superlinear Processes

Let \mathbb{Z} denote the integers and J a countable set; let F_j , $j \in J$, be distribution functions with means 0 and unit variances; let $\xi_{i,j}$, $i \in \mathbb{Z}$, $j \in J$, be independent random variables defined on a probability space (Ω, \mathcal{A}, P) ; and suppose $\xi_{i,j} \sim F_j$ are identically distributed for each j . Next let $c_{i,j}$ be a square summable array and consider processes of the form

$$X_k = \sum_{j \in J} \sum_{i=0}^{\infty} c_{i,j} \xi_{k-i,j} = \sum_{j \in J} \sum_{i \leq k} c_{k-i,j} \xi_{i,j}. \quad (1)$$

Here X_k converges *w.p.1* and in $L^2(P)$ by the Three Series Theorem and defines a strictly stationary process. We will call X_k a *superlinear process*, as it is the sum (superposition) of linear processes. Let \mathcal{F}_0 be the sigma algebra generated by $\xi_{i,j}$, $i \leq 0$, $j \in J$, $S_n = X_1 + \cdots + X_n$, $\sigma_n^2 = E(S_n^2)$, suppose that $\sigma_n \rightarrow \infty$, and let F_n be a regular conditional distribution function for S_n/σ_n given \mathcal{F}_0 ,

$$F_n(\omega; z) = P\left[\frac{S_n}{\sigma_n} \leq z | \mathcal{F}_0\right](\omega).$$

Necessary and sufficient conditions are given for the convergence in probability of F_n to the standard normal distribution Φ . The conditions resemble the Lindeberg Feller Conditions and involve an interesting interplay between the coefficients $c_{i,j}$ and the distribution functions F_j .

Wei Biao Wu

U. Chicago

Recursive Estimation of Time-Average Variance Constants

For statistical inference of means of stationary processes, one needs to estimate their time-average variance constants (TAVC) or long-run variances. For a stationary process, its TAVC is the sum of all its covariances and it is a multiple of the spectral density at zero. The classical TAVC estimate which is based on batched means does not allow recursive updates and the required memory complexity is $O(n)$. I will discuss a faster algorithm which recursively computes the TAVC, thus having memory complexity of order $O(1)$ and the computational complexity scales linearly in n . Under short-range dependence conditions, we establish moment and almost sure convergence of the recursive TAVC estimate. Convergence rates will also be discussed.