Graphical Models with Edge and Vertex Symmetreis

Lauritzen, Steffen; Højsgaard, Søren

Published in:
Scientific Program; Abstracts of the 7th World Congress in Probability and Statistics

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Lauritzen, S., & Højsgaard, S. (2008). Graphical Models with Edge and Vertex Symmetreis. In *Scientific Program; Abstracts of the 7th World Congress in Probability and Statistics* (pp. 135)

Abstracts

1 Fluid queues with input controlled by exponential timer

[CS 29,(page 26)]

Vijayakumar A., *Department of Mathematics Anna University Chennai, India*
Krishnakumar B., *Department of Mathematics Anna University Chennai, India*
SOPHIA, *Department of Mathematics Anna University Chennai, India*

Fluid queue models play a vital role in the performance analysis of computer and telecommunication networks. We study a fluid system with single on/off source in which the active (on) and silent periods (off) follow General and Exponential distributions respectively. During active periods, which are controlled by an exponential timer, there is a steady inflow into the fluid buffer at rate c0 and there is a steady outflow at rate c1 from the system when the buffer is non-empty. The system is analysed for different active periods corresponding to the service time of an M/G/1 queue and the busy period of an M/G/1 queue. The steady-state distribution of the buffer content and related performance measures are obtained. Special cases such as Erlang, exponential and deterministic service times are discussed. Some of the known results are deduced from the results obtained. Numerical results are illustrated for the models under consideration.

2 Pair-copula constructions of multiple dependence.

[IS 15,(page 5)]

Kjersti AAS, *Norwegian Computing Center, Norway*
Claudia CZADO, *Technische Universität, München, Germany*
Arnoldo FRIGESSI, *Statistics for Innovation, Norwegian Computing Center, and Department of Biostatistics, University of Oslo, Norway*
Henrik BAKKEN,

Building on the work of Bedford, Cooke and Joe, we show how multivariate data, which exhibit complex patterns of dependence in the tails, can be modelled using a cascade of pair-copulae, acting on two variables at a time. We use the pair-copula decomposition of a general multivariate distribution and propose a method to perform inference. The model construction is hierarchical in nature, the various levels corresponding to the incorporation of more variables in the conditioning sets, using pair-copulae as simple building blocks. Pair-copula decomposed models also represent a very flexible way to construct higher-dimensional copulae. We apply the methodology to a financial data set. Our approach represents the first step towards the development of an unsupervised algorithm that explores the space of possible pair-copula models, that also can be applied to huge data sets automatically.

3 Designing fuzzy time series model and its application to forecasting inflation rate

[CS 30,(page 27)]

Agus Maman ABADI, *Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, Yogyakarta State University, Indonesia*
SUBANAR, *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Indonesia*
WIDODO, *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Indonesia*
Samsubar SALEH, *Department of Economics, Faculty of Economics, Gadjah Mada University, Indonesia*

Fuzzy time series is a dynamic process with linguistic values as its observations. Modelling fuzzy time series developed by some researchers used discrete membership function and table lookup scheme methods from training data. Table lookup scheme is a simple method that can be used to overcome the conflicting rules by determining each rule degree. This paper presents new method to modelling fuzzy time series combining table lookup scheme and singular value decomposition methods which use continuous membership function. Table lookup scheme is used to construct fuzzy rules from training data and then singular value decomposition of firing strength matrix is used to reduce fuzzy rules. Furthermore, this method is applied to forecast inflation rate in Indonesia based on six-factors high-order fuzzy time series. This result is compared with neural network method and the proposed method gets a higher forecasting accuracy rate than the neural network method.

4 Robust design in the presence of non-normality and contaminated data

[CS 22,(page 22)]

Amani M. ABDULHALIM, *University Putra Malaysia (UPM)*

Kassim B. HARON, *University Putra Malaysia (UPM)*

This article concentrating on Robust design in the presence of non-normality and contaminated data. Three models are constructed for simulated data, we used mean and variance, sample median and median absolute deviation (MAD),sample median and interquartile range (IQR) in the first, second and third models. We tried to mix the Robust design modeling with optimization. Then compare the 3 models with a new model using Hodges Lehmann. We show that the second and third models are more resistant to both assumptions.

5 Constructing continuous model trinomial option pricing

[CS 24,(page 24)]

ABDURAKHMAN, *Dept. of Mathematics, Gadjah Mada University, Indonesia*

In this paper we study the Trinomial model for European option pricing theory using least squarehedge strategy in particular for call options as a kind of derivative securities. We use pseudoinverse matrix to obtain pseudoprobability. We find that risk neutral probability from Cox, Ross, Rubinstein (CRR) include in our pseudoprobability.

6 Modeling kurtosis and asymmetric data via maximum entropy distributions

[CS 38,(page 34)]

Sukru ACITAS, *Anadolu University*
Aladdin SHAMILOV, *Anadolu University, Science Faculty, Department of Statistics, Eskisehir, Turkey*
Ilhan USTA, *Anadolu University*
Yeliz MERT KANTAR, *Anadolu University*

Modeling kurtosis and asymmetry is crucial since data with skewness and kurtosis characteristics occurs in various areas such as statistics, economics, finance, biology, physics and other disciplines. In literature, generalized error (GE), Student t, skewed t, g and h, generalized beta of the second kind (GB2) and Pearson type IV distribution have been used to model these types data. These distributions may be satisfactory in many cases however in general, they are not enough to account for the kurtosis and asymmetry in the data. Recently, the maximum entropy (MaxEnt) distributions obtained from the MaxEnt method have been used to model asymmetry and kurtosis in the data, owing to the fact that the MaxEnt method based on Shannon entropy measure is a flexible and powerful tool for density estimation. Furthermore, the MaxEnt method covers normal, gamma, beta, GE, Student t and Pearson type IV distributions as special cases. In this study, the MaxEnt method and the MaxEnt distributions are introduced for data with skewness and kurtosis characteristics. Moreover, the MaxEnt distributions based on certain moment conditions are compared for modeling asymmetry and kurtosis via various statistical tests and criteria such as the Kolmogorov Smirnov test and the Akaike information criterion. Consequently, it is obtained that the MaxEnt distributions indicate a high degree of fitting in modeling kurtosis and skewed data.

7 Use of segment level prior hyperparameter in HB analysis in behavioral research.

[CS 59,(page 47)]

Atanu ADHIKARI, *ICFAI University*

Application of hierarchical Bayes techniques in researching consumer behavior is about couple of decades old. Authors have used hierarchical Bayes methodology in estimating unit level heterogeneity in parameter estimation. However, while using hierarchical Bayes, these researchers have used HB model considering single prior hyper-parameter in estimation process. Khatri and Rao (1992) showed that one hyper-parameter may be inadequate in individual parameter estimation if the population is heterogeneous. Population level prior hyper-parameter does not eliminate inter-segment variance which may severely affect individual estimate. Superiority of such estimates reduces if the heterogeneity in the population increases as the variability of the true mean value increases (Khatri and Rao, 1992). Since the intra-segment variability (nuisance parameter) decreases in formation of homogeneous groups within the heterogeneous population, the two stage prior distribution is supposed to give better parameter estimates than considering common prior distribution with same hyper-parameter.

This research uses choice based conjoint analysis through a multinomial logit model to estimate the parameter. Hierarchical Bayes method is used to capture unit level heterogeneity in parameter estimation. In this research, the researcher segments the sample in several homogeneous groups and considers mean values within the segment have common prior distribution with certain hyper-parameter which differs from segment to segment. The prior hyper-parameters of several such segments are assumed to have a common distribution with certain parameter at whole population level. The model is tested with simulated data and it is found that parameter estimates considering segment level hyper-parameters are significantly better than estimate using one hyper-parameter. A comparison between latent segment method and OLS method is also shown.

8 Kernel-wavelet estimation and empirical distribution of wavelet coefficients for regularity regression function

[PS 1,(page 4)]

Mahmoud AFSHARI, *Department of Statistics and Mathematics, School of Sciences, Persian Gulf University*

We consider the the kernel estimation of regression function r when the model is regularity and the observations are taken on the regular grid $x_i = \frac{i}{n}, i = 1, 2, .., n$. We propose kernel-estimation of regression function by algorithm of wavelet decomposition and empirical distribution of Häar wavelets coefficients are investigated.

9 Restricted Combined Ridge-Stein (Liu-type) Estimator in Semiparametric Regression Model

[CS 36,(page 33)]

Fikri AKDENIZ, *Cukurova University*
Esra AKDENIZ, *Gazi University*

In this paper, we consider the following semiparametric regression model

$$y = X\beta + f + \epsilon$$

We introduced a combined Ridge-Stein (CRS) estimation of a semiparametric regression model. Least squares for regression corresponds to minimizing the sum of squared deviations objective:

$$\|(I-S)y - (I-S)X\beta\|^2$$

adding to least squares objective, a penalized function of the squared norm

$$\|d\beta^* - \beta\|^2$$

for the vector of regression coefficients yields a conditional objective:

$$L = \arg\min_{\beta} \left\{ \|(I-S)y - (I-S)X\beta\|^2 + \|d\beta^* - \beta\|^2 \right\}$$

where S is a smoother matrix, which depends on smoothing parameter α. The first order condition of objective function minimized by the vector β is

$$\frac{\partial L}{\partial \beta} = 0.$$

The solution of this equation gives the CRS estimator of β in the semi-

parametric regression model. Firstly, the CRS estimators of both β and f are attained without a restrained design matrix. Secondly, the CRS estimator of β is compared with two-step estimator in terms of the mean square error.

We also discussed two different estimators β_1^* and β_2^*

$$\beta_1^* = \left[X'(I-S)X\right]^{-1} X'(I-S)y,$$
$$f_1^* = S(y - X\beta_1^*)$$

and

$$\beta_2^* = \left[X'(I-S)'(I-S)X\right]^{-1} X'(I-S)'(I-S)y,$$
$$f_2^* = S(y - X\beta_2^*)$$

respectively. Provided that $X'(I-S)X$ and $X'(I-S)'(I-S)X$ are invertible. Note that, there is small but subtle difference between β_1^* and β_2^* in the estimate for the parametric part of the semiparametric regression model.

We also established the estimators under $R\beta = r$ restrictions for the parametric component in the semiparametric regression model:

$b_{1r}^* = \beta_1^* + D^{-1}R'(RD^{-1}R')^{-1}(r - R\beta_1^*)$ with $D_1 = X'(I-S)X$, and $b_{2r}^* = \beta_2^* + Z_1^{-1}R'(RZ^{-1}R')^{-1}(r - R\beta_2^*)$ with $Z_1 = X'(I-S)'(I-S)X$.

Using the following CRC estimators, β_{1d}^* and β_{2d}^*

$$\beta_{1d}^* = \left[X'(I-S)X + I\right]^{-1}\left[X'(I-S)y + d\beta_1^*\right], \quad 0 < d < 1$$
$$\beta_{2d}^* = \left[X'(I-S)'(I-S)X + I\right]^{-1}\left[X'(I-S)'(I-S)y + d\beta_2^*\right]$$

in semiparametric regression model, we proposed the restricted CRS (Liu-type) estimators which are given below respectively:

$$b_{1rd}^* = \beta_{1d}^* + D_2^{-1}R'(RD_2^{-1}R')^{-1}(r - R\beta_{1d}^*),$$
$$b_{2rd}^* = \beta_{2d}^* + Z_2^{-1}R'(RZ_2^{-1}R')^{-1}(r - R\beta_{2d}^*).$$

where $D_2 = D_1 + I$ and $Z_2 = Z_1 + I$.

References

1. Akdeniz, F. and Kaçıranlar, S. (1995).On the almost unbiased generalized Liu estimator and unbiased estimation of the Bias and MSE, *Communications in Statistics – Theory and Methods* 24(7) 17891797.
2. Akdeniz, F. and Tabakan, G. (2008) Restricted ridge estimators of the parameters in semiparametric regression model. *Communications in Statistics –Theory and Methods*(Revised)
3. Liu, K. J. (1993).A new class of biased estimate in linear regression, *Communications in Statistics Theory and Methods* 22 393402.

10 The beta-Rayleigh distribution in reliability measure

[CS 12,(page 13)]

Alfred AKINSETE, *Marshall University*
Charles LOWE, *Marshall University*

The problem of estimating the reliability of components is of utmost importance in many areas of research, for example in medicine, engineering and control systems. If X represents a random strength capable of withstanding a random amount of stress Y in a component, the quantity $R = P(Y < X)$ measures the reliability of the component. In this work, we define and study the beta-Rayleigh distribution (BRD), and obtain a measure of reliability when both X and Y are beta-Rayleigh distributed random variables. Some properties of the BRD are discussed, including for example, the moments and parameter estimation. The BRD generalizes the reliability measures in literature.

11 An efficient computation of the generalized linear mixed models with correlated random effects

[CS 35,(page 33)]

Moudud ALAM, *Department of Economics and Social Sciences, Dalarna University and ESI, Orebro University*

This paper presents a two-step pseudo likelihood estimation technique for the generalized linear mixed models (GLMM), with random effects being correlated (possibly between subjects). Due to the use of the two-step estimation technique, the proposed algorithm outperforms the conventional pseudo likelihood algorithms *e.g.*, Wolfinger and O'Connell (1993, *Journal of Statistical Computation and Simulation 48*, 233-243), in terms of computational time. Moreover, it does not require any reparametrisation of the model such as Lindstrom and Bates (1989, *Journal of the American Statistical Association 43(404)*, 1014-1022). Multivariate Taylor's approximation has been used to approximate the intractable integrals in the likelihood function of the GLMM. Based on the analytical expression for the estimator of the covariance matrix of the random effects, a condition has been presented as to when such a covariance matrix can be estimated through the estimates of the random effects. An application of the estimation technique, with a binary response variable, is presented using a real data set on credit defaults.

12 Heuristic approaches to the political districting problem in Kuwait

[CS 79,(page 58)]

Shafiqah A. AL-AWADHI, *Kuwait University*
R. M'HALLAH, *Kuwait University*

This work models the political districting problem of Kuwait in search for cutting pattern that optimizes a set of criteria including population and voting equity, social, religious , ethnic and educational homogeneity, and geographical contiguity. The goal programming model which classifieds the criteria as hard and soft constraints is solved using tree-search based heuristic which combines integer programming and constraint propagation techniques. The heuristic takes the advantages of the orthogonal but complementary strengths of constraint and integer programming in stating and solving the district problem.

13 Dimension and measure of SLE on the boundary

[IS 29,(page 24)]

Tom ALBERTS, *Courant Institute of Mathematical Sciences, New York University*
Scott SHEFFIELD, *Courant Institute of Mathematical Sciences, New York University*

In the range $4 < \kappa < 8$, an SLE curve intersects the real line at a random set of points. The resulting set is sufficiently irregular to have a fractional dimension, which is now known to be $2 - 8/\kappa$. A primary focus of this talk will be describing how this dimension number is arrived at. The most technically demanding part is establishing an upper bound for the probability that an SLE curve hits two intervals on the real line, as the interval width goes to zero. I will present a proof of the correct upper bound by myself and Scott Sheffield, and I will also review work by

Oded Schramm and Wang Zhou that gives a similar bound but with different methods. Using well known tools, I will show how the two interval result establishes the lower bound on the Hausdorff dimension.

I will also present recent joint work with Scott Sheffield that uses the Schramm-Zhou techniques and an abstract appeal to the Doob-Meyer decomposition to construct a natural measure, that we call the "conformal Minkowski measure", on the boundary set. The construction is motivated by a similar one on the SLE curve itself that is due to Lawler and Sheffield. I will explain why our measure is a natural extension of the boundary measure for discrete models, such as percolation exploration paths, and present numerical evidence that the discrete measure, properly scaled, converges to our continuum measure.

14 Combining two Weibull distributions using a mixing parameter

[CS 6,(page 8)]

Ali Abdul Hussain Salih AL-WAKEEL, *Univesiti Kebagsaan Malaysia*
Ahmad Mahir bin RAZALI, *Universiti Kebangsaan Malaysia*

This paper aims to combine two Weibull distributions and produce a mixture distribution by including a mixing parameter which represents the proportions of mixing the two component Weibull Distributions.

The mixture distribution produced from the combination of two or more Weibull distributions, has a number of parameters. These parameters include; shape parameters, scale parameters, location parameters, in addition to the mixing parameter (w).

A mixture distribution is even more useful because multiple causes of failure can be simultaneously modelled. In this paper we shall concentrate on the estimation of the mixing parameter using maximum likelihood estimation. This method has been chosen as it is more important than the other methods.

The mixing parameter (w; 0<w<1) can take different values in the same distribution. Also, these values vary from one distribution to the other.

A number of samples; small, medium and large will be considered to estimate values for the mixing parameter (w).

As a measure, we shall use the average and standard deviation to assess the accuracy of the estimation.

15 Estimators of intraclass correlation coefficient for binary data

[CS 65,(page 49)]

Abdella Zidan AMHEMAD, *Department of Statistics, Faculty of Science, Al-Fateh University, Tripoli – Libya*
Abdesalam Omran NAFAD, *Department of Statistics, Faculty of Science, Al-Fateh University, Tripoli – Libya*

The correlation between observations in the same cluster is known to be the intraclass correlation coefficient (ICC). It is well known as a quantitative measure of the resemblance among observations within clusters, and it has a lengthy history of application in several different fields of research such as epidemiologic research, reliability theory and in genetics plays a central role in estimating the heritability of selected traits in animal and plant populations. In this paper we investigate a number of different estimators of the ICC for the binary data. Binary data are generated from several models where we compare the different methods of estimation and the effect of the underlying models. It is found that the performance of the intraclass estimator of the binary data depends on the model assumed for the data and other factors such as the number of clusters and the sample size of each cluster.

16 Estimation of multiparameter random fields in the presence of fractional long-memory noises

[CS 2,(page 7)]

Anna AMIRDJANOVA, *University of Michigan*
Matthew LINN, *University of Michigan*

An interesting estimation problem, arising in many dynamical systems, is that of filtering; namely, one wishes to estimate a trajectory of a "signal" process (which is not observed) from a given path of an observation process, where the latter is a nonlinear functional of the signal plus noise.

In the classical framework, the stochastic processes are parameterized by a single parameter (interpreted as "time"), the observation noise is a martingale (say, a Brownian motion), and the best mean-square estimate of the signal, called the optimal filter, has a number of useful representations and satisfies the well-known Kushner-FKK and Duncan-Mortensen-Zakai stochastic partial differential equations.

However, there are many applications, arising, for example, in connection with denoising and filtering of images and video-streams, where the parameter space is multidimensional. Another level of difficulty is added if the observation noise has a long-memory structure, which leads to "nonstandard" evolution equations. Each of the two features (multidimensional parameter space and long-memory observation noise) does not permit the use of the classical theory of filtering and the combination of the two has not been previously explored in mathematical literature.

This talk focuses on nonlinear filtering of a signal in the presence of long-memory fractional Gaussian noise. We start by introducing the evolution equations and integral representations of the optimal filter in the one-parameter case. Next, using fractional calculus and multiparameter martingale theory, the case of spatial nonlinear filtering of a random field observed in the presence of a persistent fractional Brownian sheet will be explored. New properties of multiple integrals with respect to Gaussian random fields will be discussed and their applications to filtering will be shown.

17 Bayesian study of stochastic volatility models with STAR volatilities and leverage effect.

[CS 33,(page 32)]

Esmail AMIRI, *Department of Statistics, IK International University*

The results of time series studies present that a sequence of returns on some financial assets often exhibit time dependent variances and excess kurtosis in the marginal distributions. Two kinds of models have been suggested by researchers to predict the returns in this situation: observation-driven and parameter driven models. In parameter-driven models, it is assumed that the time dependent variances are random variables generated by an underlying stochastic process. These models are named stochastic volatility models (SV). In a Bayesian frame work we assume the time dependent variances follow a non-linear autoregressive model known as *smooth transition autoregressive* (STAR) model and leverage effect between volatility and mean innovations is present. To estimate the parameters of the SV model, Markov chain Monte Carlo (MCMC) methods is applied. A data set of log transformed Pound/Dollar exchange rate is analyzed with the proposed method. The result showed that SV-STAR performed better than SV-AR.

18 Some bivariate distributions and processes useful in financial data modelling

[CS 1,(page 6)]

Mariamma ANTONY, *Department of Statistics, University of Calicut, Kerala 673 635, India*

Empirical analysis of some bivariate data, especially in the fields of Biology, Mathematical Finance, Communication Theory, Environmental Science etc. shows that bivariate observations are asymmetric and heavy tailed with different tail behavior. Kozubowski et al. (2005) considered a bivariate distribution related to Laplace and Linnik distribution, namely marginal Laplace and Linnik distribution, which can be applied for modeling bivariate data with this character. In this paper, geometric marginal asymmetric Laplace and asymmetric Linnik distribution is introduced and studied. Note that, the geometric marginal asymmetric Laplace and asymmetric Linnik distribution arise as the limit distribution of geometric sums of asymmetric Laplace and asymmetric Linnik random variables. Time series models with geometric marginal asymmetric Laplace and asymmetric Linnik distributions are introduced. Also in this paper, we study the properties of geometric marginal asymmetric Laplace - asymmetric Linnik distribution. A bivariate time series model with this marginal distribution is developed and studied. Geometric bivariate semi-Laplace distribution is also introduced and studied in this Chapter. Heavy tailed bivariate distributions with different tail index are used for modeling bivariate data. A bivariate extension of geometric asymmetric Linnik distribution is introduced and its properties are studied. Geometric bivariate semi-Laplace distribution is also introduced and studied.

19 A note on double k-class estimators under elliptical symmetry

[CS 23,(page 22)]

M. ARASHI, *Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Iran*
S. M. M. TABATABAEY, *Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Iran*

In this paper, estimation of the regression vector parameter in the multiple regression model $y = \boldsymbol{X}\beta + \epsilon$ is considered, when the error term belongs to the class of elliptically contoured distributions (ECD), say, $\epsilon \sim EC_n(0, \sigma^2\boldsymbol{V}, \psi)$, where σ^2 is unknown and $\boldsymbol{V}$ is a symmetric p.d known matrix with the density generator ψ. It is well-known that UMVU estimator of β has the form $(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}y$. In this paper using integral series representation of ECDs, the dominance conditions of double k-class estimators given by

$$\hat{\beta}_{k_1,k_2} = \left[1 - \frac{k_1\hat{\epsilon}'\boldsymbol{V}^{-1}\hat{\epsilon}}{y'y - k_2\hat{\epsilon}'\boldsymbol{V}^{-1}\hat{\epsilon}}\right]\hat{\beta}$$

over UMVUE, have been derived under weighted quadratic loss function.

20 Selection methods among spatial models

[CS 62,(page 48)]

N. Miklos ARATO, *Eotvos Lorand University, Budapest*

In the last years the development and the use of Markov chain Monte Carlo (MCMC) methods gave a possibility to characterize and to fit to increasingly large classes of spatial models. In practical applications at the final step of use these models one must decide about the used model. On the basis of the Bayesian models it is possible to apply Bayesian information criterion or deviance information criterion for this goal and for the comparison of the models. Nowadays instead of using only classified general databases there one can use individual datasets too to make new comparisons among Bayesian and non-Bayesian spatial and non-spatial models. In the proposed presentation an example from household insurances is analyzed using data from more than half a million contracts. We pay special attention to the simultaneous estimation of the spatial factors and all other type of factors. Direct estimation for all the 3111 localities is obviously not viable, except perhaps for the capital Budapest. We applied 4 different approaches: generalized linear model without spatial dependence, non- Bayesian spatial smoothing, Clayton-Kaldor and BYM model. In the Bayesian case a Markov Chain Monte Carlo simulation is used for the estimation. The challenge in the implementation of the MCMC algorithm is the very high dimension. For model selection we used cross-validation method. We propose some different distance statistics between real and predicted data and we analyze their properties.

21 Statistical analysis for a four station tandem queue with blocking and infinite queue in front of every station

[CS 13,(page 13)]

A.D. Jerome STANLEY, *Loyola College (Autonomous), Chennai 600034, India*
Pichika CHANDRASEKHAR, *Loyola College (Autonomous), Chennai 600034, India*

A maximum likelihood estimator (MLE), a consistent asymptotically normal (CAN) estimator and asymptotic confidence limits for the probability that all the four stations are busy in a four station tandem queue with blocking and infinite queue capacity in front of every station are obtained. An Application scenario is simulated with numerical work.

22 Maximum equality estimators of various distributions

[CS 5,(page 8)]

Senay ASMA, *Anadolu University*
Ozer OZDEMIR, *Anadolu University*

In this study, maximum equality method is taken into account in order to obtain the estimators of known probability density functions. Moreover, these estimators are illustrated on simulated data.

23 Conditioned ψ super-Brownian motion

[IS 17,(page 30)]

Siva ATHREYA, *Indian Statistical Institute*
Tom SALISBURY, *York University*

We extend earlier results on conditioning of super-Brownian motion to general branching rules. We obtain representations of the conditioned process, both as an h-transform, and as an unconditioned superprocess with immigration along a branching tree. Unlike the finite-variance branching setting, these trees are no longer binary, and strictly positive mass can be created at branch points. This construction is singular in the case of stable branching, and we analyze this singularity by approaching the stable branching function via analytic approximations. In this context the singularity of the stable case can be attributed to blowup of the mass created at the first branch of the tree.

24 State estimation in noisy quantum homodyne tomography

[IS 25,(page 51)]

Jean-Marie AUBRY, *University of Paris-Est (Créteil)*

The optical approach to quantum computing relies on the accurate estimation of the quantum state of a low intensity laser. Such a state is represented by either a trace 1, self-adjoint positive operator ρ acting on $L^2(\mathbf{R})$ or, in the equivalent Wigner representation, by a (non compactly supported) function W_ρ of two variables.

Given a phase $\phi \in [0, \pi)$, the measurement technique called homodyne tomography yields a random variable whose density is given by the Radon transform at angle ϕ of W_ρ. Practically, a Gaussian noise with variance $\frac{1-\eta}{2}$ is added by the physical device, where $0 < \eta \leq 1$ is a known efficiency parameter. The statistical model consists in n independent, identically distributed couples (ϕ_ℓ, Y_ℓ), where the ϕ_ℓ are uniformly random phases and Y_ℓ are the corresponding noisy measurements.

We propose nonparametric projection estimators of ρ, viewed as an infinite matrix in the Fock basis, and kernel type estimators of W_ρ. We compute their asymptotic performance in L^2 risk and compare their behaviour. The estimator of ρ can be easily projected on the space of quantum states, while the estimator of W_ρ allows for detection of non-classical features of the quantum state.

25 Existence of density in the lognormal market model (BGM).

[CS 78,(page 57)]

E. AZMY, *Monash University*
Mark JOSHI, *University of Melbourne*
Fima KLEBANER, *Monash Uiversity*

Existence of densities in financial models is important from both theoretical as well as practical considerations. It is well known that Hormander condition is sufficient for the existence of density, but unfortunately this condition is very hard to check in general, as also commented in the recent paper by Davis M.H.A. and Mataix-Pastor V. (2007), where existence of density was shown only in the case of two dimensions, n=2 for the Swap Market Model. Here we consider the LMM (BGM) model for interest rates, which consists of n stochastic differential equations driven by a single Brownian motion. We prove that the Hormander's condition for the case n=3 holds by showing that certain curves on which determinants vanish do not intersect. We also conjecture a similar condition for the case n=4 and the case of general n.

26 Stochastic models for geometric routing in mobile ad hoc networks

[IS 31,(page 51)]

Francois BACCELLI, *Institut National de Recherche en Informatique et Automatique and Ecole Normale Supérieure, Paris*

Stochastic geometry provides a natural way of defining (and computing) macroscopic properties of mobile ad hoc networks, by some averaging over all potential geometrical patterns for the mobiles.

The talk will survey recent results obtained by this approach for evaluating the performance of a class of multihop routing protocols used in this context and often referred to as geometric or geographic routing.

In a first typical example, when a given packet is located at some tagged mobile node, the next hop on the packet's route is chosen as the nearest among the nodes which are closer from the packet's destination than the tagged node. In a second typical example, one selects as next hop the node being the closest to the destination among the set of nodes which receive this packet successfully when broadcasted by the tagged node.

Such algorithms can be used both in the point to point case and in the multicast case.

We will review a few analytical results which were recently obtained on such routing algorithms using stochastic geometry. Within this framework, random routes are built on realizations of homogeneous Poisson point processes of the Euclidean plane. The geometry of these routes can be analyzed thanks to the locality of the next hop definition and the mean performance of the algorithms can then be characterized via averages computed over all Poisson configurations.

27 Partial linear models for longitudinal data based on quadratic inference functions

[PS 2,(page 17)]

Yang BAI, *The University of Hong Kong*
Zhongyi ZHU, *Fudan University*
Wing Kam FUNG, *The University of Hong Kong*

In this paper we consider improved estimating equations for semiparametric partial linear models (PLM) for longitudinal data, or clustered data in general. We approximate the nonparametric function in the PLM by a regression spline, and utilize quadratic inference functions (QIF) in the estimating equations

to achieve a more efficient estimation of the parametric part in the model, even when the correlation structure is misspecified. Moreover, we construct a test which is an analog to the likelihood ratio inference function for inferring the parametric component in the model. The proposed methods perform well in simulation studies and real data analysis conducted in this paper.

28 Limiting Properties of eigenmatrices of Large sample covariance matrices

[IS 16,(page 55)]

Zhidong BAI, *National University of Singapore*

The spectral theory of large dimensional random matrices has been widely applied to many disciplines, such as multivariate statistical analysis, wireless communications, finance and economics statistics, bioinformatics, etc. But most results of the theory are concerned with the limiting behavior of eigenvalues of large random matrices while fewer reults are about the matrices of orthonormal eigenvectors (or eigenmatrix for short) of large dimensional random matrices. However, from the point of view of multivariate analysis, eigenmatrices play a more important role than eigenvectors, especially in principal component analysis, factor analysis etc. In this talk, I will introduce some known results on eigenmatrices of large sample covariance matrices and some further considerations in ths direction.

29 Constructing length and coverage-based prior distributions

[CS 42,(page 37)]

Paul D. BAINES, *Department of Statistics, Harvard University*

X.-L. MENG, *Department of Statistics, Harvard University*

Interval estimation procedures are traditionally evaluated in terms of their coverage properties. Although the criterion is Frequency-based, the Bayesian paradigm provides a potential route to the construction of intervals with ‘optimal‘ coverage properties. From the original work by Welch and Peers, much progress has been made on the theoretical foundation of Probability Matching Priors (PMPs) i.e., priors with exact or asymptotic Frequentist validity.

Despite the progress in the theory of PMPs, their implementation is often plagued with many daunting computational and theoretical challenges. Alternative routes to obtaining approximate Frequentist validity have often proven to be theoretically simpler and more computationally manageable. Perhaps as a result of this, use of PMPs in applied statistical work remains limited. In addition to the difficulties involved in implementation, PMPs are defined according to an asymptotic property, making them ineffective in many small-sample cases. We propose an alternative criteria for the evaluation and derivation of Frequency-motivated prior distributions. The criteria invokes no asymptotic assumptions, and incorporates a simultaneous trade-off between interval coverage and interval length. By incorporating length/volume considerations into our criteria we seek to produce interval estimation procedures more closely aligned with their practical usage; which may depend on coverage alone, or a natural length-coverage trade-off. The criterion can be used to select ‘optimal‘ priors in the general sense, or from a pre-specified parametric family. Prior information may be incorporated on any subset of the nuisance parameters.

This work is motivated by a standard problem in High-Energy Particle Physics: a particularly challenging example for the standard techniques. Simulation studies are used to compare a selection of Bayesian and Frequentist interval estimation procedures in terms of their coverage properties, and also in terms of our proposed new criterion. In some settings, analytic results may be obtained, we illustrate this in the context of a simple Gaussian model.

30 On monotonocity and maximum fuzziness of fuzzy entropy and directed divergence measures

[CS 38,(page 34)]

Rakesh Kumar BAJAJ, *Jaypee Univsersity of Information Technology, Waknaghat, India*

D. S. HOODA, *Jaypee Institute of Engineering and Technology, Guna, India*

In the present communication, we have investigated the monotone property and maximum fuzziness of parametric generalized measures of fuzzy entropy and directed divergence. Comparison of monotonicity between the corresponding probabilistic measures and the generalized measures of fuzzy entropy and directed divergence under consideration has been studied. Particular cases have also been discussed.

31 Fourier coefficients and invariance of random fields on homogeneous spaces of a compact group

[CS 27,(page 26)]

Paolo BALDI, *Department of Mathematics, University of Rome Tor Vergata*
Domenico MARINUCCI, *Department of Mathematics, University of Rome Tor Vergata*
V.S. VARADARAJAN, *UCLA*

Recently applications to cosmology have prompted a renewed interest on the subject of the study of rotationally invariant random fields on the sphere. In this talk we give some results concerning the characterization of the Fourier coefficients of an invariant r.f., proving namely that an invariant r.f. on the sphere whose Fourier coefficients are invariant is necessarily Gaussian. Extensions to general homogeneous spaces of compact groups are also given.

Reference

1. Baldi P., Marinucci D., Varadarajan V.S. (2007) *On the characterization of isotropic Gaussian fields on homogeneous spaces of compact groups*, Electronic J.Probab. 12, 291–302.

32 On the cluster size distribution for percolation on some general graphs

[CS 71,(page 53)]

Antar BANDYOPADHYAY, *Indian Statistical Institute, New Delhi, India*
Jeffrey STEIF, *Chalmers University of Technology, Gothenburg, Sweden*
Adam TIMAR, *University of British Columbia, Vancouver, Canada*

We obtain various results concerning the cluster size distribution conditioned on being finite in the supercritical regime on general graphs. Our primary interest is to investigate the qualitative difference which may occur in the decay rate of the tail of this distribution, when the underlying graphs have different qualitative properties with respect to amenability and/or transitivity.

33 Estimation of hazard of HIV-1 infection of vertically transmitted children by using parametric and semi-parametric survival models for doubly censored failure times and fixed covariates

[CS 81,(page 59)]

Tanushree BANERJEE, *Department of Statistics, University of Delhi, Delhi, India*
Gurprit GROVER, *Department of Statistics, University of Delhi, Delhi, India*
Dipankar BANERJEE, *Institute of Human Behaviour and Allied Sciences, Delhi, India*

In many epidemiological studies, the survival time of interest is the elapsed time between two related events, the originating event and the failure event, and the times of occurrences of both events are right or interval censored. To estimate the effect of covariates on survival data when the times of both originating and failure events can be interval or right-censored this paper utilizes a method for applying the proportional hazards model given by Kim, De Gruttola and Lagakos along with Log-Linear models on a data of 130 vertically infected human immunodeficiency virus type 1 (HIV-1) children visiting the paediatrics clinic. The semi-parametric procedure was extended by inclusion of additional covariates viz. antiretroviral (ARV) therapy, age, change in CD4+T cell count. Our findings suggest that ARV therapy had a significant effect on risk of death (p-value $<$ 0.001). We further investigated the effect of age and change in CD4+T cell count on the risk of death. These covariates also exhibited a possible association with risk of death (p-value $<$ 0.0001). The effect of number of years under ARV therapy with diagnosis year as a confounding factor was directly related to longevity. The Log-Linear model was utilized by applying the imputation technique to estimate the sero-conversion time and time of death. Again the effect of treatment, age and change in CD4+T cell count was studied on the risk of death. The estimates obtained by the two procedures showed no significant difference (p-value $>$ 0.05). Linear regression analysis revealed that on an average there was a decrease of nearly 14 units in the CD4+T cell count with increase in the time period at 6 months succession in the ARV therapy group while in the Non-ARV therapy group a decrease of 34 units in the CD4+T cell count was noted with increase in the time period. Based on our analysis, we can suggest that in case the intervals are narrow in length, to avoid computational complexities as in case of semi-parametric procedures involving analysis of interval censored data, a parametric approach could be used instead. Future research needs to be carried out to test the statistical difference between the estimated coefficients obtained by the regression analysis for the two groups.

34 A large investor trading at market indifference prices

[IS 27,(page 10)]

Peter BANK, *Technische Universit*
Dmitry KRAMKOV, *Carnegie Mellon University and University of Oxford*

We develop a nonlinear model of a financial market where a couple of market makers fill orders posted by a large investor. Transaction prices are quoted in such a way that, after all market makers' have re-hedged their position to arrive at a new Pareto-optimal allocation of wealth, their expected utilities do not change. The dynamics of this market are thus described by a sequence of Pareto allocations which accommodate the investor's orders. Mathematically, this amounts to a nonlinear dynamic system which we show can be best described by a nonlinear SDE for the market makers' expected utility process. This allows for an easy proof of absence of arbitrage for the large investor and, under suitable assumptions, it also permits the computation of hedging strategies and replication prices in such illiquid financial markets.

35 Profit analysis of an electric supply system working under different weather conditions subject to inspection

[CS 12,(page 13)]

Mahabir Singh BARAK, *Department of Statistics, M.D.University Rohtak-124001, India*
Suresh Chander MALIK, *Department of Statistics, M.D.University, Rohtak-124001, India*

The paper has been designed with a view to develop two reliability models for an electric supply system operating under two weather conditions- normal and abnormal systems has a single-unit which may fail completely either directly from normal mode or via partial failure .There is a single server who visits the system immediately when ever needed and plays the dual role of inspection and repair. In model I, server inspects the units only at its complete failure to examine the feasibility of repair while in model II inspection of the unit is done both at its partial and complete failure. If inspection reveals that repair of the unit is not feasible, it is replaced by new one. The operation, inspection and repair of the system are stopped in abnormal weather. The distributions of failure time and change of weather conditions follow negative exponential while that of inspection and repair times are taken as arbitrary with different probability density functions. All random variables are assumed as mutually independent and uncorrelated. The different measures of system effectiveness are obtained to facilitate the research out comes by using semi-Markov process and regenerative point technique. Graphics are also plotted to depict the behaviour of MTSF and Profit of the models for a particular case.

36 Modeling the evolution of the precipitations in a region of Romania

[CS 14,(page 14)]

Alina BARBULESCU, *Ovidius University of Constanta, Romania*

The study of the precipitations evolutions is important point of view of dimensioning the hydro-amelioration works, as well as for preventing the desertification in some regions of the earth. Dobrudgea is one of the Romanian regions where the unequal distribution of precipitations was remarked and this phenomena is correlated with the temperatures increasing in the last 20 years. Taking account of the previous reasons, we made the analysis of spatial an temporal evolutions of the precipitations in this region, along 41 years, trying to predict their future evolution. The long range dependence and the break of the phenomena were analyzed. The seasonal decomposition were performed, revealing the evolution trend and determining a possible scenario for the global evolution in the next period.

37 Symmetric random walks in random environments

[Medallion Lecture 3,(page 60)]

Martin T. BARLOW, *University of British Columbia*

In this talk I will survey recent progress on the *random conductance model*. Let $(\mathbf{Z}^d, \mathbf{E}_d)$ be the Euclidean lattice, and let μ_e, $e \in \mathbf{E}_d$ be i.i.d.r.v on $[0, \infty)$. Let $Y = (Y_t, t \in [0, \infty), P^x_\omega, x \in \mathbf{Z^d})$ be the continuous time Markov chain with jump probabilities from x to y proportional to μ_{xy}. Thus Y has generator

$$Lf(x) = \gamma_x^{-1} \sum_{y \sim x} \mu_{xy}(f(y) - f(x)),$$

where γ is a 'speed' measure. Two natural choices for γ are $\gamma_x = 1$ for all x, and $\gamma_x = \mu_x = \sum_{y \sim x} \mu_{xy}$. In the first case the mean waiting time at x before a jump is $\mu_x = \sum_y \mu_{xy}$, and in the second it is 1. We can call these respectively the 'variable speed' and 'fixed speed' continuous time random walks – VSRW and CSRW for short.

A special case of the above is when $\mu_e \in \{0, 1\}$, and so Y is a random walk on a (supercritical) percolation cluster.

In this talk I will discuss recent work, by myself and others, on invariance principles for Y, and Gaussian bounds for the transition densities

$$q_t^\omega(x, y) = \gamma(y)^{-1} P_\omega^x(Y_t = y).$$

I will also discuss the connection with 'trap' models and 'ageing' phenomena.

38 On a problem of optimal allocation

[CS 29,(page 26)]

Jay BARTROFF, *University of Southern California*
Larry GOLDSTEIN, *University of Southern California*
Ester SAMUEL-CAHN, *Hebrew University of Jerusalem*

A problem of optimally allocating partially effective ammunition to be used on randomly arriving enemies to maximize a bomber's probability of survival, known as the bomber problem, was first posed by Klinger and Brown (1968). They conjectured a set of seemingly obvious properties of the optimal allocation function K, some of which have been proved or disproved by other authors since then, while others remain unsettled. We will introduce the problem, summarize the state of these conjectures, and present a new approach to the long-standing, unproven conjecture of monotonicity of K in the bomber's initial ammunition.

39 Central limit theorem for inelastic Kac equation

[CS 72,(page 54)]

Federico BASSETTI, *Dipartimento di Matematica, Universita degli Studi di Pavia, Italy*
Lucia LADELLI, *Dipartimento di Matematica, Politecnico di Milano, Italy*
Eugenio REGAZZINI, *Dipartimento di Matematica, Universita degli Studi di Pavia, Italy*

This talk deals with a one–dimensional model for granular materials, which reduces to an inelastic version of the Kac kinetic equation, with inelasticity parameter $p > 0$. In particular, we provide bounds for certain distances – such as specific weighted *chi*–distances and the Kolmogorov distance – between the solution of that equation and the limit. It is assumed that the even part of the initial datum (which determines the asymptotic properties of the solution) belongs to the domain of normal attraction of a symmetric stable distribution with characteristic exponent $a = 2/(1 + p)$. With such initial data, it turns out that the limit exists and is just the aforementioned stable distribution. A necessary condition for the relaxation to equilibrium is also proved. Some bounds are obtained without introducing any extra–condition. Sharper bounds, of an exponential type, are exhibited in the presence of additional assumptions concerning either the behaviour, near the origin, of the initial characteristic function, or the behaviour, at infinity, of the initial probability distribution function. The methods used to obtain the results are essentially inspired to previous work of Harald Cramér and its developments due to Peter Hall. In fact, our results on the solution of the inelastic equation heavily depend on a deep study of the convergence in distribution of certain sums of a random number of random weighted random variables.

40 Modelling precipitation in Sweden using non-homogeneous multiple step Markov chains

[CS 14,(page 14)]

Anastassia BAXEVANI, *Department of Mathematical Sciences, University of Gothenburg, Chalmers University of Technology, Sweden*
Jan LENNARTSSON, *Department of Mathematical Sciences, University of Gothenburg, Chalmers University of Technology, Sweden*
Deliang CHEN, *Department of Earth Sciences, University of Gothenburg, Sweden*

In this paper, we propose a new method for modelling precipitation in Sweden. We consider a chain dependent stochastic model that consist of a component that models the probability of occurrence of precipitation in a station and a component that models the amount of precipitation in the same station when precipitation does occur. For the first model, we show that in most of the Swedish stations a Markov chain of order higher than one is required. For the second model we proceed in two steps. First we model the dependence structure of the amount precipitation process using copula, and then we compute the marginal distribution using a composite of the empirical distribution below a threshold and the generalized Pareto distribution for the excesses above the given threshold. The derived models are validated by computing different weather indices.

41 A discrete class of probability distributions with variety of applications

[CS 41,(page 37)]

Rousan Ara BEGUM, *Darrang College, Tezpur-784001, Assam, India*

A discrete class of probability distributions having wide flexibility and important implications has been studied. All these distributions derived in this investigation using Lagranges expansions of type-I and type-II fall on a certain type of distribution known as Lagrangian Probability distribution. The probabilistic structures of these distributions and some of their important properties have been studied. It is believed that Lagrangian Poisson (LP) and Lagrangian negative-binomial (LNB) distributions should give better fit than their classical forms. In general, it is also conceivable that discrete data occurring in biology, ecology, home injuries and accidents data, mycology, etc. could be statistically modeled more perfectly and more accurately, because of it wide flexibility.

42 Bivariate Dirichlet distributions based on Beta marginals

[CS 76,(page 57)]

A BEKKER, *University of Pretoria, South Africa*
JJJ ROUX, *University of Pretoria, South Africa*
R EHLERS, *University of Pretoria*

When modeling a bivariate random variable, (X1,X2) it is possible that one knows the marginal distributions of each of the variables, but very little about the dependence structure between them. The task is to find a bivariate distribution having the required marginals. In this paper generalized bivariate Dirichlet distributions are derived with marginals having different Beta type distributions.

43 Stochastic calculus for convoluted Lévy processes

[CS 77,(page 57)]

Christian BENDER, *Technische Universität Braunschweig*
Tina MARQUARDT, *Technische Universität München*

In recent years fractional Brownian motion and other Gaussian processes, which are obtained by convolution of an integral kernel with a Brownian motion, have been widely studied as a noise source with memory effects. Potential applications for these noise sources with memory are in such diverse fields as telecommunication, hydrology, and finance, to mention a few.

In [2] fractional Lévy processes were introduced. While capturing memory effects in a similar fashion as a fractional Brownian motion does, the convolution with a Lévy process provides more flexibility concerning the distribution of the noise, e.g. it allows for heavy tails. In this talk we discuss a larger class of processes which are obtained by convolution of a rather general Volterra type kernel with a centered pure jump Lévy process. These convoluted Lévy process may have jumps and/or memory effects depending on the choice of the kernel. We motivate and construct a stochastic integral with respect to convoluted Lévy processes. The integral is of Skorohod type, and so its zero expectation property makes it a possible choice to model an additive noise. As a main result we derive an Itô formula for these integrals. The Itô formula clarifies the different influences of jumps and memory effects, which are captured in different terms.

The talk is based on a joint paper [1] with T. Marquardt (Munich).

References

1. Bender, C., Marquardt, T. (2008) Stochastic calculus for convoluted Lévy processes. *Bernoulli*, forthcoming.

2. Marquardt, T. (2006) Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli* **12** 1090–1126.

44 A large deviations estimate for simple random walk via a sharp LCLT

[CS 50,(page 43)]

Christian BENEŠ, *Brooklyn College of the City University of New York*

Donskerś theorem suggests that if B is a d-dimensional Brownian motion and S is a d-dimensional simple random walk, $P(|B(n)| \geq r\sqrt{n})$ and $P(|S(dn)| \geq r\sqrt{n})$ may decay at the same rate. We derive a sharp local central limit theorem for simple random walk which is more precise than those currently available for points that are unusually distant from the origin. This allows us to show that the decay of the large deviations probability for random walk is more rapid than for Brownian motion: If S is one-dimensional simple random walk, then there exists a constant C such that for every $r, n \geq 1$,

$$P(|S(n)| \geq r\sqrt{n}) \leq C \exp\left\{ - \sum_{l=1}^{\left[\frac{\log n}{\log n - 2\log r}\right]} \frac{1}{l(2l-1)} \frac{(r/\sqrt{2})^{2l}}{n^{l-1}} \right\}. \qquad (1)$$

We show that this difference in behavior exists for general d.

45 Limit theorems for change-point estimation

[CS 74,(page 54)]

Samir BEN HARIZ , *Université du Maine, France*

Let $(X_i)_{i=1..n}$ be a sequence such that $law(X_i) = P_n$ if $i \leq n\theta$ and $law(X_i) = Q_n$ if $i > n\theta$ where $0 < \theta < 1$ is the location of the change-point to be estimated.

Assume that the sequence satisfies the following condition :

$$\sup_{f \in \mathcal{F}} E\left(\sum_{i=1}^{n} f(X_i) - E(f(X_i))\right)^2 \leq Cn^{2-\theta},$$

where $\theta > 0$ and $\mathcal{F}$ a family of functions used to construct a norm on the space of probabilities measures. We show for a wide class of parametric and nonparametric estimators that the optimal $1/n$ rate is achieved.

Next, we focus on a family of cumulative-sum change-point estimators. Since cumulative-sum estimators compare differences between empirical means, it seems natural that ergodicity is a minimal assumption for consistent change-point estimation. Surprisingly, we show that change-point estimation can be consistently performed even when the law of large numbers fails.

We determine the rate of convergence for sequences under very general conditions including nonergodic cases. In particular, we determine the rate of convergence for sequences in which the correlations decay to zero arbitrarily slowly or even do not decay to zero at all.

References

1. Ben Hariz, S., Wylie, J.J., Zhang Q., (2007) *Optimal Rate of Convergence for Nonparametric change-point estimators for non-stationary sequences* . Ann. Statist. Volume 35, Number 4 (2007), 1802-1826.
2. Dumbgen, L. (1991), The asymptotic behavior of some nonparametric change-point estimators, *Ann. Statist.* **19**, 1471–1495.

46 New sign correlations related to Kendall's tau and Spearman's rho for measuring arbitrary forms of association

[CS 18,(page 19)]

Wicher BERGSMA, *London School of Economics and Political Science*

Kendall's tau and Spearman's rho are sign correlation coefficients which measure monotonic association in bivariate data sets. In this paper we introduce two new coefficients, related to these, which measure arbitrary forms of association. The first one is based on "concordance" and "discordance" of four points in the plane, the latter one of six points. We propose tests of independence based on either of the new coefficients as an alternative to the chi-square test for ordinal categorical and continuous data, and argue that this leads to a significant increase in power in most practical situations.

47 MCMC in infinite dimensions

[IS 10,(page 18)]

Alexandros BESKOS, *University of Warwick, Department of Statistics*
Gareth ROBERTS, *University of Warwick, UK*
Andrew STUART, *University of Warwick, Mathematics Institute*
Jochen VOSS, *University of Warwick, Mathematics Institute*

In infinite-dimensional spaces, one can naturally select a linear law as a-priori description of a complex system. The a-posteriori distribution can then be many times written as a change of measure from a Gaussian law. We develop local Metropolis-Hastings algorithms for these targets and study their behaviour. The complexity of the state space forces a more involved approach for defining Random-Walk and Langevin-type algorithms. Central in our considerations is the construction of a Hilbert-space valued SDE invariant under the target law. The use of an implicit scheme for it's numerical solution is critical for the development of practical algorithms. This MCMC methodology has been motivated by high-dimensional applications in molecular dynamics, signal filtering, geophysics, lagrangian data assimilation.

48 Efficient emulators of computer experiments using compactly supported correlation functions

[IS 34,(page 10)]

Derek BINGHAM, *Department of Statistics, Simon Fraser University*
Cari KAUFMAN, *University of California, Berkeley*

Building an emulator for a computer simulator using a Gaussian process (GP) model can be computationally infeasible when the number of trials is large. Here, we propose using a compactly supported correlation function in the specification of the GP model. This has the effect of introducing zeroes into the correlation matrix, so that it can be efficiently manipulated using sparse matrix algorithms. Following the usual product correlation form used for computer experiments, we explain how to impose restrictions on the range of each correlation function. This has the advantage of giving sparsity, while also allowing the ranges to trade off against one another (i.e., isotropy). The methodology is demonstrated via simulation and also on a real computer experiment conducted in cosmology.

49 Pricing of European options using empirical characteristic functions.

[CS 24,(page 24)]

Karol BINKOWSKI, *Department of Statistics, Division of Economic and Financial Studies, Macquarie University, Sydney, Australia*

Pricing problems of financial derivatives are among the most important questions of quantitative finance. Since 1973 when Nobel prize winning model was introduced by Black, Merton and Scholes the Brownian Motion (BM) process gained huge attention of professionals. It is known, however, that market stock log-returns are not fit well by the very popular BM process. Derivatives pricing models which are based on more general Lévy processes tend to perform better. Carr and Madan (1999) and Lewis (2001) (CML) developed a method for vanilla options valuation based on a characteristic function of asset log-returns. Recognizing that the problem is in modeling the distribution of the underlying price process, we use instead a nonparametric approach in the CML formula, and base options valuation on Empirical Characteristic Functions (ECF). We consider four modifications of this model based on the ECF. The first modification requires only historical log-returns of the underlying price process. The other three modifications of the model need, in addition, real option prices. We compare their performance based on data of DAX index and ODAX options written on the index between 1st of June 2006 and 17th of May 2007. Resulted pricing errors shows that our model performs better in some cases, than that of CML.

50 Bayesian clustering of high dimensional data via a directional approach

[CS 11,(page 12)]

Natalia BOCHKINA, *University of Edinburgh, UK*

I will present a novel statistical approach for clustering variables by their correlation in high dimensional data where the number of variables is larger than the number of observations. Knowledge of such classes is important, on the one hand, to confirm or to initiate biological discovery of genes/proteins/metabolites that act together in a cell, and, on the other hand, for use as a single unit in variable selection to avoid ambiguities.

I propose to discover such classes by estimating a square root of the covariance matrix which can be chosen to represent the directions of the residuals for each variable in the space of observations using spectral decomposition. Then, correlated variables "look" in the same direction. This approach restricts the class of covariance matrices to be block matrices of rank at most equal to the number of observations. To identify the clusters, I apply Bayesian nonparametric techniques ased on the Dirichlet Process prior and study conditions for convergence of such estimators.

51 Orthogonal series and asymptotic analysis of canonical U-statistics based on dependent observations

[CS 80,(page 58)]

Igor S. BORISOV, *Sobolev Institute of Mathematics, 630090, Novosibirsk, Russia*
Nadezhda V. VOLODKO, *Sobolev Institute of Mathematics, 630090, Novosibirsk, Russia*

Let $\{X_i;\, i \geq 1\}$ be a stationary sequence of r.v.'s. in a separable metric space $\mathcal{X}$, and let P be the distribution of X_1. Denote by $\{X_i^*;\, i \geq 1\}$ i.i.d. copies of X_1. Introduce the normalized m-variate U-statistic

$$U_n(f) := n^{-m/2} \sum_{1 \leq i_1 \neq \ldots \neq i_m \leq n} f(X_{i_1}, \ldots, X_{i_m}),\ m \geq 2,$$

where $\mathbf{E}f^2(X_1^*, ..., X_m^*) < \infty$ and this kernel satisfies the so-called *degeneracy* condition $\mathbf{E}_{X_k^*} f(X_1^*, ..., X_m^*) = 0$ *a.s.* for every $k = 1, ..., m$, where $\mathbf{E}_{X_k^*}$ is the conditional expectation given $\{X_i^* : i \neq k\}$. Such kernels and the corresponding U-statistics are called *canonical.* Let $\{e_i(t);\ i \geq 0\}$ be an orthonormal basis of the separable Hilbert space $L_2(\mathcal{X}, P)$ such that $e_0(t) \equiv 1$. Every canonical kernel from $L_2(\mathcal{X}^m, P^m)$ admits the representation

$$f(t_1, ..., t_m) = \sum_{i_1,...,i_m=1}^{\infty} f_{i_1...i_m} e_{i_1}(t_1)...e_{i_m}(t_m), \quad (1)$$

where the multiple series $L_2(\mathcal{X}^m, P^m)$-converges. Notice that, due to the degeneracy condition, the multiple sum in (1) does not contain the element $e_0(t)$. We also introduce the following restriction on all m-dimensional distributions of $\{X_i\}$:
(AC) *For all* $j_1 < ... < j_m$, *the distribution of the vector* $(X_{j_1}, ..., X_{j_m})$ *is absolutely continuous relative to the distribution* P^m *of the vector* $(X_1^*, ..., X_m^*)$.

We study non-Gaussian weak limits for canonical U-statistics (and for canonical Von Mises statistics as well) based on samples from stationary sequences satisfying α or φ-mixing conditions.

Theorem. *Let* $\{X_i\}$ *be a* φ*-mixing stationary sequence with the following restriction on the corresponding coefficient:* $\sum_k \varphi^{1/2}(k) < \infty$. *Let a canonical kernel* $f(t_1, ..., t_m)$ *satisfy the conditions* $\sum_{i_1,...,i_m=1}^{\infty} |f_{i_1...i_m}| < \infty$ *and* $\sup_i \mathbf{E}|e_i(X_1)|^m < \infty$. *If, in addition,* (**AC**) *is fulfilled then*

$$U_n(f) \xrightarrow{d} \sum_{i_1,...,i_m=1}^{\infty} f_{i_1...i_m} \prod_{j=1}^{\infty} H_{\nu_j(i_1,...,i_m)}(\tau_j), \quad (2)$$

where the multiple series in (2) *converges almost surely,* $\{\tau_i\}$ *is a Gaussian sequence of centered random variables with covariance matrix*

$$\mathbf{E}\tau_k\tau_l = \mathbf{E}e_k(X_1)e_l(X_1) + \sum_{j=1}^{\infty} \big[\mathbf{E}e_k(X_1)e_l(X_{j+1}) + \mathbf{E}e_l(X_1)e_k(X_{j+1})\big],$$

$\nu_j(i_1, ..., i_m) := \sum_{k=1}^m \delta_{i_k,j}$ *(here* $\delta_{i,j}$ *is the Kronecker symbol), and* $H_k(x) := (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}$ *are the Hermite polynomials.*

Remark. In general, to obtain relation (2), condition (**AC**) as well as the restriction of the Theorem on the coefficients $\{f_{i_1...i_m}\}$ cannot be omitted.

In the case of i.i.d. $\{X_i\}$, an analog of this theorem was proved by H. Rubin and R. A. Vitale (*Ann. Statist.*, 1980, **8**(1), 165-170).

52 Reinforcement learning — a bridge between numerical methods and monte carlo

[CS 40,(page 37)]

Vivek Shripad BORKAR, *School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*

This talk advocates the viewpoint that reinforcement learning algorithms, primarily meant for approximate dynamic programming, can also be cast as a technique for estimating stationary averages and stationary distributions of Markov chains. In this role, they lie somewhere between standard deterministic numerical schemes and Markov chain Monte Carlo, and capture a trade-off between the advantages of either - lower per iterate computation than the former and lower variance than the latter. The specific algorithms we consider will be as follows.

1. A reinforcement learning scheme is presented for solving the Poisson equation associated with an irreducible finite state Markov chain based on a simulation run. The objective is to estimate the stationary average of a prescribed function of the Markov chain. This scheme exploits the algebraic relationship provided by the Poisson equation to replace the averaging operation in MCMC by a conditional averaging operation. An approximate version based on linear function approximation, aimed at addressing the 'curse of dimensionality', is also presented. Convergence of both versions is established.

2. A reinforcement learning scheme is presented for solving the *multiplicative* Poisson equation associated with an irreducible finite state Markov chain based on a simulation run. The objective is to estimate the stationary distribution of the Markov chain. This scheme exploits the algebraic relationship provided by the multiplicative Poisson equation to replace the averaging operation in MCMC by a conditional averaging operation. Again, an approximate version based on linear function approximation, aimed at addressing the 'curse of dimensionality', is presented and convergence of both schemes is established.

Two possible modifications of the basic schemes for accelerating convergence are also considered.

53 Financial time series modelling by Levy potential diffusions

[CS 78,(page 57)]

Svetlana BOROVKOVA, *Vrije Universiteit Amsterdam, The Netherlands*
Ferry Jaya PERMANA, *Universitas Katolik Parahyangan, Indonesia*
Ilya PAVLYUKEVICH, *Humboldt Universitat Berlin, Germany*

Electricity prices from liberalized markets exhibit the so-called *price spikes*: very large but shortlasting price jumps, which are notoriously diffucult to model. Motivated by this problem, we suggest a class of stochastic models - potential Lévy diffusions - for modelling commodity and other asset prices that are subject to price spikes. These models are stochastic diffusion processes driven by α-stable (not necessarily symmetric) Lévy processes, with the drift given by a potential function. Increments of α-stable Lévy processes, which model price jumps, have heavy tails observed in empirical data. A drift given by a potential function attracts the price process back to its mean level, while allowing for a continuum of mean-reversion rates.

We develop model estimation procedures based on the theoretical properties of potential Lévy diffusions, such as the law of the process' exit time from a bounded interval. We apply our model to electricity prices from major European power exchanges. We illustrate the main applications of the model, such as scenario simulation and risk management. Finally, we demonstrate that, if an α-stable distribution in the Lévy process' specification is replaced by a hyperbolic distribution, the model can be used for derivatives pricing.

54 Stochastic target problems with controlled loss

[IS 27,(page 10)]

Bruno BOUCHARD, *CEREMADE, University Paris Dauphine*
Romual ELIE, *CEREMADE, University Paris Dauphine*
Nizar TOUZI, *CMAP, Ecole Polytechnique Paris*

We provide a direct dynamic programming principle for stochastic target problems where the target has to be reached with a given probability or so that a given loss funtion is above a prescribed level. As an application, we provide a pde formulation for the quantile hedging problem in finance which does not use the dual formulation of Follmer and Leukert. As a by-product we obtain new results on stochastic target problems in the case where the controls are unbounded

55 A poor man Wilks phenomenon

[IS 23,(page 35)]

Stephane BOUCHERON, *Laboratoire Probabilit*
Pascal MASSART, *D*

We investigate the Wilks phenomenon in the Statistical Learning Framework. Thanks to recent moment inequalities for functions of independent random variables, we show that the excess empirical risk (that is the difference between the empirical risk of the best classifier in the model and the minimum of the empirical risk in the model) satisfies a Bernstein like inequality where the variance term is a function of the learning rate.

56 Sharp asymptotics for nucleation times in Kawasaki dynamics in large volumes

[IS 33,(page 56)]

Anton BOVIER, *Weierstrass Institute and Technische Universität Berlin*

In this talk we present results on metastability in large volumes at low temperatures for conservative (Kawasaki) dynamics of on Ising lattice gas.

Let β denote the inverse temperature and let be a square box with $\Lambda_\beta \subset Z^2$ be a square box with periodic boundary conditions such that $\lim_{\beta\to\infty} |\Lambda_\beta| = \infty$. We run the dynamics on Λ_β starting from a random initial configuration where all the droplets are small. For large β, and for interaction parameters that correspond to the metastable regime, we investigate how the transition from the metastable state (with only small droplets) to the stable state (with one or more large droplets) takes place under the dynamics. This transition is triggered by the appearance of a single *critical droplet* somewhere in Λ_β. Using potential-theoretic methods, we compute the *average nucleation time* (= the first time a critical droplet appears and starts growing) up to a multiplicative factor that tends to one as $\beta \to \infty$. It turns out that this time grows as $K\beta e^{\Gamma\beta}/|\Lambda_\beta|$ for Kawasaki dynamics, where Γ is the local grand-canonical energy to create a critical droplet and K is a constant reflecting the geometry of the critical droplet. The fact that the average nucleation time is inversely proportional to $|\Lambda_\beta|$ is referred to as *homogeneous nucleation*, because it says that the critical droplet for the transition

appears essentially independently in small boxes that partition Λ_β.

(Joint work with Frank den Hollander and Cristian Spitoni)

57 Randomized load balancing for general service times with the FIFO discipline

[IS 31,(page 51)]

Maury BRAMSON, *University of Minnesota*
Yi LU, *Stanford University*
Balaji PRABHAKAR, *Stanford University*

We consider the randomized load balancing scheme where each arriving job joins the shortest of d randomly chosen queues from among a pool of n queues, where d is fixed and n goes to infinity. Works by Mitzenmacher (1996) and Vvekenskaya et al. (1996) considered the case with Poisson input and exponentially distributed service times, where an explicit formula for the equilibrium distribution was derived when the system is subcritical. For general service times, the service discipline can affect the nature of the equilibrium distribution. Here, we examine its behavior for the FIFO service discipline and service distributions with fat tails.

58 High-dimensional variable selection: from association to intervention

[IS 9,(page 18)]

Peter BÜHLMANN, *ETH Zürich*

High-dimensional (supervised) data where the number of covariates p greatly exceeds the sample size n occur in numerous applications nowadays. Often, one would like to infer the relevant variables or features which relate to a response of interest. In the high-dimensional setting, this is a difficult task due to the huge number 2^p of possible subsets of variables. When considering association-relations between a response and many variables, substantial progress, in computation and theory, has been achieved over last few years. The situation is fundamentally different when focusing on relations arising from interventions. Such relations are of direct interest when pursuing or planning intervention experiments, for example knocking down some genes in a biological system. We illustrate and discuss the computationally feasible FCI- and PC-algorithm (Spirtes, Glymour & Scheines, 2000) for inferring intervention effects from observational data only. Under some assumptions (e.g. that the data is generated from a directed acyclic graph or an ancestral graph) and even if $p \gg n$, it is possible to consistently estimate some minimal bounds for intervention effects and in ideal cases, these bounds allow to infer unique intervention effects. We illustrate the method for discovering genetic modifications in Bacillus Subtilis to improve Riboflavin production rate.

59 Time gap model for the bombings in region XII, Mindanao, Philippines

[CS 79,(page 58)]

Epimaco A. , Jr. CABANLIT, *Mindanao State University, General Santos City, Philippines*
Robert R. KIUNISALA, *South Cotabato Police Provincial Office, Philippines*
Romeo C. , Jr. GALGO, *3rd Company, Regional Mobile Group, Region 12, Philippines*
Steiltjes M. CABANLIT, *Mindanao State University, General Santos City, Philippines*

From the year 2000 to 2006, a total of 96 bombing incidents occured, 101 bombs exploded, 86 were killed and 674 were injured in Region XII, Mindanao, Philippines. This paper presents a time gap model for these bombings. The mixture of exponential and uniform distribution shows a good fit.

60 Weak limits of infinite source Poisson pulses

[CS 1,(page 6)]

Mine ÇAĞLAR, *Koc University*

Convergence in distribution of scaled input processes to a fractional Brownian motion (fBm) or a Lévy process has been studied by several researchers over the last decade. The motivation comes from data traffic in telecommunications which can be modeled as an infinite source Poisson process in particular [1]. On the other hand, both fBm and Lévy processes have recently become prevalent in finance. In this talk, we construct a general stochastic process based on a Poisson random measure N, prove weak convergence results and then interpret its application as a market model in finance.

Inspired by [2], we consider a stochastic process of the form

$$Z_n(t) \\ = \int_{-\infty}^{\infty} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{q}{a_n}\, u \left[f\left(\frac{t-s}{u}\right) - f\left(\frac{-s}{u}\right)\right] N_n(ds, du, dq).$$

where f is a deterministic function, (a_n) is a scaling sequence as $n \to \infty$, and q and u are marks of the resulting Poisson point process. The process Z is scaled but not centered as it is constructed to have zero mean by either the pulse f or the distribution of q. The mean measure of N is taken to have a regularly varying form in u to comply with the long-range dependence property of teletraffic or financial data. The correlation structure of the input process is as in fBm even before taking the limit. However, we can obtain either fBm or stable Lévy motion depending on the particular scaling of N_n and the factor a_n. In the latter case, the process has independent increments and self-similarity exists without long-range dependence. Although the pulse f has a general form for fBm limit, we can show Lévy limit through specific pulses.

As for application in finance, the process Z can be interpreted as the price of a stock. Under the assumption of positive correlation between the total net demand and the price change, we assume that a buy order of an agent increases the price whereas a sell order decreases it. Each order has an effect f proportional to its quantity q. The logarithm of the stock price is found by aggregating the effects of orders placed by the agents. The duration of the effect of an agent's transaction is assumed to follow a heavy tailed distribution. As a semi-martingale, our process does not allow for arbitrage. It is a novel model which is alternative to existing agent based constructions that are semi-Markov processes [3].

References

1. M. Çağlar (2004) *A Long-Range Dependent Workload Model for Packet Data Traffic* . Mathematics of Operations Research, 29: 92-105.
2. R. Cioczek-Georges, B. B. Mandelbrot (1996) *Alternative Micropulses and Fractional Brownian Motion* . Stochastic Processes and their Applications, 64: 143-152.
3. E. Bayraktar, U. Horst, R. Sircar (2006) *A limit Theorem for Financial Markets with Inert Investors* . Mathematics of Operations Research, 31: 789-810.

61 Exact Matrix Completion via Convex Optimization

[IS 20,(page 50)]

Emmanuel J. CANDES, *California Institute of Technology*

This talk considers a problem of considerable practical interest: the recovery of a data matrix from a sampling of its entries. In partially filled out surveys, for instance, we would like to infer the many missing entries. In the area of recommender systems, users submit ratings on a subset of entries in a database, and the vendor provides recommendations based on the user's preferences. Because users only rate a few items, we would like to infer their preference for unrated items (this is the famous Netflix problem). Formally, suppose that we observe m entries selected uniformly at random from a matrix M. Can we complete the matrix and recover the entries that we have not seen?

We show that perhaps surprisingly, one can recover low-rank matrices exactly from what appear to be highly incomplete sets of sampled entries; that is, from a comparably small number of entries. Further, perfect recovery is possible by solving a simple convex optimization program, namely, a convenient semidefinite program (SDP). This result hinges on powerful techniques in probability theory. There are also some connections with the recent literature on compressed sensing, since this says that objects (e.g. information tables) other than signals and images can be perfectly reconstructed from very limited information

62 Analysis of extreme values in meteorology

[CS 14,(page 14)]

F. CASANOVA DEL ANGEL, *Instituto Politecnico Nacional. Mexico*

The objective of this work is to determine wind dynamic action with real statistical data, as well as the basic speeds applicable to distinct recurring periods, pursuant to the hypothesis that the structure has a lineal, elastic behavior of which the contribution of superior modes to the first hypothesis to the structure answer is negligible. Consequently, the gust factor is constant, and the variation of the medium wind speed with the height responds to a distribution likelihood function of wind speed over a period of time. Starting with wind speed meteorological information, it was found that the hourly distribution of wind speed showed a logarithmical, exponential growth in the North Zone of Mexico City (NZMC).

The analyzed information has been input at the IPN Experimental Meteorological Station (Casanova. 2001) located in the NZMC where speed was measured in m/sec, and the direction of the wind in degrees. The first data is that of June 6, 2001, and the last data analyzed is that of April 28, 2006.

Conclusions. It was found that the hourly distribution of wind speed showed a logarithmical, exponential growth composition in the NZMC. Probability distributions were determined of extreme values for Gumbell, Weibul, and Gamma distributions. The Weibull probability density function for maximum values is the only function that does not adjust to the case of the data studied. With respect to minimum values, the three functions applied adjust quite well.

63 Risk hull method for inverse problems.

[IS 22,(page 23)]

Laurent CAVALIER, *Universite Aix-Marseille 1*
Yuri GOLUBEV, *CNRS, Universite Aix-Marseille 1*

We study a standard method of regularization by projections of the linear inverse problem $Y = Af + \epsilon$, where ϵ is a white Gaussian noise, and A is a known compact operator with singular values converging to zero with polynomial decay. The unknown function f is recovered by a projection method using the SVD of A. The bandwidth choice of this projection regularization is governed by a data-driven procedure which is based on the principle of the risk hull minimization. We provide non-asymptotic upper bounds for the mean square risk of this method and we show, in particular, that in numerical simulations, this approach may substantially improve the classical method of unbiased risk estimation.

64 Nonlinear distributions constructed on abstract Wiener spaces

[CS 71,(page 53)]

Uluğ ÇAPAR, *FENS, Sabanci University, Tuzla, Istanbul, Turkey*

For the study of Brownian or white-noise functionals which are more irregular than those in their L^2 spaces, usually Meyer-Watanabe distributions (in abstract Wiener spaces) or Hida distributions (in white-noise spaces) or different variants thereof such as Kubo-Takenaka, Kondratiev distributions etc. are resorted. (cf. also [4] for an improvement). However Meyer-Watanabe distributions are rather narrow compared to Hida distributions and in both of them nonlinear problems can not be tackled satisfactorily without employing concepts such as renormalization, Wick products and S-transforms etc.

In this study we attempt to construct a simplified version of infinite dimensional Colombeau distributions directly on an abstract Wiener space. The work is a continuation and an extension of [1] and [2] and based on the ideas of asymptotic algebras of generalized functions as introduced in [3]. Starting with a particular class of E-valued test functionals

$$D^\infty(E) = \cap_{s>0} \cap_{1<p<\infty} D^p_s(E)$$

we construct a moderate class $\mathcal{E}_{M,s}(W,H,\mu)$ of functionals (s stands for " simplified " and (W,H,μ) is the abstract Wiener space) and a space of null germs $\mathcal{N}_s(W,H,\mu)$. For $E = \mathcal{R}$, $\mathcal{E}_{M,s}$ turns out to be an algebra and $\mathcal{N}_s$ an ideal and the class of generalized functionals is defined as the quotient algebra

$$\mathcal{E}_{M,s}(W,H,\mu)/\mathcal{N}_s(W,H,\mu).$$

This class possesses a natural rule of multiplication and contains Meyer-Watanabe distributions although there will be no privileged inclusion. However this situation is remedied by the fact that different inclusions are within the same equivalence class. Also using asymptotic exponential scale instead of polynomial scale, algebras stable under exponentiation can be obtained. Among applications we consider for fixed t, $\dot{B}(t), \dot{B}^n(t), \exp^{\dot{B}(t)}$ and products of Donsker's delta function.

References

1. Çapar, U., Aktuğlu, H., A New Construction of Random Colombeau Algebras, *Stat. and Prob. Letters* 54(2001), 291-299.
2. Çapar, U., Aktuğlu, H., An Overall View of Stochastics in Colombeau Related Algebras. In *Progress in Probability,* vol. 53 , 67-90 (ed. by Çapar, U., Üstünel, A.S.) Birkhauser (2003).
3. Delcroix, A., Scarpalezos, D., Topology on Asymptotic Algebras of Generalized Functions and Applications , *Monatshefte für Mathematik*, 129 (2001), 1-14.
4. Körezlioğlu, H., Üstünel, A.S. , A New Class of Distributions on Wiener Spaces, *Stoch. Analysis and Related Topics II* 106-121, Springer Verlag (1990).

65 Universality of the REM for dynamics of mean-field spin glasses

[IS 33,(page 56)]

Jiří ČERNÝ, *ETH Zürich, Switzerland*
Gérard BEN AROUS, *Courant Institute, New York, USA*
Anton BOVIER, *Weierstrass Institute and Technische Universität Berlin*

Aging has become one of the main paradigms to describe the long-time behaviour of glassy systems, in particular of spin glasses. In the last decade aging was studied in simple phenomenological models (trap models), and in the simplest mean-field spin-glass model, the Random Energy Model (REM). In these studies an almost universal scheme for aging emerged, based on convergence of the so-called clock process (time-change process describing the dynamics) to a stable subordinator, and on the related arcsine law. All these studies, however, used substantially the independence of the underlying random environment.

We now show, for the first time, that the same aging scheme is relevant in correlated spin glasses. We consider a particular dynamics of a p-spin Sherrington–Kirkpatrick model that can be seen as a time change of simple random walk on the N-dimensional hypercube. We show that, for all $p \geq 3$ and all inverse temperatures $\beta > 0$, there exists a constant $\gamma_{\beta,p} > 0$, such that for all exponential time scales, $\exp(\gamma N)$, with $\gamma < \gamma_{\beta,p}$, the properly rescaled clock process converges to an α-stable subordinator, $\alpha = \gamma/\beta^2 < 1$. This implies that the dynamics exhibits aging at these time scales and the asymptotic behaviour of time-time correlation functions can be described using the arcsine law. In other words, on these time scales (that are much shorter than the equilibration time of the system) the dynamics of p-spin models ages in the same way as the REM, confirming the universality of the REM aging scheme. The Sherrington-Kirkpatrick model (the case $p = 2$) seems to belong to a different universality class.

66 The asymptotic distribution of the domination number of proportional edge proximity catch digraphs

[CS 82,(page 59)]

Elvan CEYHAN, *Department of Mathematics, Koç University, Sarıyer, 34450, Istanbul, Turkey*

We derive the asymptotic distribution of the domination number of a new family of random digraph called proportional edge proximity catch digraph (PCD). The proportional edge PCD is a parametrized digraph based on two sets of points on the plane, where sample size of the elements of one class is assumed to be much smaller than sample size of the other class. Furthermore, the locations of the members of the first class (i.e., the class whose sample size is smaller) are assumed to be fixed while the elements of the second class are randomly distributed over a region of interest. PCDs are constructed based on the relative allocation of the random set of points with respect to the Delaunay triangulation of the other set whose size and locations are fixed. We also use the domination number as a statistic for testing spatial point patterns of segregation and association.

67 Variable window discrete scan statistics

[CS 82,(page 59)]

Jie CHEN, *University of Massachusetts, Boston, USA*
Joseph GLAZ, *University of Connecticut, USA*

In this talk approximations for the distributions of two-dimensional maximum scan score-type statistic and minimum p-value scan statistic are derived for independent and identically distributed Poisson distribution when the total number of events is known. Numerical results are presented to compare the power of these variable window type scan statistics with fixed single window scan statistics.

68 Three-person red-and-black game

[CS 51,(page 43)]

May-Ru CHEN, *Institute of Statistical Science, Academia Sinica*

In the traditional red-and-black game, the player with positive integral fortune can stake any positive integral units in his possession, winning an amount equal to the stake with probability w $(0 < w < 1)$ and losing the stake with probability $1 - w$. The player seeks to maximize the probability of reaching a fixed fortune by gambling repeatedly with suitably chosen stakes.

In this talk, a brief survey about generalizations of red-and-black game to more than one person is given. For the two-person case, we show that, under some conditions, it is the unique Nash equilibrium for each player to adopt strategy σ of playing timidly or boldly according as the game is either in his favor or not (assuming the other also plays σ). For the three-person case, we also show that, under some conditions, the profile (σ, σ, σ) is a Nash equilibrium.

69 Inference of Principal Fitted Components (PFC) models in dimension reduction

[CS 26,(page 25)]

Xin CHEN, *School of Statistics, University of Minnesota, Minneapolis, USA*
R. Dennis COOK, *School of Statistics, University of Minnesota, Minneapolis, USA*

Construction of a few important regressors for the use in the regression without loss of information is very helpful especially when there are many predictors. There are many methods to reduce predictors dimension such as ordinary least square regression, partial least square regression, principal component regression, LASSO and the inverse regression like SIR and SAVE. Cook (2007) developed a new method - principal fitted components (PFC) models in sufficient dimension reduction area. In this presentation, I would like to talk our recent works such as robustness of estimators, prediction and variable selection based on PFC models.

70 Predicting bankruptcy using a discrete-time semiparametric hazard model

[CS 24,(page 24)]

Kuang Fu CHENG, *Biostatistics Center, China Medical University, Taichung, Taiwan, and Institute of Statistics, National Central University, Jhongli, Taiwan*
C.K. CHU, *Department of Applied Mathematics, National Dong Hwa University, Hualien, Taiwan*
Ruey-Ching HWANG, *Department of Finance, National Dong Hwa University, Hualien, Taiwan*

The usual bankruptcy prediction models are based on single period data of firms. These models ignore the fact that the characteristics of firms change through time, and thus they may suffer from the loss of prediction power. In recent years, a discrete-time parametric hazard model has been proposed for bankruptcy prediction using panel data. This model has been proved by many examples to be more powerful than the traditional models. In this paper, we have proposed an extension of this approach allowing for more flexible choice of hazard function. The new method does not require assuming a specific parametric hazard function in the analysis. It also provides a tool for checking the adequacy of the parametric model, if necessary. We use real panel datasets to illustrate the proposed method. The empirical results confirm that the new model compares favorably with the well-known discrete-time parametric hazard model.

71 Nonlinear regression analysis for models with incomplete data

[CS 45,(page 39)]

Tsung-Lin CHENG, *Department of Mathematics, National Changhua Unviersity of Education, Taiwan*
Hen-hui LUE, *Tunghai University, Taiwan*
Xuewen LU, *University of Calgary, Canada*
Jia-Wei SUN, *National Changhua Unviersity of Education, Taiwan*

In many regression problems the regressor is usually measured with errors. Errors in the predictors can cause severe bias in estimation unless some auxiliary adjustment has been made. Nonlinear measurement error models have received progressive attention since Carroll and Li's (1992) paper appeared. Besides, regression techniques, e.g. scatterplot, can suggest the forms of the models. But for censored responses, these techniques might fail. In order to resolve the problem caused by unobserved censored data, Fan and Gijbel (1994) proposed a Kaplan-Meier like approach. Based on regression calibration proposed by Carroll and Li (1992) in dealing with the measurement error model as well as a Kaplan-Meier like Transformation for censored data proposed by Fan and Gijbel (1994) in dealing with the nonparametric regression model with censored response, we may consider several general models for those defective data sets. In seeking to reach these objectives, we modify both of Lu and Cheng's (2007) and Lue's (2004) approach to simultaneously overcome the difficulty of estimations caused by censored responses and error-prone regressors. Moreover, we generalize both of their works. The illustrative simulation results verify the validity of our method.

72 Nonparametric estimation of cause-specific cross hazard ratio with bivariate competing risks data

[CS 28,(page 26)]

Yu CHENG, *Departments of Statistics and Psychiatry at the University of Pittsburgh*
Jason FINE, *Department of Biostatistics, University of North Carolina at Chapel Hill*

We propose an alternative representation of the cause-specific cross hazard ratio for bivariate competing risks data. The representation leads to a simple plug-in estimator, unlike an existing ad hoc procedure. The large sample properties of the resulting inferences are established using empirical process techniques. Simulations and a real data example demonstrate that the proposed methodology may substantially reduce the computational burden of the existing procedure, while maintaining similar efficiency properties.

73 Universal and M-S optimality criteria and efficiency of neighbouring design

[CS 22,(page 22)]

Santharam CHENGALVARAYAN, *Loyola College (Autonomous) Chennai, India*

This paper deals with the universal optimality, M-S optimality and efficiency of Nearest Neighbour Balanced Block Design (NBBD) using Generalized Least Square Method of estimation for different correlated models (AR(1), MA(1) and ARMA(1,1)) NNBD turns out to be universal optimal for AR(1) model and the performance of NNBD is quite satisfactory for remaining models whereas NNBDS turns out to be not MS optimal for all the correlated models. The efficiency of the proposed design in comparison to the regular block design is substantial for the above correlated models.

74 Profiling time course expression of virus genes — An illustration of Bayesian inference under shape restrictions

[CS 69,(page 52)]

Li-Chu CHIEN, *Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan*
I-Shou CHANG, *Institute of Cancer Research and Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan*
Chao A. HSIUNG, *Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan*

Because microarray experiments offer feasible approaches to the studies of genome-wide temporal transcriptional program of viruses, which are generally useful in the construction of gene regulation network, there have been many genome-wide expression studies of virus genes. They considered different viruses and/or different host cells and the samples were taken at different time points post infection. It seems that all the biological interpretations in these studies are directly based on the normalized data and crude statistics, which provide only naive estimates of limited features of the profile and may incur bias; in fact, there are some discrepancies reported in the literature. It is desirable that appropriate statistical methods can be developed and employed for these studies. Because some of these studies make use of different but related strains of viruses and/or different and related host cells, it is of great interests to compare transcriptional studies using similar/related strains of viruses or host cells.

Motivated by the above reason, the purpose of this paper is to illustrate a hierarchical Bayesian shape restricted regression method for the inference on the genome-wide time course expression of virus genes. The prior, introduced by Bernstein polynomials, takes into consideration the geometric constraints on the profile and has its parameters determined by data; the hierarchical modeling takes advantage of the correlation between the genes so as to enjoy the shrinkage effects. A customized Markov chain Monte Carlo algorithm is used to simulate the posterior distributions for inference. This method offers the possibility of comparing genome-wide expression studies with different designs, as long as there are enough of them to capture their respective main features. One specific advantage of this method is that estimates of many salient features of the expression profile like onset time, inflection point, maximum value, time to maximum value, etc. can be obtained immediately. Another specific advantage of this method is that it is especially useful in assessing the strength of the evidence provided by the data in favor of a hypothesis on the shape of the time course expression curve; for example, the hypothesis that it is unimodal.

75 Near optimality of dyadic random forests

[CS 40,(page 36)]

Choongsoon BAE, *Department of Statistics, U.C.Berkeley*
Peter BICKEL, *Department of Statistics, U.C.Berkeley*

Random Forests has been one of the most popular methods in both classification and regression. It is computationally fast and behaves well empirically. But there are few analytic results.

In this presentation, we introduce a modified version of Random Forests for regression problems. With aggregation by exponential weighting and applying oracle inequality, we show that it attains a near optimal convergence rate for a Lipschitz continuous regression function. We have also extended these results for general function classes.

76 The reliability of Type II censored reliability analyses for Weibull data

[CS 12,(page 13)]

See Ju CHUA, *School of Business & Economics, Swansea University, UK*

Alan John WATKINS, *School of Business & Economics, Swansea University, UK*

From a statistical viewpoint, the analysis of the complete data set is to be preferred, but, in practice, some censoring - such as Type I or Type II - is often inevitable; some items are expensive to test; some failures may take years to observe; and some experiments may be hazardous to run for prolonged periods. In practice, an experimenter may wish to know the smallest number of failures at which a trial can be reasonably or safely terminated, but where the censored analysis still provides a reliable guide to the analysis of the complete data. In this paper, we examine the roles of censoring number r and sample size n for Type II censoring.

It is particularly relevant in practical applications to make inferences on the running time for the experiment, or some percentile of lifetimes, since time is often directly linked to costs; for example, estimating the 10^{th} percentile of failure times. We illustrate the problem with reference to the classic ball-bearings data set, modelled (Kalbfleisch, 1979), by the Weibull distribution, with the percentile function $(0 < p < 1)$

$$B_p\left(\theta, \beta\right) = \theta \left\{-\ln\left(1-p\right)\right\}^{\frac{1}{\beta}} ,$$

where $\theta > 0$ and $\beta > 0$ are, respectively, the usual scale and shape parameters. With $n = 23$ items under test, we obtain the following maximum likelihood estimates under Type II censoring for various r; note that

$$\hat{B}_{0.1,r} = B_{0.1}\left(\hat{\theta}_r, \hat{\beta}_r\right) = \hat{\theta}_r\left(-\ln 0.9\right)^{\frac{1}{\hat{\beta}_r}}$$

converges to its complete counterpart as $r \to n$.

r	8	12	16	20	23
$\hat{\theta}_r$	67.641	75.217	76.696	78.967	81.878
$\hat{\beta}_r$	3.228	2.624	2.469	2.354	2.102
$\hat{B}_{0.1,r}$	33.686	31.906	30.833	30.356	28.069

We can now quantify the effect of censoring at, say, $r = 8$ and $r = 16$. For $r = 8$, the test stops after 51.84 million revolutions, while, with $r = 16$, we need to wait roughly 30 million revolutions longer. We can also assess the changes related to the final few failures by taking $r = 20$, when we intuitively expect estimates to be more consistent with final values than with $r = 8$ or 16. More generally, we can consider the precision with which we can make statements on final estimates, based on interim estimates. This approach requires an assessment of the extent to which $\hat{B}_{0.1,r}$ can be regarded as a reliable guide to $\hat{B}_{0.1,n}$, and hence we study the relationship between $\hat{B}_{0.1,n}$ and $\hat{B}_{0.1,r}$.

We develop some recent work (Chua & Watkins, 2007) on estimates of Weibull parameters, and present asymptotic results on the correlation between the two estimates of $B_{0.1}$; this, in turn, yields 95% confidence limits for the final estimate given the interim estimate. We illustrate our results using published data and simulation experiments, indicating the extent to which asymptotic results apply in samples of finite size.

77 Uniform design over convex input regions with applications to complex queueing models

[CS 29,(page 27)]

Shih-Chung CHUANG, *Graduate Institute of Statistics National Central University, Taiwan*
Ying-Chao HUNG, *Graduate Institute of Statistics National Central University, Taiwan*

The Uniform Design (UD) was first proposed in 1980 and has been recognized as an important space-filling design robust to model selections. Over the last two decades, applications have been found in diverse areas such as system engineering, pharmaceutics, military sciences, chemical engineering, quality engineering, survey design, computer sciences, and natural sciences. The basic idea of UD is to choose the input points in the experimental domain so that some measure of uniformity (such as discrepancy, dispersion, etc) is optimized. However, most UD methods introduced in literature are developed for input regions that can be reasonably transformed into a unit cube (e.g. rectangles). In this study, we propose a new measure of uniformity called "Central Composite Discrepancy" that is suitable for any convex input regions (such as convex polytope, ball, simplex, etc). This new approach has the following advantages: (i) it is easy to implement; (ii) the optimal design solution is invariant under coordinate rotation. We also introduce some direct applications of the proposed UD method to a complex queueing model.

78 Random recurrence equations and ruin in a Markov-dependent stochastic economic environment

[CS 33,(page 32)]

Jeffrey F. COLLAMORE, *University of Copenhagen*

The objective of this talk will be to explore random recurrence equations for Markov-dependent processes which arise in various applications, such as insurance mathematics and financial time series modeling.

We begin by considering an insurance company in a stochastic economic environment, where the investment returns are Markov-dependent, governed e.g. by an ARMA process or a stochastic volatility model. Such investment processes can be viewed as Markov chains in general state space. In this setting, we develop sharp large deviation asymptotics for the probability of ruin, showing that this probability decays at a certain polynomial rate, which we characterize.

Next, we briefly draw a connection of this result to a related problem in financial time series modeling, where it is of interest to characterize the tails of the standard ARCH(1) and GARCH(1,1) time series models, but with dependence in the driving sequence, as may occur e.g. under regime switching.

Our estimates build upon work of Goldie (Ann. Appl. Probab., 1991), who has obtained similar asymptotics applicable for independent sequences of random variables. Here we adopt a general approach for Harris recurrent Markov chains and develop, moreover, certain new recurrence properties for such chains based on a nonstandard “Gartner-Ellis¨-type assumption on the driving process.

79 The Vine-copula and Bayesian belief net representation of high dimensional distributions

[IS 15,(page 6)]

Roger M. COOKE, *Resources for the Future and T.U. Delft*
Dorota KUROWICKA, *T.U.Delft*

Regular vines are a graphical tool for representing complex high dimensional distributions as bivariate and conditional bivariate distributions. Assigning marginal distributions to each variable and (conditional) copulae to each edge of the vine uniquely specifies the joint distribution, and every joint density can be represented (non-uniquely) in this way. From a vine-copulae representation an expression for the density and a sampling routine can be immediately derived. Moreover the mutual information (which is the appropriate generalization of the determinant for non-linear dependencies) can be given an additive decomposition in terms of the conditional bivariate mutual informations at each edge of the vine. This means that minimal information completions of partially specified vine-copulae can be trivially constructed. The basic results on vines have recently been applied to derive similar representations for continuous, non-parametric Bayesian Belief Nets (BBNs). These are directed acyclic graphs in which influence (directed arcs) are interpreted in terms of conditional copulae. Interpreted in this way, BBNs inherit all the desirable properties of regular vines, and in addition have a more transparent graphical structure. New results concern ’optimal’ vine-copulae representations; that is, loosely, representations which capture the most dependence in the smallest number of edges. This development uses the mutual information decomposition theorem, the theory of majorization and Schur convex functions.

80 Statistics and Malliavin calculus

[CS 77,(page 57)]

José Manuel CORCUERA, *University of Barcelona*

When we do statistics in a Wiener space, for instance when observations come from a solution of stochastic differential equation driven by a Brownian motion, Malliavin Calculus can be used to obtain expressions for the score function as a conditional expectation. These expressions can be useful to study the asymptotic behavior of the model and estimators. For instance we can derive the local asymptotic normality property. We proceed from very simple examples to more complex ones where proceses under observation can have a jump component.

81 Bayesian partitioning for modelling residual spatial variation in disease risk based upon a matched case-control design.

[PS 3,(page 29)]

Deborah A COSTAIN, *Lancaster University*

Methods for modelling spatial variation in disease risk continue to motivate much research. Not least, knowledge of the dynamics of sub-regions in which the risk appears to be atypical can be used to filter resources, generate clues as to aetioliogy and guide further research.

In general, variation due to known risk factors, such as age and gender, is not of intrinsic interest and thus the ability to accommodate known confounders is paramount. In addition, methods which make few assumptions about the underlying form of the risk surface are motivated.

Assuming point referenced case-control data, confounding can be handled at the analysis stage, by means of covariate adjustment or stratification, or alternatively, at the design stage by using the confounders as stratifying factors for matching the cases and controls.

A consequence of matching is that the selection process needs to be accommodated in the analysis, which in turn results in consistency issues since the number of parameters increases with the sample size. In general, analysis thus proceeds on the basis of the matched conditional likelihood (Breslow and Day, 1980) in which the 'nuisance' parameters are eliminated.

This paper presents a Bayesian partition model formulation for the analysis of matched case control data. Random Voronoi tessellations are used to decompose the region of interest and a Poisson model re-formulation, which is proportional to the matched conditional likelihood, is used. Key features of the approach include a relaxation of the customary assumption of a stationary, isotropic, covariance structure, and moreover, the ability to detect spatial discontinuity.

Reversible Jump MCMC is used to sample from the posterior distribution; various methods, including transformation, are developed to provide more efficient means of candidate generation. The methodology developed maintains the conditional independence structure underpinning the partitioning approach, negating the need to monitor the locations of, and compute covariate differences between, the case and the m-1 matched controls in each matched group, at each iterative step. The methodology is also generalised to handle additional 'non-matched-for' covariate information.

Simulations demonstrate the capacity to recover known smooth and discontinuous risk surfaces and the estimated covariate effects were consistent with the underlying truth. Although computationally intensive run times are good.

The methodology is demonstrated on matched perinatal mortality data derived in the North-West Thames region of the UK. Infants were individually matched on the basis of sex and date-of-birth. A measure of social deprivation, Carstairs' index, known to be associated with infant death, was available and was included in the analysis.

As anticipated social deprivation was found to be a highly significant predictor of perinatal mortality. The unadjusted analysis (ignoring Carstairs' index) resulted in surface estimates which exhibited regions of atypical risk. When adjusted for Carstairs' index, however, there was no evidence of residual spatial variation in perinatal risk.

82 Using hierarchical Dirichlet process to classify high-dimensional data. Application to exposure to pesticides in French diet

[CS 11,(page 12)]

Amélie CRÉPET, *Afssa*
Jessica TRESSOU, *Inra-Met@risk and HKUST-ISMT*

This work introduces a specific application of the hierarchical Dirichlet process [Ferguson, 1973, Ann. Statist., 1:209–230] methodology in the food risk analysis context. Since different contaminant may have combined effects, namely interactions between substances may result in a greater toxic effect than the one predicted from individual substance assessments, it would be relevant to study the problem from a multivariate perspective. The main goal of this work is to determine groups of pesticides simultaneously present in the French diet at doses close to the acceptable ones (called Acute Reference Doses, ARfD) so as to build relevant toxicological experiments for measuring the possible combined effects of multiple residues of pesticides.

The exposure to the chemical is estimated from the available data, generally built from consumption data and contamination data since exposure to chemicals is very seldom investigated directly. In our pesticide study case, $P = 70$ different pesticides are considered, several residue levels being obtained from surveillance plans of the French administration for a wide range of food products. The "'INCA"' national survey on individual food consumptions provides the detailed consumptions of $n = 3003$ individuals on 7 days. For each individual i and each pesticide p, an empirical distribution (M values) of daily exposure is computed by randomly sampling all the food consumptions observed on a day (acute risk) and plausible pesticide residue levels of the different foods, accounting for the correlation among the different pesticides (through the Iman & Conover methodology). Finally, scaling problems are ruled out by dividing these exposures by the associated individual body weight and associated pesticide ARfD before passing to log scale. The final relative risk observations are viewed as multivariate functional data.

We propose to develop a hierarchical Bayesian

Nonparametric model, similar to the one proposed by [Teh et al, 2006, JASA, 101(416):1566–1581] with one extra hierarchical level due to the specific structure of our data. This approach has several advantages namely because no parametric assumption on the shape of the exposure distribution is required, the number of clusters is automatically determined through the estimation process. A stick-breaking approximation will be retained to account for the complexity of the hierarchy within an effective algorithm, [Ishwaran & James, 2001, JASA, 96:161–173].

83 Convergence of simple random walks on random discrete trees to Brownian motion on the continuum random tree

[IS 18,(page 19)]

David CROYDON, *University of Warwick*

It is well known that the continuum random tree can be defined as the scaling limit of a family of random graph trees. In this talk, I will discuss how the natural diffusion, or Brownian motion, on the continuum random tree arises as the scaling limit of the simple random walks associated with a suitably convergent family of random graph trees. In particular, the result provides a description of the scaling limit of the simple random walks associated with the family trees of Galton-Watson processes with a finite variance offspring distribution conditioned on the total number of offspring. Further properties of the Brownian motion on the continuum random tree will also be outlined.

84 Graphical modeling for discrete random variables with application to Tissue Microarray (TMA) experiments

[CS 69,(page 52)]

Corinne DAHINDEN, *Seminar for Statistics, ETH Zurich*
Peter BÜHLMANN, *ETH Zurich*

Tissue microarrays (TMA) are composed of hundreds of tissue sections from different patients arrayed on a single glass slide. With the use of immunohistochemical staining, they provide a high-throughput method of analyzing potential biomarkers on large patient samples. The assessment of the expression level of a biomarker is usually performed by the pathologist on a categorical scale.

The analysis of the interaction of these biomarkers and in particular causal relations are of biological importance. How are the pathways changing in cancer tissues compared to the fairly well understood pathways in cell lines? To address this question, we fit an L1-regularized log-linear model assuming a multinomial sampling scheme in order to obtain the conditional independence graph. The regularization becomes necessary as after cross-tabulation of the samples in contingency tables, many cell entries remain zero, leading to so-called sparse contingency tables, where standard procedures fail to work.

85 Testing for two states in a hidden Markov model

[CS 6,(page 8)]

Joern DANNEMANN, *Institute for Mathematical Stochastics, University of Goettingen*
Hajo HOLZMANN, *Institute for Stochastics, University of Karlsruhe*

We investigate likelihood inference, in particular the likelihood ratio test (LRT), for hidden Markov models (HMMs) in situations for which the standard regularity conditions are not satisfied. Specifically, we propose a test for two against more states of the underlying regime in an HMM with state-dependent distributions from a general one-parameter family. For HMMs, choosing the number of states of the underlying Markov chain, is an essential problem. Our test for this problem is an extension to HMMs of the modified likelihood ratio test for two states in a finite mixture, as proposed by Chen et al. (2004). Its asymptotic distribution theory under the null hypothesis of two states is derived. The test is based on inference for the marginal mixture distribution of the HMM. In order to illustrate the additional difficulties due to the dependence structure of the HMM, we also show how to test general regular hypotheses on the marginal mixture of HMMs via a quasi LRT. The testing procedures are illustrated by examples from economics and biology.

References

1. Chen, H., Chen, J. and Kalbfleisch, J. D. (2004). Testing for a finite mixture model with two components. J. Roy. Stat. Soc. Ser. B Stat. Methodol., 66, 95 – 115.
2. Dannemann, J. and Holzmann, H. (2007), Likelihood ratio testing for hidden Markov models under non-standard conditions. Scand. J. Statist. OnlineEarly.
3. Dannemann, J. and Holzmann, H. (2008), Testing

for two states in a hidden Markov model. Canad. J. Statist. In revision.

86 Estimating hazard rate of earthquake risk

[CS 3,(page 7)]

Sutawanir DARWIS, *Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
Agus Yodi GUNAWAN, *Industrial and Financial Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
I Wayan MANGKU, *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia*
Nurtiti SUNUSI, *Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
Sri WAHYUNINGSIH, *Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*

Predicting earthquake risk is one of research questions in seismic hazard modeling. Estimating hazard rate plays significant role and a large number of studies based on maximum likelihood of point process have been published in the literature. An alternative approach is significantly needed to contribute to estimation methodology. This proposed approach is based on hazard rate estimation developed in actuarial study adapted to earthquake risk estimation. The proposed approach will be applied to some data taken from an earthquake catalogue.

87 Bifurcation results of Nutrient-Phytoplankton-Zooplankton Model with delayed recycling

[CS 75,(page 56)]

Kalyan DAS, *B.P.Poddar Institute of Management & Technology,137, VIP Road, Kolkata:700 052,India.*

In the present study we consider a nutrient-based model of phytoplankton-zooplankton interaction with a delayed nutrient recycling and incorporating the average number of zooplankton attacking the phytoplankton follow Poission distribution (random zooplankton attack).We have derived the condition for asymptotic stability of the steady state and also estimated the length of the delay preserving the stability. The criterion for existence of Hopf-type small and large amplitude periodic oscillations of phytoplankton biomass and zooplankton numbers are derived. Finally, all the analytical results are interpreted ecologically and compared with the numerical results.

88 Model choice and data approximation

[IS 11,(page 23)]

P. L. DAVIES, *University of Duisburg-Essen, Germany, Technical University of Eindhoven, Holland, SFB 475 University of Dortmund, Germany, Eurandom, Eindhoven, Holland.*

The talk will concentrate on the problem of model choice in situations where the model could in principle have generated the data. Most, but not all, paradigms of model choice involve balancing the fidelity of the model to the data against some measure of complexity of the model. This is accomplished by minimizing an expression involving a real-valued measure of fidelity and a real-valued measure of complexity. Such an approach restricts the models to a certain chosen class of models and does not allow for the possibility that no model in the class is adequate. An approach to model choice based on a concept of data approximation will be expounded: a model will be said to approximate a data set if typical data sets generated under the model look like the real data set. The words ‘typical’ and ‘look like’ will be operationalized so that the concepts can be applied to data sets. In particular it allows for the possibility that no model in a class of models is adequate for the data. The approach will be exemplified in the context of non-parametric regression.

89 Upper estimates for Pickands constants

[CS 61,(page 47)]

Krzysztof DEBICKI, *Mathematical Institute, University of Wroclaw, Poland*
Pawel KISOWSKI, *Mathematical Institute, University of Wroclaw, Poland*

Pickands constants H_{B_α} play a significant role in the extreme value theory of Gaussian processes. Recall that

$$\mathrm{H}_{B_\alpha} := \lim_{T\to\infty} \frac{\mathrm{E}\,\exp\left(\sup_{t\in[0,T]}(\sqrt{2}B_\alpha(t)-t^\alpha)\right)}{T},$$

where $\{B_\alpha(t), t \ge 0\}$ is a fractional Brownian motion with Hurst parameter $\alpha/2$ and $\alpha \in (0,2]$.

In this talk we present new upper bounds for H_{B_α} and $\alpha \in (1,2]$.

90 Multifractal embedded branching processes

[CS 66,(page 49)]

G. DECROUEZ, *The University of Melbourne*
Owen D. JONES, *Dept. of Mathematics and Statistics, University of Melbourne*

On-line simulation of self-similar and multifractal processes is problematic because of the long-range dependence inherent in these processes. In the presence of long-range dependence, simulation of $X(n+1)$ given $X(1), \ldots, X(n)$ implicitly requires the full conditional distribution, which becomes impractical when n is large. In other words, it is generally not possible to simulate $X(n+1)$ once the first n samples have been generated.

We propose a new class of discrete-scale invariant process, called Multifractal Embedded Branching Processes (MEBP), which can be efficiently simulated on-line. MEBP are defined using the crossing tree, an ad-hoc space-time description of the process, and are characterized as processes whose crossing tree is a Galton-Watson branching process. For any suitable branching process there is a family of discrete-scale invariant process—identical up to a continuous time change—for which it is the crossing tree. We identify one of these as the canonical Embedded Branching Process (EBP), and then construct MEBP from it using a multifractal time change. To allow on-line simulation of the process, the time change is constructed from a multiplicative case on the crossing tree.

Time-changed self-similar signals, in particular time-changed Brownian motion, are popular models in finance. Multiscale signals also find applications in telecommunications and turbulence. Brownian motion can be constructed as a canonical EBP, so MEBP include a class of time changed Brownian motion, suggesting their suitability for modeling purposes.

Proofs of the existence and continuity of MEBP are given, together with an efficient on-line algorithm for simulating them (a Matlab implementation is freely available from the web page of Jones). Also, using an approach of Riedi, an upper bound on the multifractal spectrum of the time change is derived.

91 Semiparametric regression with kernel error model

[CS 36,(page 33)]

Jan G. DE GOOIJER, *University of Amsterdam, The Netherlands*
Ao YUAN, *Howard University, USA*

We propose and study a class of regression models, in which the mean function is specified parametrically as in the existing regression methods, but the residual distribution is modeled nonparametrically by a kernel estimator, without imposing any assumption on its distribution. This specification is different from the existing semiparametric regression models. The asymptotic properties of such likelihood and the maximum likelihood estimate (MLE) under this semiparametric model are studied. The nonparametric pseudo-likelihood ratio has the Wilks property as the true likelihood ratio does. Simulated examples are presented to evaluate the accuracy of the proposed semiparametric MLE method. Next, the problem of selecting the optimal functional form among a set of non-nested nonlinear mean functions for a semiparametric kernel based regression model is considered. To this end we apply Rissanen's minimum description length (MDL) principle. We prove the consistency of the proposed MDL criterion. Its performance is examined via simulated data sets of univariate and bivariate nonlinear regression models.

92 Student's t-statistic for martingales

[CS 34,(page 32)]

Victor H. DE LA PENA, *Department of Statistics Columbia University*

In this talk we introduce an inequality for self-normalized martingales that can be viewed as an extension of Student's (1908) t-Statistic and provide an application to testing for the variance of a moving average.

93 Trimming methods in model checking

[CS 17,(page 19)]

Eustasio DEL BARRIO, *Universidad de Valladolid*

This talk introduces an analysis of similarity of distributions based on measuring some distance between trimmed distributions. Our main innovation is the use of the impartial trimming methodology, already considered in robust statistics, which we adapt to the setup of model checking. By considering trimmed probability measures we introduce a way to test whether the core of the random generator underlying the data fits a given pattern. Instead of simply removing mass at non-central zones for providing some robustness to the similarity analysis, we develop a data-driven trimming method aimed at maximizing similarity between distributions. Dissimilarity is then measured in terms of the distance between the optimally trimmed distributions. Our main choice for applications is the Wasserstein metric, but other distances might be of interest for different applications. We provide illustrative examples showing the improvements over previous approaches and give the relevant asymptotic results to justify the use of this methodology in applications.

94 On the epidemic of financial crises: a random graph approach

[CS 33,(page 32)]

Nikolaos DEMIRIS, *MRC Biostatistics Unit, Cambridge, UK*
L. V. SMITH, *Centre for Financial Analysis and Policy, University of Cambridge, UK*

The propagation of financial crises has generated a considerable amount of interest recently. However, the contagion literature does not directly model the inherent dependencies in the spread of crises. We propose a stochastic epidemic model where the population of countries is explicitly structured and the crisis may be transmitted locally and globally. The approach will be illustrated using historical data. The likelihood for such data is numerically intractable and we surmount the problem by augmenting the parameter space with a random graph that describes the underlying contact process. The results indicate an increasing trend for global transmission over time. Policy implications and model determination issues shall also be described.

95 Steady-state Lévy flights in a confined space

[PS 2,(page 17)]

S. I. DENISOV, *Institut für Physik, Universität Augsburg, Universitätsstraße 1, D-86135 Augsburg, Germany*
Werner HORSTHEMKE, *Department of Chemistry, Southern Methodist University, Dallas, Texas 75275-0314, USA*
Peter HÄNGGI, *Institut für Physik, Universität Augsburg, Universitätsstraße 1, D-86135 Augsburg, Germany*

For many physical, biological, social and other systems that interact with a fluctuating environment, the temporal evolution of the relevant degrees of freedom can be described by the Langevin equation

$$\dot{x}(t) = f(x(t),t) + g(x(t),t)\xi(t)$$

driven by a white noise $\xi(t)$. We show that the generalized Fokker-Planck equation associated with this equation can be written in the form

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}f(x,t)P(x,t) + \mathcal{F}^{-1}\{P_k(t)\ln S_k\}.$$

Here, $u_k = \mathcal{F}\{u(x)\}$ and $u(x) = \mathcal{F}^{-1}\{u_k\}$ are the Fourier transform pair, and S_k is the characteristic function of the white noise generating process at $t = 1$. For Lévy flights, i.e., stable Lévy processes, Eq. (2) turns into a fractional Fokker-Planck equation, which we solve analytically for a confined domain. We find that steady-state Lévy flights in an infinitely deep potential well are distributed according to the beta probability density,

$$P(x) = (2d)^{1-\alpha}\frac{(d+x)^{-\nu}(d-x)^{-\mu}}{B(1-\nu,1-\mu)},$$

where $2d$ is the well width, $B(a,b)$ is the beta function, and the exponents ν and μ are related to the parameters of the stable distribution, viz., the index of stability α and the skewness parameter. Our results clarify the origin of the preferred concentration of flying objects near the boundaries of the well.

96 On infinite divisibility of probability distribution with density function of normed product of multi-dimensional Cauchy densities

[CS 10,(page 12)]

Dodi DEVIANTO, *Department of Mathematics, Faculty of Science, Ibaraki University, JAPAN*
Katsuo TAKANO, *Department of Mathematics, Faculty of Science, Ibaraki University, JAPAN*

From probability distributions with density function of normed product of the Cauchy densities, it is constructed a mixing density of the d-dimensional distribution

$$g(d;u) = \frac{c\pi^{d/2}}{u^{d/2+2}} \cdot \exp\{-\frac{b^2}{u}\} \int_0^1 \exp\{\frac{(b^2-a^2)v}{u}\} \cdot v^{(d+1)/2-1}(1-v)^{(d+1)/2-1}dv,$$

where $0 < a < b$, d is dimension and c is a constant. We conjecture that its Laplace transform does not have zeros in the complex plane except at the origin and therefore the corresponding probability distribution is infinitely divisible.

97 The application of neural networks model in forcasting oil production based on statistical inference: a comparative study with VAR and GSTAR models

[CS 70,(page 52)]

Urwatul Wutsqa DHORIVA, *Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, UNY, Indonesia*
SUBANAR, *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Indonesia*
Guritno SURYO, *Department of Mathematics, Gadjah Mada University, Indonesia*
Soejoeti ZANZAWI, *Department of Mathematics, Gadjah Mada University, Indonesia*

This article aims to investigate an appropriate space and time series model for oil production forecasting. We propose a new method using neural networks (NN) model based on the inference of R2 incremental contribution to solve that problem. The existing studies in NN modeling usually use descriptive approach and focus on univariate case, whereas our method has employed statistical concepts, particularly hypothesis test and has accommodated multivariate case including space and time series. This method is performed on bottom up or forward scheme, which starts from empty model to gain the optimal neural networks model. The forecast result is compared to those from the linear models GSTAR (Generalized Space-Time Autoregressive) and VAR (Vector Autoregressive). We show that our method outperforms to these statistical techniques in forecasting accuracy. Thus, we suggest that the NN model is the proper model for oil production forecasting.

98 Multilevel functional principal component analysis

[CS 64,(page 48)]

Chongzhi DI, *Department of Biostatistics, Johns Hopkins University*
Ciprian M. CRAINICEANU, *Department of Biostatistics, Johns Hopkins University*

We develop the framework of multilevel functional principal component analysis (MFPCA) for multilevel functional data, which contains multiple functions per subject. Functional principal component analysis (FPCA) is a key technique to characterize modes of variations and reduce dimensions for functional data. However, it is designed for a sample of independent functions so far. Possible correlation among functions in multilevel functional data complicates principal component analysis. We propose MFPCA based on a combination of functional analysis of variance and standard FPCA. The idea is to decompose the total variation into between subject and within subject variations, and carry out principal component analysis at both levels. Two methods of estimating principal component scores are proposed and compared. Our research was motivated by the Sleep Heart Health Study, which contains electroencephalographic (EEG) signal series for each subject at two visits. The proposed methodology is general, with potential broad applicability to many modern high-throughput studies.

99 Umbral methods in statistics

[CS 15,(page 15)]

Elvira DI NARDO, *Dipartimento di Matematica e Informatica, Università degli Studi della Basilicata, Viale dell'Ateneo Lucano 10, I-85100 Potenza, Italy*
Domenico SENATO, *Dipartimento di Matematica e Informatica, Università degli Studi della Basilicata, Viale dell'Ateneo Lucano 10, I-85100 Potenza, Italy*

The talk presents the classical umbral calculus as a precision algebraic tool for handling moments, cumulants and factorial moments of a random variable. The symbolic expressions we obtain for the corresponding unbiased estimators are particularly suited to be implemented with some specialized package, like MAPLE, speeding up existing procedures. These results are contained in two forthcoming papers: Di Nardo E., G. Guarino, D. Senato An unifying framework for k-statistics, polykays and their generalizations, Bernoulli, and Di Nardo E., G. Guarino, D. Senato On symbolic computation of moments of sampling distribution. Comp. Stat. Data Analysis.

The basic device of the umbral calculus is to represent an unital sequence of numbers by a symbol α, named umbra, that is, to associate the sequence $1, a_1, a_2, \ldots$ to the sequence $1, \alpha, \alpha^2, \ldots$ of powers of α through an operator E that resembles the expectation operator of a random variable. Dealing a number sequence by means of a symbol has been used in a variety of mathematical subjects since 1964, due to its founder G.C. Rota.

Recently, the method has led to a finely adapted language for random variables. This because an umbra looks as the structure of a random variable but with no reference to a probability space. We have

focused the attention on cumulant sequences because a random variable is often better described by its cumulants than by its moments, as it happens for the family of Poisson random variables. Moreover, due to the properties of additivity and invariance under translation, cumulants are not necessarily connected with moments of any probability distribution. The notion of umbra indexed by a multiset allows to umbrally represent multivariate moments and multivariate cumulants so that many classical identities can be recovered by simply characterizing a suitable multiset indexing.

As applications, we show how the classical umbral calculus provides a unifying framework for k-statistics, the unbiased estimators of cumulants, and their multivariate generalizations by a really simplified symbolic representation of some symmetric polynomials in the data points. These symmetric polynomials allow to handle algebraic expressions such as moments of sampling distributions in such a way that the overall complexity appear to be reduced with a substantial advantage in computational time. The new algorithm we propose to compute k-statistics and their generalizations turns to be very fast when compared with the ones of Andrews and Stafford (see Symbolic Computation for Statistical Inference. Oxford Univ.Press 2000) and the ones of MATHSTATICA.

100 Parameters of stochastic diffusion processes estimated from observations of first hitting-times: application to the leaky integrate-and-fire neuronal model

[IS 32,(page 36)]

Susanne DITLEVSEN, *Department of Mathematics, University of Copenhagen*
Petr LANSKY, *Institute of Physiology, Academy of Sciences of the Czech Republic*
Ove DITLEVSEN,

The first hitting-time to a constant threshold of a diffusion process has been in focus for stochastic modeling of problems where a hidden random process only shows when it reaches a certain level that triggers some observable event. Applications come from various fields, e.g. neuronal modeling, survival analysis and mathematical finance. In many applications where renewal point data are available, models of first hitting-times of underlying diffusion processes arise. Despite of the seemingly simplicity of the model, the problem of how to estimate parameters of the underlying stochastic process has resisted its solution. The few attempts have either been unreliable, difficult to implement or only valid in subsets of the relevant parameter space. In this talk a newly developed estimation method that overcomes these difficulties is presented, it is computationally easy and fast to implement, and also works surprisingly well on small data sets. It is a direct application of the Fortet integral equation. The method is illustrated on simulated data and applied to recordings of neuronal activity.

101 Asymptotical properties of heterogeneity test statistics with random sample size

[PS 1,(page 4)]

A. A. DJAMIRZAEV, *National University of Uzbekistan, Tashkent/Uzbekistan*

Let $X_N = (x_1, x_2, \dots, x_N)$ and $Y_N = (y_1, y_2, \dots, y_N)$ be two independent samples which consist of N independent observations and $N = N_n$ is a positive integer-valued random variable. Here, X_N and Y_N are samples with random size from distribution $L(\xi)$ with distribution function (d.f.) $F_1(x)$ and $L(\eta)$ with d.f. $F_2(x)$ respectively. Let $\nu_N = (\nu_1, \nu_2, \dots, \nu_k)$ and $\mu_N = (\mu_1, \mu_2, \dots, \mu_k)$ be frequency vector corresponding for sample points $x_1, x_2, \dots, x_N$ and $y_1, y_2, \dots, y_N$ that correspond to intervals of $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k$ groups that have no points in common. Here, $\nu_1 + \nu_2 + \dots + \nu_k = N$ and $\mu_1+\mu_2+\dots+\mu_k = N$. We shall consider the following statistics:

$$\chi^2_N = \sum_{i=1}^{k} \frac{(\nu_i - \mu_i)^2}{\nu_i + \mu_i}$$

It's known [1] that when $P\{N = n\} = 1$ holds and $H_0 : F_1(x) = F_2(x)$ true, then:

$$\lim_{n\to\infty} P\{\chi^2_n < x\} = H_{k-1}(x).$$

Where, $H_{k-1}(x)$ - chi-square distribution with $k-1$ degrees of freedom.

Theorem 1. Let $\frac{N}{n} \xrightarrow{P} N_0$ when $n \to \infty$, where N_0 - positive random variable. Then under H_0, $\forall A$ with $P(A) > 0$:

$$\lim_{n\to\infty} P\{\chi^2_N < x|A\} = H_{k-1}(x).$$

From Theorem 1, it follows that χ^2_n and χ^2_N is the mixing sequence of random variables by means of A. Renyi with limit d.f. $H_{k-1}(x)$ (refer to [2]).

Theorem 2. In the condition of Theorem 1, $\forall x \in R$ and $\forall y \in C(F)$:

$$\lim P\{\chi_N^2 < x, \quad \frac{N}{n} < y\} = H_{k-1}(x) \cdot P\{N_o < y\}.$$

References

1. G.I. Ivchenko, Yu.I. Medvedev. "Mathematicheskaya statistika". Moskva, Vys'shaya shkola, - 1984. (in Russian)
2. A.A. Djamirzaev. "About property of mixing by means of A. Renyi for the number of positive sums". Acta Scientarum Math., v.41 (1979) pp.47-53

102 Regularity of SPDEs in bounded domains and analysis of first exit times for non-Markov Ito processes

[CS 77,(page 57)]

Nikolai DOKUCHAEV, *Department of Mathematics, Trent University*

Problems arising for non-Markov processes and their first exit times will be discussed. It is suggested a way of obtaining the representation theorems for non-Markov processes via backward SPDEs (stochastic partial differential equations). For this purpose, some new regularity is obtained for solutions of backward SPDEs. More precisely, an analog of the second energy inequality and the related existence theorem are obtained for equations in domains with boundary. In particular, we establish conditions of differentiability of the diffusion coefficient that is included into the solution of backward SPDEs. These results lead to many applications for non-Markov processes in domains.

103 Power of tests for multivariate normality based on moments

[PS 1,(page 4)]

Czeslaw DOMANSKI, *University of Lodz*
Dariusz PARYS, *University of Lodz*

There are many methods of construction of multivariate normality tests. The current review of the literature proves that there are at least 60 procedures of verification of the hypothesis about multivariate normality of variable and random distributions. We can indicate a few factors which prove an analysis of this class

104 Uniform in bandwidth consistency of the kernel-based Hill estimator

[CS 49,(page 42)]

Julia DONY, *Free University of Brussels (VUB)*

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random variables in $\mathbf{R}^d \times \mathbf{R}$, and denote the common density function of X by f_X. We start by considering a kernel estimator $\hat{\varphi}_{n,h}(t)$, where $t \in \mathbf{R}^d$ is fixed and $\varphi : \mathbf{R} \to \mathbf{R}$ is a measurable function with finite second moment. If $h \equiv h_n$ is a deterministic sequence such that $h_n \to 0$ and $nh_n^d / \log\log n \to \infty$, it is well–known that $\hat{\varphi}_{n,h_n}(t)$ estimates consistently $m_\varphi(t) f_X(t)$, where $m_\varphi(t) = \mathbb{E}[\varphi(Y)|X = t]$ is the regression function. As an extension, we recall a result in which additional assumptions are imposed to make $\hat{\varphi}_{n,h}(t)$ a consistent estimator uniformly for a certain range of bandwidths $a_n \leq h \leq b_n$. As an application, we consider real valued random variables $Y_1, \ldots, Y_n$ for which the common distribution function has regularly varying upper tails of exponent $-1/\tau < 0$, and study the asymptotic behavior of a kernel–based version of the Hill estimator for the tail index τ, defined as

$$\hat{\tau}_{n,h} := \frac{\frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{j}{nh}\right)}{\sum_{j=1}^{n} \frac{j}{nh} K\left(\frac{j}{nh}\right)\{\log Y_{n-j+1:n} - \log Y_{n-j:n}\}},$$

where $Y_{j:n}, j = 1, \ldots, n$ denote the order statistics of $Y_1, \ldots, Y_n$, and where K is a kernel function and h is the bandwidth. The results that we obtain for the process $\hat{\varphi}_{n,h}(t)$ are used to establish the weak consistency of this estimator, uniformly for a certain range of bandwidths tending to zero at particular rates. This "uniform in bandwidth" result permits to consider estimators $\hat{\tau}_{n,h}$ based upon data–dependent bandwidths or bandwidths depending on the location.

105 Choosing optimal penalty term for Markov chain order estimator

[PS 1,(page 3)]

Chang C. Y. DOREA, *Universidade de Brasilia*

The Efficient Determination Criterion (EDC) generalizes the AIC and BIC criteria and provides a class of consistent estimators for the order of a Markov chain. Several authors have addressed the question of choosing between AIC or BIC estimates. The first tends to overestimate the order and the second, though consistent, may lead to underestimation. All these estimators are based on penalized maximum log-likelihood functions. In this work, we study the choice of the optimal penalty term and show that corresponds to neither AIC nor BIC estimators. We prove the strong consistency for EDC without assumption of a finite upper bound and provide the optimal choice for a class of estimators.

106 Particle Markov chain Monte Carlo

[IS 10,(page 18)]

Arnaud DOUCET, *Department of Statistics, University of British Columbia, Canada*
Christophe ANDRIEU, *Department of Mathematics, University of Bristol, UK*
Roman HOLENSTEIN, *Department of Computer Science, University of British Columbia, Canada*

Markov chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) methods have emerged as the two main tools to sample from high-dimensional probability distributions. Although asymptotic convergence of MCMC algorithms is ensured under weak assumptions, the performance of these latters is unreliable when the proposal distributions used to explore the space are poorly chosen and/or if highly correlated variables are updated independently. We show here how it is possible to build efficient high-dimensional proposal distributions using SMC methods. This allows us to design effective MCMC algorithms in complex scenarios where standard strategies fail. We demonstrate these algorithms on a nonlinear non-Gaussian state-space model, a stochastic kinetic model and Dirichlet process mixtures.

107 First passage densities and boundary crossing probabilities for diffusion processes

[CS 20,(page 20)]

Andrew DOWNES, *The University of Melbourne*
Konstantin BOROVKOV, *The University of Melbourne*

Calculating the probability $P(-\infty, g)$ that a diffusion process $\{X_t\}$ will stay under a given curvilinear boundary g (or the probability $P(g_-, g_+)$ that it will stay between two such boundaries $g_- < g_+$) during a given time interval $[0,T]$ $(T < \infty)$ is a classical problem of great importance for applications including financial mathematics (pricing barrier type options) and sequential analysis. Since no closed-form solutions are known except for a limited number of special cases and numerical computations based on solving respective boundary problems for PDE's are rather tedious, finding approximate solutions is of substantial interest. One possible approach is to approximate the given curvilinear boundaries $g_\pm$ with close ones, $f_\pm$, of a form that makes the computation of $P(f_-, f_+)$ feasible (for example, with piecewise linear $f_\pm$ when $\{X_t\}$ is a Wiener process, see e.g. Borovkov and Novikov (2005)). To use that approximation, one needs to control the approximation error

$$|P(g_-, g_+) - P(f_-, f_+)|.$$

The main results of the paper include upper bounds for (107). We show that under mild regularity conditions the difference between the probabilities does not exceed a multiple of the uniform distance between the original and approximating boundaries, the coefficient being an explicit function of the Lipschitz coefficients of the boundaries.

Much work has also been done in the area of calculating the density of the first crossing time. Explicit formulae can be obtained for a limited number of specific pairs of diffusions and boundaries. In the process of deriving an upper bound for (107), we establish the existence of the first crossing time densities and provide new sharp bounds for them. In the case when the diffusion interval coincides with the whole real line, we are we are able to establish upper and lower bounds for the density for both upper $(g(0) > X_0)$ and lower $(g(0) < X_0)$ boundaries. For diffusions restricted to $(0, \infty)$ we establish upper bounds for upper boundaries first crossing time densities. Employing a similar approach enables us to obtain further interesting results, including new sharp bounds for transition densities of diffusion process.

Acknowledgement. This research was supported by the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems.

108 Discrete chain graph models

[IS 7,(page 5)]

Mathias DRTON, *University of Chicago*

The statistical literature discusses different types of Markov properties for chain graphs that lead to four possible classes of chain graph Markov models. The different models are rather well-understood when the observations are continuous and multivariate normal, and it is also known that one class, referred to as models of LWF (Lauritzen-Wermuth-Frydenberg) or block concentration type, yields discrete models for categorical data that are smooth. We consider the structural properties of the discrete models based on the three alternative Markov properties. It is shown by example that two of the alternative Markov prop-

erties can lead to non-smooth models. The remaining model class, which can be viewed as a discrete version of multivariate regressions, is proven to comprise only smooth models. The proof employs a simple change of coordinates that also reveals that the model's likelihood function is unimodal if the chain components of the graph are complete sets.

109 Probability problems arising from genetics and ecology : philosophy and anecdotes.

[Wald Lecture 1,(page 3)]

Richard DURRETT, *Department of Mathematics, Cornell University*

The Wald Lectures provide me with an opportunity to reflect on 20 years of trying to use probability to shed light on questions that arise from biology. The philosophical question I will address is: what is good applied probability? To attempt to answer this question and to describe some of the challenges that come from trying to publish work that mathematicians think is trivial and biologists find incomprehensible, I will describe some of the research I have done recently with Nathanael Berestycki, Deena Schmidt, and Lea Popovic.

110 Probability problems arising from genetics and ecology : recent work in genetics with Jason Schweinsberg.

[Wald Lecture 2,(page 28)]

Richard DURRETT, *Department of Mathematics, Cornell University*

The first half of the talk concerns the problem of approximating selective sweeps, which in turn leads to a consideration of coalescents with multiple collisions and their application to species with heavy-tailed family sizes (e.g., marine species and domesticated animals). In the second half, we will cover work done with Deena Schmidt on regulatory sequence evolution and a model for multistage carcinogenesis, showing the power of mathematical abstraction to realize that these two applications are special cases of one problem in population genetics. How long do we have to wait until some member of the population has experienced a prespecified sequence of m mutations?

111 Probability problems arising from genetics and ecology : coexistence in stochastic spatial models.

[Wald Lecture 3,(page 46)]

Richard DURRETT, *Department of Mathematics, Cornell University*

For the much of the last twenty years I have worked to understand when competing species can coexist. An answer is provided by the competitive exclusion principle: the number of coexisting species cannot exceed the number of resources. Unfortunately, it is often not clear how many resources a system has. I will describe five examples, being with the Cornell Ph.D. theses of Claudia Neuhauser and Glen Swindle, work on bacterial competition models with Simon Levin, and ending with recent research with Nicolas Lanchier and Ben Chan.

112 Hoeffding decompositions and urn sequences

[CS 80,(page 58)]

Omar EL-DAKKAK, *Université Paris VI, Laboratoire de Statistique Théorique et Appliquée*
Giovanni PECCATI, *Université Paris VI, Laboratoire de Statistique Théorique et Appliquée*

Let $X_1, X_2, ...,$ be a sequence of random variables. The sequence is said to be Hoeffding decomposable if, for all $n \geq 2$, every symmetric statistic $T(X_1, ..., X_n)$ admits a unique representation as a sum of $n+1$ uncorrelated U-statistics. Introduced in the pioneering work of Hoeffding (see Hoeffding [1948]), Hoeffding decompositions are one of the central tools for establishing distributional limit results. They have been widely studied for i.i.d. sequences and have found many applications. In the dependent case, only extractions without replacement have been studied (see Bloznelis and Götze [2001, 2002]), until Peccati [2004] extended the theory of Hoeffding-decompositions to general exchangeable sequences with values in a Polish space. In this last reference, a characterization of Hoeffding decomposable exchangeable sequences has been obtained in terms of the notion of weak independence. While offering deep insights into the structure of Hoeffding decomposable exchangeable sequences, the work of Peccati left the following crucial question unanswered: can one characterize the Hoeffding-decomposable exchangeable sequences in terms of their de Finetti measure? In a recent paper (El-Dakkak, Peccati [2008]), we focus on the case of exchangeable sequences with values in a finite set D. When $D = \{0, 1\}$, we show that the sequence is Hoeffding decomposable if, and only if, it is either i.i.d. or a two-colour Pólya sequence. In the case in which D is an arbitrary finite set (of cardinality m), we obtain a partial generalization of this result, namely that if the the directing measure (de Finetti measure) of the sequence is the law of a vector of normalized positive infinitely divisible random variables, then the sequence is Hoeffding decomposable if, and only if, it is an m-colour Pólya sequence.

References

1. M. Bloznelis et F. Götze (2001). Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics. *The Annals of Statistics* **29** (3), 353-365.
2. M. Bloznelis and F. Götze (2002). An Edgeworth expansion for finite population statistics. *The Annals of Probability* **30**, 1238-1265.
3. O. El-Dakkak and G. Peccati (2008). Hoeffding decompositions and urn sequences. To appear in: *The Annals of Probability.*
4. W. Hoeffding (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**(2), 293-325.
5. G. Peccati (2004). Hoeffding-ANOVA decompositions for symmetric statistics of exchangeable observations. *The Annals of Probability,* **32** (3A), 1796-1829.

113 Some interactions between random matrices and statistics

[IS 16,(page 55)]

Noureddine EL KAROUI, *Department of Statistics, UC Berkeley*

Many widely used methods of multivariate statistics rely at their core on spectral decompositions of certain matrices. Nowadays, it is not uncommon to encounter in practice data matrices for which the number of variables (p) is of the same order of magnitude as the number of observations (n). In this setting, results from random matrix theory become relevant to theoretical statistics. In this talk, I will talk about the interaction between random matrix theory and statistics. More specifically, I will talk about estimation problems: how random matrix results can be used in practice, and how random matrix ideas can help frame problems in theoretical statistics. Time permitting, I will also speak about potential statistical limitations of some of the most widely studied random matrix models.

114 Impact of dimensionality and independence learning

[Laplace Lecture,(page 60)]

Jianqing FAN, *Princeton University*
Yingying FAN, *Harvard University*

Model selection and classification using high-dimensional features arise frequently in many contemporary statistical studies such as tumor classification using microarray or other high-throughput data. The impact of dimensionality on classifications is largely poorly understood. We first demonstrate that even for the independence classification rule, classification using all the features can be as bad as the random guessing due to noise accumulation in estimating population centroids in high-dimensional feature space. In fact, we demonstrate further that almost all linear discriminants can perform as bad as the random guessing. Thus, it is paramountly important to select a subset of important features for high-dimensional classification, resulting in Features Annealed Independence Rules (FAIR). The connections with the sure independent screeing (SIS) and iterative SIS(ISIS) of Fan and Lv (2007) in model selection will be elucidated and extended. The methods essentially utilize the concept of correlation learning. Further extension of the correlation learning results in independence learning for feature selection in general loss functions. The choice of the optimal number of features, or equivalently, the threshold value of the test statistics are proposed based on an upper bound of the classification error. Simulation studies and real data analysis support our theoretical results and demonstrate convincingly the advantage of our new classification procedure.

115 Some notes about Gaussian mixtures

[CS 6,(page 8)]

M. M. FELGUEIRAS, *CEAUL e Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Portugal*
D. D. PESTANA, *CEAUL e Faculdade de Ciências da Universidade de Lisboa, Portugal*

We investigate Gaussian mixtures with independent components, whose density function is

$$f_X(x) = \sum_{j=1}^{N} w_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right\}, \quad \sigma_j > 0,$$

where $w_j > 0$ and $\sum_{j=1}^{N} w_j = 1$, very effective in modeling real data since they can accomodate multimodality and heavier tailweight than the Gaussian. Our main purpose is to identify useful approximations to F_X, namely of the Pearson family.

While the general problem is hard, restrictions such as equal means or equal variances may lead to useful results. When $\mu_j = \mu$, for $i = 1, .., N$, the mixture can be approximated by a shifted t-student

$$\alpha X \overset{\circ}{\sim} t_{(n)} + \mu^*.$$

This can be used, as well, to develop a test for the equality of the means.

When $\sigma_j^2 = \sigma^2$, for $i = 1, .., N$, a rather trivial result allows us to write the mixture as a convolution of a zero mean Gaussian with a discrete variable. Some interesting results arise when special discrete distributions are considered. For instance, a convolution of Poisson distribution with a zero mean Gaussian is infinitely divisible, and can be seen has an infinite Gaussian mixture. For the simplest case, unimodal mixture of two subpopulations with the same variance,

$$X \overset{\circ}{\sim} Beta(a, b, p, q).$$

for $w \in [0.2113; 0.7887]$. This can be used to test $\sigma_1^2 = \sigma_2^2$.

116 Indirect genomic effects on survival from gene expression data

[CS 9,(page 11)]

Egil FERKINGSTAD, *Statistics for Innovation, Norwegian Computing Center, and Department of Biostatistics, University of Oslo, Norway*
Arnoldo FRIGESSI, *Statistics for Innovation, Norwegian Computing Center, and Department of Biostatistics, University of Oslo, Norway*
Heidi LYNG, *Department of Radiation Biology, Institute for Cancer Research, Norwegian Radium Hospital, Oslo, Norway*

In cancer, genes may have indirect effects on patient survival, mediated through interactions with other genes. Methods to study the indirect effects that contribute significantly to survival are not available. We propose a novel methodology to detect and quantify indirect effects from gene expression data. We discover indirect effects through several target genes of transcription factors in cancer microarray data, pointing to genetic interactions that play a significant role in tumor progression.

117 Weak link correction for a graph based spatial scan cluster detection algorithm

[CS 82,(page 59)]

Sabino J. FERREIRA, *Statistics Department, Universidade Federal de Minas Gerais, Brazil*
Luiz DUCZMAL, *Statistics Department, Universidade Federal de Minas Gerais, Brazil*
Marcus Vinicius SOARES, *Statistics Department, Universidade Federal de Minas Gerais, Brazil*
Eliane Dias GONTIJO, *Social and Preventive Medicine Department, Universidade Federal de Minas Gerais, Brazil.*

Many spatial cluster finder algorithms do not have adequate procedures for controlling the shapes of the clusters found. The cluster solution may sometimes spread through large portions of the map, making it difficult for the practitioner to assess its geographical meaning. Whenever the Kulldorffs spatial scan statistic is used, some kind of correction needs to be used to avoid the excessive irregularity of the clusters. Geometric and topologic corrections have been recently proposed. A weak link is defined as a relatively unpopulated region within a cluster, such that its removal disconnects the cluster. We argue that the existence of weak links is the chief factor impacting the geographic consistency of a cluster, being more important than its shape irregularity. We present a novel scan statistic algorithm employing a cohesion function based on the graph topology to penalize the presence of weak links in candidate clusters. By applying this weak link penalty cohesion function, the most geographically meaningful clusters are sifted through the immense set of possible irregularly shaped candidate clusters solutions. Numerical tests show that the weak link correction has advantages over the previous geometric correction, boosting the power to detect elongated clusters, and presenting good sensitivity and positive predictive value. A multi-objective genetic algorithm is used to compute the solutions, consisting of the Pareto-set of clusters candidates. The goal of the cluster finder algorithm is to maximize two objectives: the scan statistic and the cohesion of the graph structure. The statistical significances of the clusters in the Pareto-set are estimated through Monte Carlo simulations. A real data application for Chagas disease in Brazil is presented.

118 Probabilistic image segmentation without combinatorial optimization

[IS 3,(page 55)]

Mario A. T. FIGUEIREDO, *Instituto de Telecomunicacoes, Instituto Superior Tecnico, Lisboa, Portugal*

The goal of segmentation is to partition an image into a set of regions which are homogeneous in some (e.g., statistical) sense; it is thus an intrinsically discrete problem. Probabilistic approaches to segmentation use priors (such as Markov random fields) to impose spatial coherence. The discrete nature of segmentation demands priors defined on discrete-valued fields, thus leading to difficult combinatorial optimization problems.

In this talk, I will present a formulation which allows using continuous priors, namely Gaussian fields, or wavelet-based priors, for image segmentation. Our approach completely avoids the combinatorial nature of standard approaches to segmentation and is completely general, in that it can be used in supervised, unsupervised, or semi-supervised modes, with any probabilistic observation model.

119 Quicksort: what we know, what we think we know (but don't really), and what we don't know

[IS 13,(page 30)]

James Allen FILL, *Department of Applied Mathematics and Statistics, The Johns Hopkins University*

`Quicksort`, invented by C. R. (Tony) Hoare in 1962, is an elegantly simple, yet quite efficient, recursive randomized algorithm for sorting a list of numbers. It is the standard sorting procedure in `Unix` systems, and has been cited as one of the ten algorithms most influential on the development and practice of science and engineering in the last hundred years.

The runtime of `Quicksort` is measured well by the number of comparisons C_n needed to sort a list of n numbers. It is known that C_n, after normalization, converges in distribution. Focusing on this convergence and the limiting distribution, I will survey what is known (in some cases rigorously and in other cases non-rigorously), and what remains unknown, about `Quicksort`.

120 A class of conditional probability distributions and its application to the statistical regression theory.

[CS 23,(page 22)]

Jerzy FILUS, *Department of Mathematics and Computer Science, Oakton Community College, Des Plaines, IL 60016, USA*

Lidia FILUS, *Department of Mathematics, Northeastern Illinois University, Chicago, IL 60625, USA*

The topic of our presentation is basically associated with some new aspects of the stochastic modeling. Originally, our work was related to description of a relatively new kind of continuous stochastic dependences that seem to cover a significantly wide range of various real life phenomena. The general pattern for that dependence description is obtained by means of some unexpectedly simple method of conditioning. This method relies on a kind of randomization of an originally constant (the no influence case) parameter(s) of any pdf g0(y) of a random variable, say Y, through considering that parameter(s) as any continuous (in general, nonlinear) function of some independent explanatory random variables X1 ,, Xk; each of them having some known probability distribution. This procedure yields to the determination of a wide set of classes of the conditional probability densities

gk (y — x1, ,xk) of random variables say Y, given realizations of the independent (explanatory) random variables X1, ,Xk . The striking simplicity of the obtained conditional densities and a promising wide range of anticipated possible applications, suggest possibility of employing these models in the area of statistical regression. More specifically, we try to extend and develop the idea of the quantile regression introduced by Koenker and Bassett in 1978, and continued by many authors up to the recent. Similarly like ours, this approach is aimed to complement the classical linear regression analysis mainly by replacing the typical investigations centered on the conditional expectation E[Y— x1, ,xk], by a more general considerations on the conditional quantiles directly associated with the (whole) conditional probability distribution. The main advantage of the extension is a possibility as to obtain more statistical information than by using the traditional method of regression. Consequently, one can produce more accurate stochastic predictions.

As we expect, our contribution to the quantile regression theory could be: 1. reinforcement of the existing theory by creation a wide supply of the new effective models (i.e., the conditional distributions), 2. introducing more parametric settings, in place of nonparametric, 3. both the assumptions: one of the normality, and the other of linearity can be relaxed in a natural way.

Moreover, regardless of significant generality of the used methods, as well as of wideness of the resulting class of the obtained models, most of underlying analytical calculations are easy, or relatively easy, to handle.

121 Dependence analysis with reproducing kernel Hilbert spaces

[IS 23,(page 35)]

Kenji FUKUMIZU, *Institute of Statistical Mathematics*

A nonparametric methodology of characterizing independence and conditional independence of random variables is discussed, and its application to dimension reduction for regression is shown. The methodology uses the framework of reproducing kernel Hilbert spaces (RKHS), which are defined by positive definite kernels. In this methodology, a random variable is mapped to a RKHS, thus a random variable on the RKHS is considered. The framework of RKHS enables to derive an easy method of estimating various linear statistics such as mean and covariance with a finite sample, based on the reproducing property that the value of a function at a point is given by the inner product of that function and the positive definite kernel. It is shown that the basic statistics such as mean and covariance of the mapped random variables on the RKHS captures all the information on the underlying probability and dependence of random variables under some assumptions on the positive definite kernels. In particular, the dependence of two variables is sufficiently expressed by the covariance operator on the RKHS. Using the covariance operators, a characterization of conditional independence among variables is also derived. Practical methods are proposed based on the above theoretical results. First, a new measure of independence and conditional independence is given with a method of estimation with a finite sample. A remarkable point of the measure is that the population value does not depend on the choice of a kernel for a wide class of kernels. Some asymptotic results for a large sample limit are also shown. Second, using the characterization of conditional independence, a method of dimension reduction or feature extraction of the covariates is introduced for regression problems. A practical algorithm to extract an effective linear feature is derived. The method is of wide applicability; it does not require any strong assumptions on the type of variables or the probability of variables, which are often imposed by other methods of dimension reduction. Consistency of the estimator is proved under weak condition, and some experimental results show the method is practically competitive.

122 A stochastic heat equation with the distributions of Lévy processes as its invariant measures

[IS 28,(page 36)]

Tadahisa FUNAKI, *Graduate School of Mathematical Sciences, University of Tokyo, Komaba, Tokyo 153-8914, Japan*

It is well-known that the stochastic partial differential equation obtained by adding a space-time Gaussian white noise to the heat equation on a half line has the Wiener measure as its invariant measure. It is therefore natural to ask whether one can find a noise added to the heat equation, under which the distributions of Lévy processes on a path space are invariant. We will give a construction of such noise. Our assumption on the corresponding Lévy measure is, in general, mild except that we need its integrability to show that the distributions of Lévy processes are the only invariant measures of the stochastic heat equation. This is a joint work with Bin Xie.

123 A Bayesian view of climate change: assessing uncertainties of general circulation model projections

[IS 26,(page 35)]

Reinhard FURRER, *Colorado School of Mines, Golden, CO, USA*

Reto KNUTTI, *Swiss Federal Institute of Technology Zurich, Zurich, Switzerland*

Stephan SAIN, *National Center for Atmospheric Research, Boulder, CO, USA*

Recent work on probabilistic climate change projections has focussed mainly on the evolution of global mean temperature. However, impacts of and adaptations for climate change are determined mostly locally and thus require a quantitative picture of the expected change on regional and seasonal scales.

We present probabilistic projections for fields of future climate change using a multivariate Bayesian analysis. The statistical technique is based on the assumption that spatial patterns of climate change can be separated into a large scale signal related to the true forced climate change and a small scale signal stemming from model bias and internal variability. The different scales are represented via a dimension reduction technique in a hierarchical Bayes model. Posterior probabilities are obtained using a Markov chain Monte Carlo simulation technique. The approach is applied to simulations of climate of the twentieth century and several scenarios for the twenty first century from coupled atmosphere ocean general circulation models used in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In contrast to other techniques, the method presented here takes into account uncertainty due to the use of structurally different climate models. It explicitly models the spatial covariance of the global fields, thus providing PDFs of localized climate change that are nevertheless coherent with the distribution of climate change in neighboring locations.

We present probability density functions for the projected temperature in different regions as well as probabilities of exceeding temperature thresholds aggregated over space. By the end of the century in the SRES A1B scenario, 40% of the land regions are found to very likely (90% probability) warm more than two degrees Celsius relative to the preindustrial era in boreal winter (38% in boreal summer).

124 Markov stochastic operators of heredity

[CS 69,(page 52)]

N.N. GANIKHODJAEV, *Faculty of Science, International Islamic University Malaysia, 25200 Kuantan, Malaysia and Institute of Mathematics,*
Tashkent,700125,Uzbekistan

Consider a biological population and assume that each individual in this population belongs to precisely one species $1, \cdots, m$. The scale of species is such that the species of the parents i and j unambiguously determines the probability $p_{ij,k}$ of every species k for the first generation of direct descendants, where $p_{ij,k} \geq 0$ for all i, j, k and

$$\sum_{k=1}^{m} p_{ij,k} = 1.$$

The state of the population is described by the vector

$$\mathbf{x} \in S^{m-1} = \{\mathbf{x} = (x_1, \cdots, x_m) \in R^m : x_i \geq 0, \sum_{i=1}^{m} x_i = 1\},$$

where x_k is the fraction of the species k in the total population.In the case of random interbreeding the parent pairs i and j arise for a fixed state $\mathbf{x} = (x_1, x_2, \cdots, x_m)$ with probability $x_i x_j$.Hence the total probability of the species k in the first generation of direct descendants is defined by quadratic stochastic operator (q.s.o.)

$$V : (V\mathbf{x})_k = \sum_{i,j=1}^{m} p_{ij,k} x_i x_j, \quad (k = 1, \cdots, m) \qquad (1)$$

Let $\Pi = (q_{ij})_{i,j=1}^{m}$ be a stochastic matrix and $\mathbf{x} = (x_1, x_2, \cdots, x_m)$ be a fixed state of population.If parents pairs i and j arise with probability $x_i q_{ij}$ we call such interbreeding Π-*random interbreeding.* Under Π-panmixia the total probability of the species k in the first generation of direct descendants is defined as

$$V_\Pi : (V_\Pi \mathbf{x})_k = \sum_{i,j=1}^{m} p_{ij,k} q_{ij} x_i \quad (k = 1, \cdots, m). \qquad (2)$$

We call a transformation (2) $V_\Pi : S^{m-1} \rightarrow S^{m-1}$ a *Markov stochastic operator of heredity.*

Under Π-random interbreeding the total probability of the species k in the $(n+1)th$ generation of direct descendants is defined as

$$x_k^{(n+1)} = \sum_{i,j=1}^{m} p_{ij,k}^{(n,n+1)} q_{ij}^{(n,n+1)} x_i^{(n)} \quad (k = 1, \cdots, m),$$

where $n = 0, 1, 2, \cdots$, and $\mathbf{x}^{(0)} = \mathbf{x}$.

If $\Pi^{(n,n+1)} = (q_{ij}^{(n,n+1)})_{i,j=1}^{m}$ is defined as $q_{ij}^{(n,n+1)} = x_j^{(n)}$ then Markov stochastic operator of heredity(2) is reduced to q.s.o. (1).

A transformation $V : S^{m-1} \rightarrow S^{m-1}$ is called *ergodic* if for each $\mathbf{x} \in S^{m-1}$ the limit

$$\lim_{k\to\infty} \frac{1}{k} \sum_{n=0}^{k-1} V^n(\mathbf{x})$$

exists.

In [2] Ulam conjectured that any q.s.o. is ergodic transformation. In 1977 Zakharevich [3] showed that this conjecture is false in general. Later in [1] necessary and sufficient conditions were established for the ergodicity of a q.s.o. on S^2

$$V : (x, y, z) \rightarrow (x(1 + ax - bz), y(1 - ax + cz), z(1 + bx - cy)). \quad (3)$$

Theorem The quadratic stochastic operator (3) is non ergodic if and only if the parameters a, b, c have the same sign and each is non-zero.

We will consider asymptotic behavior of trajectories and the ergodic hypothesis for Markov stochastic operators of heredity.

References

1. Ganikhodjaev N.N., Zanin D.V., *Russian Math.Surveys* **59**:3,571-572,(2004)
2. Ulam S.,*A collection of mathematical problems*, Interscience Publishers,New-York-London 1960.
3. Zakharevich M.I.,*Russian Math.Surveys* **33**:6,265-266 (1978)

125 Instrumental variable estimation of regression coefficients in a restricted ultrastructural measurement error model

[CS 7,(page 9)]

Gaurav GARG, *Indian Institute of Technology Kanpur, Kanpur - 208016, Uttar Pradesh, INDIA*

We consider a multiple linear regression model where all the variables are observed with additive measurement errors. Such models are usually called measurement error models. When there are no measurement errors in the data, ordinary least squares estimator is the best linear unbiased estimator of regression coefficients. However, in the presence of measurement errors, it becomes biased and inconsistent. It is well known that in measurement error models, some additional information is required to estimate regression coefficients consistently. Two popular types of such additional information in multivariate models are covariance matrix of measurement error vectors and reliability matrix. Availability of such information is a big constraint in obtaining the consistent estimators. Frequently, this information is either not available or is not widely shared. When such additional information is not available, the instrumental variable approach can be utilized. In many situations, some prior information is available about the unknown regression coefficients. We assume that this prior information is expressible in the form of exact linear restrictions on regression coefficients. Using the instrumental variable approach, we obtain some consistent estimators of regression coefficients which are consistent as well as satisfy the given restrictions. We derive asymptotic properties of proposed estimators without assuming any distributional form of measurement errors or any other random term in the model. The performances of the estimators and the effect of departure from normality of distributions of measurement errors are studied in finite samples through a simulation study.

126 Weak convergence of error processes in discretizations of stochastic integrals and Besov spaces

[CS 10,(page 12)]

Stefan GEISS, *Department of Mathematics and Statistics, P.O. Box 35 (MaD), FIN-40014 University of Jyvaeskylae, Finland*

Anni TOIVOLA, *Department of Mathematics and Statistics, P.O. Box 35 (MaD), FIN-40014 University of Jyvaeskylae, Finland*

We consider weak convergence of the rescaled error processes arising from Riemann discretizations of certain stochastic integrals and relate the L_p-integrability of the weak limit to the fractional smoothness in the Malliavin sense of the stochastic integral. More precisely, consider a stochastic integral

$$g(X_1) = \mathrm{E}g(X_1) + \int_0^1 \frac{\partial G}{\partial x}(u, X_u)dX_u,$$

with an appropriate diffusion $X = (X_t)_{t\in[0,1]}$, we look at Riemann approximations along deterministic time-nets $\tau = (t_i)_{i=1}^n$ and the corresponding error process $C(\tau) = (C_t(\tau))_{t\in[0,1]}$ given by

$$\begin{aligned} C_t(\tau) := & \int_0^t \frac{\partial G}{\partial x}\left(u, X_u\right) dX_u \\ & - \sum_{i=0}^{n-1} \frac{\partial G}{\partial x}\left(t_i, X_{t_i}\right)\left(X_{t_{i+1}\wedge t} - X_{t_i \wedge t}\right). \end{aligned}$$

In Finance, the process $C(\tau)$ describes the hedging error which occurs when a continuously adjusted portfolio is replaced by a portfolio which is adjusted only at the time-knots $t_0, ..., t_{n-1}$. Given a sequence of time-nets τ^n, we are interested in the weak convergence of $\sqrt{n}C(\tau^n)$ as $n \to \infty$.

- We characterize the existence of a square integrable weak limit of $\sqrt{n}C(\tau^{n,\beta})$, where $\tau^{n,\beta} = (t_i)_{i=0}^n$ are the special refining time-nets with $t_i = 1 - (1 - i/n)^{1/\beta}$, by the condition that g or $g(\exp(\cdot - \frac{1}{2}))$ (depending on the diffusion X) belongs to the Besov space $B_{2,2}^{\beta}(\gamma)$. The parameter $\beta \in (0, 1]$ is the fractional smoothness in the Malliavin sense.

- We give nearly optimal conditions that the weak limit is L_p-integrable. As an application for the Binary option we compute the best possible L_p-integrability of the weak limit provided that the $\tau^{n,\beta}$-nets are used and show how the integrability can be improved to any $p \in [2, \infty)$ by using nets with a sufficiently high concentration of the time-knots close to the final time-point $t = 1$.

- The upper estimate for the L_p-integrability of the weak limit for the Binary option has a more general background: assuming that g has a local singularity of order $\eta \geq 0$ (measured in terms of a sharp function), we deduce an upper bound for the L_p-integrability of the weak limit. This is related to a result of Watanabe about the regularity of conditional expectations and the Donsker delta function.

References

[1] S. Geiss and A. Toivola. Weak convergence of error processes in discretizations of stochastic integrals and Besov spaces. *arXiv 0711.1439*.

127 Asymptotic tail probabilities of sums of dependent subexponential random variables

[CS 10,(page 12)]

Jaap GELUK, *The Petroleum Institute P.O. Box 2533 Abu Dhabi United Arab Emirates*

Qihe TANG, *Department of Statistics and Actuarial Science The University of Iowa, 241 Schaeffer Hall, Iowa City, IA 52242, USA*

In this paper we study the asymptotic behavior of the tail probabilities of sums of dependent and real-valued random variables whose distributions are assumed to be subexponential and not necessarily of dominated variation. We propose two general dependence assumptions under which the asymptotic behavior of the tail probabilities of the sums is the same as that in the independent case. In particular, the two dependence assumptions are satisfied by multivariate Farlie-Gumbel-Morgenstern distributions. Another application of our result is to determine the asymptotic behavior of $\Pr(S_n > x)$ $(x \to \infty)$ for $S_n = e^{Y_1} + \ldots + e^{Y_n}$ where the vector $(Y_1, \ldots, Y_n)$ has a multivariate normal distribution, extending a result of Asmussen and Rojas-Nandayapa (2007).

128 Shares of various weighted/ unweighted association and partial association components: asymptotic maximum likelihood and Bayesian analysis

[PS 1,(page 3)]

S.K. GHOREISHI, *Department of Statistics, Faculty of sciences, Razi university, Kermanshah, I.R. of Iran*

Abstract: Goodman RC(M) association models have been considered by some statistician in a number of situations. In this paper, we survey an interesting applicable association model RC(M1,M2) which includes Goodman RC(M) model as a special case. We confine our discussion as follow: 1) decomposing the association term into its polynomial orthogonal components, 2) analyzing a weighted RC(M1,M2) association model, 3) defining and determining the shares of various weighted trends, 4) finding the asymptotic distributions of these shares, 5) analyzing the partial association when there are more than two polytomous variables. Two examples illustrate the procedures and specify various trends in association.

Keywords: Association models, Contrast coefficients, Multiple association, Partial association, Polynomial trends, Weights.

129 Kullback–Leibler property of kernel mixture priors in Bayesian density estimation

[CS 19,(page 20)]

Subhashis GHOSAL, *North Carolina State University*

Yuefeng WU, *North Carolina State University*

Positivity of the prior probability of Kullback–Leibler neighborhood around the true density, commonly known as the Kullback–Leibler property, plays a fundamental role in posterior consistency in Bayesian density estimation and semiparametric problems. Therefore it is important to identify sufficients conditions on the true density and the prior distribution which lead to the Kullback–Leibler property. The Dirichlet mixture of a parametric kernel function is a popular choice of a prior, where the kernels are chosen depending on the sample space and the class of densities to be estimated. The Kullback–Leibler property of the Dirichlet mixture prior has been shown for only some very special kernels such as the normal kernel or the Bernstein polynomial. In this talk, we shall obtain easily verifiable sufficient conditions on the true density and the kernel function so that the Kullback–Leibler property holds. We illustrate our results with a wide variety of kernels used in practice, including the normal, t, histogram, Weibull and gamma densities. This gives a catalog of simple sufficient conditions leading to the Kullback–Leibler property, and can be readily used in applications.

130 Statistical methods for integrative genomic analyses

[IS 24,(page 9)]

Debashis GHOSH, *Departments of Statistics and Public Health Sciences, Penn State University*

Hyungwon CHOI, *Departments of Biostatistics and Pathology, University of Michigan*

Steve QIN, *Departments of Biostatistics and Center for Statistical Genetics, University of Michigan*

With the proliferation of high-throughput technologies for generating genomewide data, a key challenge will be to develop methodology for combining information from the various data sources. In this talk, we will focus on one example of such a data integration problem: that of combining genomewide mRNA and copy number expression datasets. We will describe several approaches to the problem, some model-based and some nonparametric. Interesting statistical issues arise, among which are the following: (1) Multiple testing procedures with discrete test statistics; (2) Multivariate extensions of familiar multiple testing procedures; (3) Use of latent variables for data integration. These issues will be described and some proposals for addressing the issues listed above will be given.

131 An application of maximum entropy spectral analysis

[CS 16,(page 16)]

Cigdem GIRIFTINOGLU, *Anadolu University, Science Faculty, Department of Statistics, Eskisehir, Turkey*
Aladdin SHAMILOV, *Anadolu University, Science Faculty, Department of Statistics, Eskisehir, Turkey*

Maximum entropy formalism enables us to obtain a distribution for a time series when the knowledge is given as the expected values of the autocovariance up to a lag m. Although the autocovariance functions and spectral density function form a Fourier transform pair, Power Spectrum cannot be exactly obtained since only a limited number of autocovariances are available. Therefore, the principle of maximum entropy can be used to extract the best estimate of power spectrum. This method which is called Maximum entropy Spectral Analysis (MESA) has the fewest possible assumptions about the unknown autocovariance functions among the traditional spectrum estimation methods. In this study, in order to obtain Maximum Entropy distribution and spectral density function, a process determined by values of lagrange multipliers is presented for a real time series that is stationary up its second-order statistics

132 Behaviour near the extinction time in stable self-similar fragmentations

[CS 73,(page 54)]

Christina GOLDSCHMIDT, *Department of Statistics, University of Oxford*
Benedicte HAAS, *CEREMADE, Université Paris-Dauphine*

A *fragmentation process* describes the way an object (here thought of as the interval $(0,1)$) falls apart. We work with a particular example called the *stable fragmentation process*, which was introduced by Miermont. Take a normalized excursion $(e(x), 0 \leq x \leq 1)$ of the height process which encodes the β-stable Lévy tree for $\beta \in (1,2)$. The state of the fragmentation process at time $t \geq 0$ is obtained by considering the sequence of lengths of the connected components of the open set $\{x : e(x) > t\}$. Call this sequence $F(t) = (F_1(t), F_2(t), \ldots)$, where we list the elements in decreasing order of size; the elements of this sequence are referred to as *blocks*. Note that $F(0) = (1, 0, 0, \ldots)$ and $\sum_{i=1}^{\infty} F_i(t) \leq 1$ for all $t \geq 0$. In the terminology of Bertoin, $(F(t), t \geq 0)$ is a *self-similar fragmentation* of index $\alpha = 1/\beta - 1$. The self-similar fragmentations are a class of Markovian fragmentation processes in which, roughly speaking, different blocks split independently and in the same way but at a rate which is "proportional" to their lengths to the power of the index. Since in our case the index is negative, we know that smaller blocks split faster than larger ones; indeed, the whole state is reduced to *dust* (i.e. blocks of length 0) in an almost surely finite time ζ. We give a full description of the asymptotics of the fragmentation process as it approaches this time. In particular, we show that there exists a non-trivial limit L taking values in the set of non-increasing, non-negative sequences with finite sum, which is such that $t^{1/\alpha} F((\zeta - t)^+) \to L$ in distribution as $t \downarrow 0$.

133 Reified Bayesian analysis for physical systems

[IS 34,(page 10)]

Michael GOLDSTEIN, *Department of Mathematical Sciences, Durham University, England*

Many physical problems are studied through the construction and analysis of mathematical models. While such analysis may be very revealing, it is also inherently limited, as even a good model will only offer an imperfect representation of the underlying physical system. The relationship between a computer implementation of a mathematical model and the actual behaviour of the physical system is only poorly understood in principle and is often ignored in practice. This discrepancy between model and system raises fundamental questions as to how we should learn about actual physical systems through the analysis of models of such systems. These questions go to the heart of the philosophy and practice of science.

We shall describe an approach to analysing such discrepancy, termed reified analysis. To reify is to consider an abstract concept to be real. In our context, therefore, reified analysis is the statistical framework which allows us to move our inferences from the abstract notion of the mathematical model to the real notion of the physical system, via one or more simulators. We will describe the general approach and illustrate the various stages involved in carrying out such a reified analysis.

134 Semi-parametric PORT-ML and PORT-MP tail index estimators: asymptotic and finite sample comparison

[CS 61,(page 47)]

M. Ivette GOMES, *University of Lisbon, Faculty of Science, DEIO and CEAUL, Portugal*
Lígia Henriques RODRIGUES, *Tomar Polytechnic, College of Technology, CEAUL, Portugal*

In Statistics of Extremes, inference is often based on the excesses over a high random threshold. Those excesses are approximately distributed as the whole set of order statistics associated with a sample from a Generalized Pareto (GP) model. We then get the so-called "maximum likelihood" (ML) extreme value index estimators (Smith, 1987; Drees et al., 2004), denoted PORT-ML. The terminology PORT, named after Araújo Santos et al. (2006), stands for Peaks Over Random Threshold. Dealing with heavy tails only, we are particularly interested in a similar ML estimator, denoted PORT-MP tail index estimator, with MP standing for modified-Pareto (Gomes and Henriques Rodrigues, 2008). Such an estimator is based also on the excesses over a high random threshold, but with a trial of accommodation of bias on the GP model underlying those excesses. At optimal levels, we proceed to an asymptotic comparison of the two estimators jointly with the recently introduced mixed moment estimator (Fraga Alves et al., 2006), as well as the classical Hill (Hill, 1975), moment (Dekkers et al., 1989) and iterated Hill (Beirlant et al., 1996) estimators. An illustration of the finite sample behaviour of the estimators is provided through a Monte Carlo simulation study and attention is drawn to the urgent need of a new testing procedure.

References

1. Araújo Santos, P., Fraga Alves, M.I., and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation. Revstat 4:3, 227-247.
2. Beirlant, J., Vynckier, P., and Teugels, J. (1996). Excess functions and estimation of the extreme-value index. Bernoulli 2, 293-318.
3. Dekkers, A., Einmahl, J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. Annals of Statistics 17, 1833-1855.
4. Drees, H., Ferreira, A., and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. Ann. Appl. Probab. 14, 1179-1201.
5. Fraga Alves, M.I., Gomes, M.I., de Haan, L. and Neves, C. (2007). Mixed Moment Estimator and its Invariant Alternatives. Notas e Comunicações CEAUL 14/07. Submitted.
6. Gomes, M.I. and Henriques Rodrigues, L. (2008). Tail index estimation for heavy tails: accommodation of bias in the excesses over a high threshold. Extremes. DOI: 10.1007/s10687-008-0059-1, 2008.
7. Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163-1174.
8. Smith, R.L. (1987). Estimating tails of probability distributions. Ann. Statist. 15:3, 1174-1207.

135 Bayesian approaches to inverse problems

[IS 22,(page 23)]

Peter J. GREEN, *Department of Mathematics, University of Bristol*
Natalia BOCHKINA, *University of Edinburgh, UK*

Formulating a statistical inverse problem as one of inference in a Bayesian model has great appeal, notably for what this brings in terms of coherence, the interpretability of regularisation penalties, the integration of all uncertainties, and the principled way in which the set-up can be elaborated to encompass broader features of the context, such as measurement error, indirect observation, etc. The Bayesian formulation comes close to the way that most scientists intuitively regard the inferential task, and in principle allows the free use of subject knowledge in probabilistic model building. We will discuss the principles of this Bayesian approach, and illustrate its power with a review of challenging examples in a range of application domains. Taking emission tomography as a canonical example for study, we will present some asymptotic results about Bayesian reconstruction (estimation), and discuss further opportunities for theoretical analysis.

136 Stochastic ion channel dynamics and neural firing patterns

[CS 47,(page 42)]

Priscilla E. GREENWOOD, *Arizona State University*
Luis F. GORDILLO, *University of Puerto Rico at Mayaguez*
Peter ROWAT, *University of California, San Diego*

Deterministic neural models become stochastic when ion channel dynamics are modeled as jump processes with voltage-dependent intensity. We investigate the resulting distributions of neural firing patterns.

137 Free Brownian motion and applications

[Lévy Lecture Lecture,(page 46)]

Alice GUIONNET, *CNRS, École Normale Supérieure de Lyon*

Free Brownian motion is an operator-valued process obtained as the renormalized limit of an NxN Hermitian matrix with independent Brownian entries. We will review its properties as a stochastic process and illustrate its use as a probabilistic tool to solve problems from combinatorics (graph enumeration) and operator algebras.

138 Indonesian call option under geometric average

[CS 68,(page 51)]

GUNARDI, *Department of Mathematics, Gadjah Mada University, Yogyakarta, Indonesia*

Indonesia Stock Exchange (www.idx.co.id) has started to trade options at September 9th, 2004. An option can be considered as an American style barrier option with immediate (forced) exercise if the price hits or crosses the barrier before maturity. The payoff of the option is based on a Weighted Moving Average (WMA) of the price of the underlying stock. The barrier is fixed at the strike price plus or minus a 10 percent. The option is automatically exercised when the underlying stock hits or crosses the barrier and the difference between strike and barrier is paid immediately. We will refer to type of this option as Indonesian option.

To calculate the price of this option contract, we have to model the WMA price. This is not easy. In this paper we study the pricing of the Indonesian call option when WMA is replaced by Geometric Average in a Black-Scholes model. We will derive analytic approximations for the option price.

139 Threshold for the coverage of a unit sphere.

[CS 10,(page 12)]

Bhupendra GUPTA, *Indian Institute of Information Technology (Design and Manufacturing)-Jabalpur, India.*

In this article, we consider 'N'spherical caps of area $4\pi p$ were uniformly distributed over the surface of a unit sphere. We are giving the strong threshold function for the size of random caps to cover the surface of a unit sphere. We have shown that for large N, if $\frac{Np}{\log N} > 1/2$ the surface of sphere is completely covered by the N caps almost surely , and if $\frac{Np}{\log N} \leq 1/2$ a partition of the surface of sphere is remains uncovered by the N caps almost surely.

140 Local asymptotic normality in quantum statistics

[IS 25,(page 51)]

Madalin GUTA, *University of Nottingham*
Jonas KAHN, *Univeriste Paris-Sud 11*

Quantum statistics deals with the processing of statistical information carried by quantum systems.One of the central problems in quantum statistics is that of estimating an unknown quantum state by performing general measurements on a (large) number of identically prepared quantum systems.

We will show how the collective statistical model for d-dimensional quantum systems converges in a statistical sense to a model consisting of a classical Gaussian model and a quantum Gaussian state of d(d-1)/2 harmonic oscillators. Both Gaussian models have fixed variance and unknown mean which can be easily estimated by means of standard (heterodyne/homodyne) measurements. This optimal measurement can be pulled back to an optimal strategy for estimating the state of the d-dimensional systems.

141 On a stopped collision local time formula

[CS 44,(page 39)]

Olympia HADJILIADIS, *Brooklyn College, City University of New York*

In this work we derive a closed form formula for the expected value of a stopped collision local time of two processes, with constant correlation, stopped at the first time that their maximum reaches a prespecified threshold. These processes are each defined to be the differences between Brownian motions with equal drifts and their running minimums. We discuss applications of this work in the problem of quickest detection of abrupt changes in random processes.

142 Using fourier descriptors to distinguish cell shapes

[CS 75,(page 56)]

Charles HAGWOOD, *National Institute of Standards and Technology*

Biological activity within a cell is important for numerous reasons, the medical reason being one of many. One response to activity that can be visualized is a cell's morphology. Using fluorescence microscopy, individual cell shapes can be imaged. A population of identical cells exhibits a distribution of responses. Our goal is to use Fourier desciptors to assess differences between populations of cell shapes

143 Mathematical models of moving pariticles and applications for queues.

[CS 13,(page 13)]

A.H. HAJIYEV, *Azerbaijan National Academy of Sciences*

Mathematical models of moving (in one direction) particles without an outrun on N equidistant points of a circle are considered. There are two particles: a leader, which can jump forward one unit with probability r or stay at the same position with probability 1-r. Motion of the other particle depends on the motion of the leader. If the distance between particles equals one unit and the leader stays at the same position, then other particles with probability 1 dont change their position. If a distance between particles is more than one unit, then the other particle jumps with probability rk, where k is the distance between particles. It is proven that for a stationary regime the leader particle, forces the other particle to make a random binomial walk with the same parameters r and l iff rN-1 =1. This means that if the leader particle is not visible and other particle is, then the visible particle makes a random binomial walk with the same probability. This result is also true for the case of three particles, but is not true for the case of four particles or more. An example four case of four particles is given. A class of probability distributions between moving particles is found for the motion of two particles. It is proven, that any discrete probability distribution can be approximated (convergence in variation) by distribution from this class. A point is fixed on a circle and we assume that Poisson flow of customers arrives for service. Each customer is waiting from the epoch of arriving into the system until the epoch when some particle crosses this point. Such model is typical for applications, particularly, in traffic systems. In the capacity of efficiency index an expectation of waiting time of customer before service is taken. For diminishing of an expectation of waiting time before service of customer the delays of beginning service are introduced. The class of systems, for which introduction of delays diminishes an expectation of waiting time before service is found. The form of optimal function, minimizing an expectation of waiting time before service is also derived. It is shown that for input Poisson flow of customers and exponential distribution intervals between services a gain in expectation of waiting time before service is 10%. The complicated systems, where delays change the next intervals between services are also investigated. Such systems have quite complicated structure and using the previous method of investigation is not possible. A class of complicated systems for which the introduction of delays is advisable is derived. It is shown that for such systems, in some cases, the gain in expectation of waiting time before service is 50%. Numerical examples demonstrating these results for complicated systems are given.

144 Bayesian probit model for credit scoring and reject inference

[CS 59,(page 47)]

Siana HALIM, *Industrial Engineering Department - Petra Christian University. Surabaya - Indonesia*

Indriati N. BISONO, *Industrial Engineering Department - Petra Christian University. Surabaya - Indonesia*

It is generally easier to predict defaults accurately if a large data set (including defaults) is available for estimating the prediction model. However, small banks tend to have small data sets. Large banks can also pose the same problem when they began to collect their own historical data or when they introduced a new rating system. Facing that situation, we proposed to use the Bayesian probit model that enables banks with small data sets to improve their default probability.

The data used in credit scoring modeling usually came from training sets consisting of applicants whom loans were approved. Learning from only approved loans yields an incorrect model because the training set is a biased sample. Defaulted data from the rejected samples regarded as missing values. Therefore, we extend this model to predict these missing values and used the accepted and rejected data to build a new unbiased model. Including the rejected samples in the learning process is called reject inference.

145 Semiparametric estimation of periodic functions

[IS 19,(page 35)]

Peter HALL, *The University of Melbourne*

Many stars in the heavens emit radiation in a periodic fashion. In such cases, physical information about the star can be accessed through the both the period and the 'light curve' – the function that describes the variation of brightness over time. Estimation of the light curve is generally viewed as a nonparametric problem, but estimation of period is a semiparametric problem. However, the rate of convergence is faster than the inverse of the square root of sample size; it is in fact the cube of this rate.

146 Managing loan customers using statistical methods and neural networks

[CS 40,(page 37)]

Maryam HAMOOLEH KHEIROLAHPOUR, *Department of Statistics, School of Math Science, Shahid Chamran University, Ahvaz*
Hamid NILSAZ DEZFOULI, *Faculty Member of Islamic Azad University, Mahshahr , Iran*

Credit scoring is one of the most successful applications of statistical and operational research modeling in financing and banking. Credit scoring techniques asses the risk in lending to a particular consumer and classify consumer loan applications. The objective of quantitative credit scoring models is to assign credit applicants to one of two groups, a good credit that is likely to repay the financial obligation or a bad credit group that should be denied credit because of a high likelihood of defaulting on the financial obligation. Credit scoring has some obvious benefits that have led to its increasing use in loan evaluation, for example it is quicker, cheaper and more objective than judgmental method. A wide range of statistical method such as discriminant analysis, regression, smoothing nonparametric methods and neural networks have been applied in credit scoring. In this paper we design a neural network credit scoring system for classifying the applicants of personal loans in bank and compare the performance of this model with discriminant analysis and logistic regression models.

147 On one inverse problem in financial mathematics

[CS 1,(page 6)]

Kais HAMZA, *Monash University*
Fima KLEBANER, *Monash University*

The Black-Scholes formula has been derived under the assumption of constant volatility in stocks. In spite of evidence that this parameter is not constant, this formula is widely used by the markets. This talk addresses the question of whether a model for stock price exists such that the Black-Scholes and Bachelier formulae hold while the volatility is non-constant.

148 Markov bases for two-way subtable sum problems

[CS 15,(page 15)]

Hisayuki HARA, *University of Tokyo*
Akimichi TAKEMURA, *Graduate School of Information Science and Technology, University of Tokyo*
Ruriko YOSHIDA, *University of Kentucky*

Since Sturmfels (1996) and Diaconis and Sturmfels (1997) showed that a set of binomial generators of a toric ideal for a statistical model of discrete exponential families is equivalent to a Markov basis and initiated Markov chain Monte Carlo approach based on a Gröbner basis computation for testing statistical fitting of the given model, many researchers have extensively studied the structure of Markov bases for models in computational algebraic statistics, In this talk, we discuss Markov bases for two-way contingency tables with fixed row sums, column sums and an additional constraint that the sum of a subtable is also fixed.

Let $X = \{x_{ij}\}$, $i = 1, \dots, R$, $j = 1, \dots, C$ be an $R \times C$ two-way contingency table and S be any nonempty proper subset of the table, i.e. $S \subset \{(i,j) \mid 1 \leq i \leq R, 1 \leq j \leq C\}$. Consider the following model for cell probabilities $\{p_{ij}\}$,

$$\log p_{ij} = \mu + \alpha_i + \beta_j + \gamma I_S,$$

where I_S is the indicator function for S. We call this model a *subtable sum model*. When S satisfies $S = \{(i,j) \mid 1 \leq i \leq R_0, 1 \leq j \leq C_0\}$ for some $R_0 < R$ and $C_0 < C$, the model is called a block interaction model or two-way change point model proposed by Hirotsu (1997). In the case where $S = \{(i,i) \mid 1 \leq i \leq min(R,C)\}$, the model is a special case of quasi-independence model (e.g. Tomizawa(2006)),

$$\log p_{ij} = \mu + \alpha_i + \beta_j + \gamma_i \delta_{ij},$$

where δ_{ij} is Kronecker's delta. We call this model a *common diagonal effect model*.

The sufficient statistics for a subtable sum model are the row sums, the column sums

$$x_{i+} = \sum_{j=1}^{C} x_{ij}, \quad i = 1, \ldots, R,$$

$$x_{+j} = \sum_{i=1}^{R} x_{ij}, \quad j = 1, \ldots, C.$$

and the sum of subtable $x(S)$

$$x(S) = \sum_{(i,j)\in S} x_{ij}.$$

Here we consider Markov bases of the model for performing the conditional test of the goodness-of-fit of the model. It has been well-known that if $S = \emptyset$, the set of square-free moves of degree two forms a Markov basis. However in the case where $S \neq \emptyset$, these moves do not necessarily form a Markov basis. Thus, we first derive a necessary and sufficient condition on S that the set of square-free moves of degree two forms a Markov basis. The result shows that the class include a block interaction model and does not include a common diagonal effect model. So we next derive an explicit form of a Markov basis of a common diagonal effect model.

Once a Markov basis is given, MCMC procedure for the exact test can be easily implemented by techniques of algebraic statistics. We illustrate the procedure with some real data sets.

149 Two point functions and the lace expansion for random walks and percolation in high dimensions

[IS 33,(page 56)]

Takashi HARA, *Faculty of Mathematics, Kyushu University*

Lace expansion is a powerful method for analyzing critical behaviour of various statistical mechanical systems, such as self-avoiding walk, percolation, lattice trees and animals, contact process, and Ising model. In this talk, recent development on the behaviour of critical two-point functions of these systems [1] are explained. In addition, taking into account broad audience, some basics of the lace expansion method is presented. (See [4] for a good review on the lace expansion.)

Main result: We consider nearest-neighbour self-avoiding walk, bond percolation, lattice trees, and bond lattice animals on Z^d. Using the lace expansion, we prove that the two-point function $G(x)$ at the critical point exhibits the asymptotic behaviour, $G(x) \sim const.|x|^{2-d}$ as $|x| \to \infty$, for $d \geq 5$ for self-avoiding walk, for $d \geq 19$ for percolation, and for sufficiently large d for lattice trees and animals.

Comments: (1) It is expected that the main result holds in the following dimensions: $d > 4$ for self-avoiding walk, $d > 6$ for percolation, and $d > 8$ for lattice trees and animals. (2) These results are complementary to those obtained in [2], where spread-out models were considered, and where the results were proved under the above optimal condition on dimensions. (3) Our method can be applied to other models, as long as one has a "reasonable" lace expansion. One good example is the Ising model, for which Alira Sakai [3] constructed the lace expansion and proved similar results.

About the proof: By the lace expansion, one can express the Fourier transform of the critical two-point function (roughly) as $\hat{G}(k) = \{\hat{\Pi}(0) - \hat{\Pi}(k)\}^{-1}$, where exact expression for $\Pi(x)$ is given by the lace expansion. This allows us to interpret the critical two point function as the two-point function of a random walk, whose transition probability is given by $\Pi(x)$.

Now, for a genral random walk on Z^d , we prove a sufficient condition on its transition probability under which the two-point function of the walk is asymptotic to $const.|x|^{2-d}$ as $|x| \to \infty$.

The main result is now proved, by showing that $\Pi(x)$ satisfies this condition in high dimensions.

References

1. T. Hara: *Ann. Prob.* **36**,(2008) 530–593.
2. T. Hara, R. van der Hofstad, and G. Slade: *Ann. Prob.* **31**,(2003) 349–408.
3. A. Sakai: *Commun. Math. Phys.* **272**,(2007) 283–344.
4. G. Slade: *The Lace Expansion and its Applications.* Lecture Notes in Mathematics 1879. Springer,(2006).

150 Sampling the palaeoclimate of the past 15000 years using Bayesian models

[CS 48,(page 42)]

John HASLETT, *Trinity College Dublin Ireland*
Andrew PARNELL, *Trinity College Dublin Ireland*
Michael SALTER-TOWNSHEND, *Trinity College Dublin Ireland*

The central difficulty in understanding future climate change - and most especially possible abrupt climate change - is that most of what we know about climate derives from data collected over only the past century or so. Much of the important data - for example, variation in ice coverage -covers no more than a few decades, since the advent of satellites. There is almost no ocean temperature data older than 100 years; similarly there is very little Southern Hemisphere of any sort older than 100 years. Access to pre-historic records is indirect and via proxies in ice-cores, or in sediment in lakes and the ocean.

This paper will report on Bayesian modelling of the European palaeoclimate for the past 15000 years using pollen data in lake sediment; specifically it will report on progress since the proof-of-concept paper read to the Royal Statistical Society in 2006. Technical interest lies in the use of high-dimensional priors based, for different aspects of the problem, on monotone, long-tailed and Gaussian stochastic processes.

The essential statistical problem may be stated thus. Modern training data (y^m, c^m) are available, as are ancient data y^a. The task is to study $[c^a|y^a, (y^m, c^m)]$, which we do by high dimensional Monte Carlo sampling. The y represent multivariate (here $dim(y) = 28$) counts of different types of pollen - here $dim(y) = 28$; c represents multivariate climate - here $dim(c) = 2$. The key scientific idea is that pollen tells us about vegetation and vegetation tells us about climate. One technical statistical task is the Bayesian inversion of a multivariate non-parametric regression, using Gaussian priors. A second is that there are many y^a in a single core representing many values in a changing climate, depth representing age. The prior models the entire climate history, jointly. A long tailed random walk prior is used. Finally, the depth age relationship is incompletely known, but is monotone; a monotone Compound Poisson Gamma process is used (Haslett and Parnell, 2008). Results will be presented that confirm an abrupt change about 10,000 years ago. Some aspects of the technical work will be discussed, given time. This will focus on the Compound Poisson Gamma process.

References

1. Haslett, J. , Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, Simon P., Allen, J. R. M.,Huntley, B. ,Mitchell, F. J. G. Bayesian palaeoclimate reconstruction, Journal of the Royal Statistical Society, Series A-Statistics in Society 169: 395-430, 2006
2. Haslett, J. , and Parnell A. A simple monotone process with application to radio-carbon dated depth chronologies, Journal of the Royal Statistical Society: Series C Forthcoming

151 Hydrodynamic limit of the stochastic ranking process and its application to long-tail economy

[CS 33,(page 32)]

Kumiko HATTORI, *Tokyo Metropolitan University*
Tetsuya HATTORI, *Tohoku University*

Internet commerce has drastically increased product variety through low search and transaction costs and nearly unlimited inventory capacity. This fact suggests the so-called Long Tail theory which claims a huge number of poorly selling goods that are now available on internet catalogues (long-tail goods) make a significant contribution to total sales. We study long-tail economy on mathematical basis focusing on online bookstores, specifically, Amazon.co.jp.

Online retailers seem to be hesitant about releasing sales data. One of the scant data publicly available is sales ranks of books, which we use for our estimate of sales.

We propose the following simple model of ranking. Consider a system of N book titles, each of which has a rank ranging from 1 to N so that no two titles have the same rank. Each title sells at random times. Every time a copy of a title sells, the title jumps to rank 1 immediately. If its rank was m before the sale, all the titles that had ranks 1 to $m - 1$ just before the sale shift to ranks 2 to m, respectively. Different titles can have different sales rates (selling well or poorly), and sell independently. Thus, the motion of a book's rank is caused by its own jumps (sales) and push-back by other titles.

We prove that under appropriate assumptions, in the limit $N \to \infty$, the random motion of each book's rank between sales converges to a deterministic trajectory. This trajectory can actually be observed as the time-development of a book's sales rank at Amazon.co.jp's website. Simple as our model is, its prediction fits well with observation and allows the estimation of, what economists call, the Pareto slope parameter. We also prove that the "hydrodynamic limit" exists, in the sense that the (random) empirical distribution of this system (sales rates and scaled

ranks) converges to a deterministic time dependent distribution.

Our result of the theoretical fit to the observed data suggests that the economic impact of long-tail products is smaller than that claimed in preceding studies.

152 Asymptotic goodness-of-fit tests for point processes based on second-order characteristcs

[CS 17,(page 19)]

Lothar HEINRICH, *University of Augsburg, 86135 Augsburg, Germany*
Stella DAVID, *University of Augsburg, 86135 Augsburg, Germany*

In many fields of application statisticians are faced with irregular point patterns or point-like objects which are randomly distributed in large planar or spatial domains. Random point processes provide appropriate models to describe such phenomena. It is often assumed and in practice at least approximately justifyable that the distribution of point fields under consideration are invariant under translations. The main aim of the talk is to establish goodness-of-fit tests for checking point process hypotheses when the hypothesized homogeneous d-dimensional point process $N(\cdot) = \sum_{i\geq 1} \delta_{X_i}(\cdot)$ possesses an intensity $\lambda := \mathsf{E}N([0,1)^d) > 0$ and a known second-order moment function $K(r) := \lambda^{-1}\,\mathsf{E}\big(N(B_r)\setminus\{o\})|N(\{o\}) = 1\big)$ (known as *Ripley's K-function*, where B_r is a ball with radius $r \geq 0$ centred at o) and the number $N(W_n)$ turns out to be asymptotically Gaussian with mean $\lambda\,|W_n|$ and asymptotic variance $\sigma^2 := \lim_{n\to\infty} n^{-d}\,\mathsf{Var}N([0,n)^d)$ when the observation window W_n runs through a growing sequence of convex sets. An unbiased, edge-corrected (so-called *Horvitz-Thompson-type*) estimator $\hat{K}_n(r)$ for $K(r)$ is calculated from a single observation of the point pattern in W_n for $0 \leq r \leq r(W_n)(\to\infty)$ and the limit distributions of appropriately scaled discrepancy measures between empirical and true K-function are used to check our point process hypothesis. Special emphasis is put on checking the Poisson hypothesis for $N(\cdot)$. To determine the weak limits of the test statistics especially in the Poisson case we need central limit theorems for U-statistics and locally dependent random fields including Berry-Esseen rates. Another crucial issue is the consistent estimation of σ^2. Finally we discuss some asymptotic results for the *integrated mean squared error* of kernel-type estimators for second-order product densities of *Brillinger-mixing* point process and their use for testing point process hypotheses.

153 Testing many families of hypotheses using partial conjunctions and the FDR

[IS 12,(page 19)]

Ruth HELLER, *Wharton School, University of Pennsylvania*

We address the problem of testing many units of interest when more than one hypothesis is tested on each unit. We introduce the overall FDR, i.e. the expected proportion of units for which at least one hypothesis has been falsely rejected, as an appropriate measure for controlling for the multiple units tested. For controlling the overall FDR, we suggest 1) first applying the BH procedure on the p-values from a "screening" hypothesis for each unit. We suggest a "screening" hypothesis that is useful for many modern biostatistics applications. This is the partial conjunction hypothesis, that asks whether at least u out of n null hypotheses are false. 2) On the discovered units, we can further test for partial conjunctions with increasing values of u with no further cost for multiplicity. We apply the method to examples from Microarray analysis and functional Magnetic Resonance Imaging (fMRI), two application areas where the problem has been identified.

154 Increment stationarity for set-indexed processes: from set-indexed fractional Brownian motion to set-indexed Lévy processes

[CS 32,(page 31)]

E. HERBIN, *Ecole Centrale Paris*
E. MERZBACH, *Bar Ilan University*

In [HeMe06], we defined and proved the existence of a fractional Brownian motion (fBm) indexed by a collection $\mathcal{A}$ of closed subsets of a measure space $\mathcal{T}$, in the frame of [IvMe00]. It is defined as a centered Gaussian process $B^H = \left\{B^H_U;\ U \in \mathcal{A}\right\}$ such that

$$\forall U,V \in \mathcal{A};$$
$$E\left[B^H_U B^H_V\right] = \frac{1}{2}\left[m(U)^{2H} + m(V)^{2H} - m(U \bigtriangleup V)^{2H}\right]$$

where $H \in (0, 1/2]$, $\bigtriangleup$ is the symmetric difference between sets, and m is a measure on $\mathcal{T}$. In order to study fractal properties of the set-indexed fractional Browian motion (sifBm), we defined properties of self-similarity and increment stationarity for $\mathcal{A}$-indexed processes. The latter has been strenghtened in [HeMe07] and allows a complete characterization of the sifBm by its fractal properties. Under this new property, if X is a set-indexed process with stationary increments, then its projection on any increasing path is a one-parameter increment stationary process in the usual sense.

The so-called m-stationarity of $\mathcal{C}_0$-increments definition is used to give a new definiton of set-indexed Lévy processes. On the contrary to previous definitions (see [AdFe84],[BaPy84]), the parameter set is not reduced to rectangles of $[0,1]^N$ and no group structure is needed to define the increment stationarity property. We will discuss some simple examples and links with infinitely divisible distributions. Consequentely, we get a Lévy-Khintchine representation formula.

References

[AdFe84] R.J. Adler and P.D. Feigin, On the cadlaguity of random measures, *Ann. Probab.* 12, 615-630, 1984.

[BaPy84] R.F. Bass and R. Pyke, The existence of set-indexed Lévy processes, *Z. Wahr. verw. Gebiete* 66, 157-172, 1984.

[HeMe06] E. Herbin and E. Merzbach, A set-indexed fractional Brownian motion, *J. of Theoret. Probab.*, Vol. 19, No. 2, pp. 337-364, 2006.

[HeMe07] E. Herbin and E. Merzbach, Stationarity and Self-similarity Characterization of the Set-indexed Fractional Brownian Motion, preprint 2007.

[IvMe00] G. Ivanoff and E. Merzbach, *Set-Indexed Martingales*, Chapman & Hall/CRC, 2000.

155 Asymptotic distribution of modified vector variance (MVV)

[CS 34,(page 32)]

Erna Tri HERDIANI, *Institut Teknologi Bandung*
Maman A. DJAUHARI, *Faculty Mathematics and Natural Sciences, ITB, Indonesia*

The most popular methods of testing hypothesis equaltiy of some correlation matrices is Likelihood Ratio Test (LRT). Computation of LRT based on computation determinant of covariance matrix. In the case of variable is large, computation of LRT will be difficulty, as a consequance Vector Variance (VV) can be used as alternative solution. In this paper, we will show asymptotic distribution of VV to test hypothesis equality of some covariance matrices with only involve elements of upper triangle diagonal or lower triangle diagonal of covariance matrices. It so called as Modified Vector Variance (MVV). The result of this paper will be more eficiency because covariance matrix in VV is a symmetry matrix.

156 Hellinger estimation of general bilinear time series models

[CS 30,(page 27)]

O HILI, *National Polytechnic Institute of Yamoussoukro*

In the present paper, minimu Hellinger distance estimates for parametors of a general bilibear time series model are presented.The probabilistic properties such as stationarity, existence of moment of the stationary distribution and trong mixing property are well know. We establish, under some milds conditions, the consistency and the asymptotic normality of the minimum Hellinger distance estimates of the parameters of the model.

157 On fair pricing of emission-related derivatives

[IS 15,(page 6)]

Juri HINZ, *National University of Singapore*

The climate rescue is on the top of the agendas today. To protect the environment, emission trading schemes are considered as one of the most promising tools. In a system of such type, a central authority allocates credits among emission sources and sets a penalty which must be paid per unit of pollutant which is not covered by credits at the end the period. This regulatory framework introduces a market for emission allowances and creates need for risk management by appropriate emission-related financial contracts. In this talk we apply methodologies from stochastic analysis to address logical principles underlying price formation of tradable pollution certificates. Based on tools from optimal control theory, we characterize the equilibrium allowance prices and show the existence of the proposed price dynamics. Further, we illustrate the computational tractability of the resulting models. In the context of the least square Monte Carlo method, we utilize fixed point arguments to derive appropriate numerical schemes, which are illustrated by examples.

158 Selection of the number of factors in Bayesian factor analysis

[CS 55,(page 44)]

Kei HIROSE, *Graduate School of Mathematics, Kyushu University*
Shuichi KAWANO, *Graduate School of Mathematics, Kyushu University*
Sadanori KONISHI, *Faculty of Mathematics, Kyushu University*
Masanori ICHIKAWA, *Tokyo University of Foreign Studies*

Factor analysis is one of the most popular methods of multivariate statistical analysis, used in the social and behavioral sciences to explore the covariance structure among a set of observed random variables by construction of a smaller number of unobserved random variables called common factors. The parameters are usually estimated by maximum likelihood methods under the assumption that the observations are normally distributed. In practical situations, however, the maximum likelihood estimates of unique variances can often turn out to be zero or negative, which makes no sense from a statistical point of view. Such estimates are known as improper solutions, and many researchers have studied these inappropriate estimates both from a theoretical point of view and also by means of numerical examples. In order to overcome this difficulty we use a Bayesian approach by specifying a prior distribution for the variances of unique factors and estimate parameters by posterior modes.

Another important issue in factor analysis model includes the choice of the number of factors. It can be viewed as a model selection and evaluation problem. In maximum likelihood factor analysis, the appropriate number of factors could be selected by the use of AIC or BIC, whereas the occurrence of improper solutions still provokes the problem of the interpretation of the model. The aim of this paper is to introduce procedures for preventing the occurrence of improper solutions and also for selecting the appropriate number of factors.

We show the cause of improper solutions from the feature of the likelihood function and introduce a prior distribution using the knowledge of information extracted from the likelihood function. In order to choose the adjusted parameters that include the hyper-parameters for the prior distribution and the number of factors, we derive a model selection criterion from a Bayesian viewpoint to evaluate models estimated by the maximum penalized likelihood method. A real data example is conducted to investigate the efficiency of the proposed procedures.

159 Accuracy assessment for the trade-off curve and its upper bound curve in the bump hunting using the new tree genetic algorithm

[CS 70,(page 53)]

Hideo HIROSE , *Kyushu Institute of Technology*
Takahiro YUKIZANE , *Kyushu Institute of Technology*
Faisal M. ZAMAN , *Kyushu Institute of Technology*

Suppose that we are interested in classifying n points in a z-dimensional space into two groups having response 1 and response 0 as the target variable. In some real data cases in customer classification, it is difficult to discriminate the favorable customers showing response 1 from others because many response 1 points and 0 points are closely located. In such a case, to find the denser regions to the favorable customers is considered to be an alternative. Such regions are called the bumps, and finding them is called the bump hunting. By pre-specifying a pureness rate p in advance a maximum capture rate c could be obtained; the pureness rate is the ratio of the number of response 1 points to the total number of points in the target region; the capture rate is the ratio of the number of response 1 points to the total number of points in the total regions. Then a trade-off curve between p and c can be constructed. Thus, the bump hunting is the same as the trade-off curve constructing. In order to make future actions easier, we adopt simpler boundary shapes for the bumps such as the union of z-dimensional boxes located parallel to some explanation variable axes; this means that we adopt the binary decision tree. Since the conventional binary decision tree will not provide the maximum capture rates because of its local optimizer property, some probabilistic methods would be required. Here, we use the genetic algorithm (GA) specified to the tree structure to accomplish this; we call this the tree GA. The tree GA has a tendency to provide many local maxima of the capture rates unlike the ordinary GA. According to this property, we can estimate the upper bound curve for the trade-off curve by using the extreme-value statistics. However, these curves could be optimistic if they are constructed using the training data alone. We should be careful in assessing the accuracy of these curves. By applying the test data, the accuracy of the trade-off curve itself can easily be assessed. However, the property of the local maxima would not be preserved. In this paper, we have developed a new tree GA to preserve the property of the local maxima of the capture rates by assessing the test data results in each evolution procedure. Then, the accuracy of the trade-off curve and its upper bound curve are assessed.

160 Monotonicity for self-interacting random walks.

[IS 17,(page 30)]

Mark HOLMES, *University of Auckland*
Remco van der HOFSTAD, *Eindhoven University of Technology*

Monotonicity properties for self-interacting random walks are often rather difficult to prove. For example, an *(once-)excited random walk* has a drift $\beta/2d$ in the first coordinate the first time the walk revisits a site (no drift on subsequent visits). The resulting speed (in dimensions $d > 1$) appearing in the law of large numbers, is obviously monotone increasing in the parameter β, however this has only recently been proved, and only in high dimensions.

We discuss this result and other monotonicity results that it may be possible to obtain using a general expansion for self-interacting random walks.

161 Bayesian multiscale analysis of differences in noisy images

[PS 2,(page 17)]

Lasse HOLMSTROM, *University of Oulu*
Leena PASANEN, *University of Oulu*

We consider the detection of features in digital images that appear in different spatial scales, or image resolutions. In particular, our goal is to capture the scale dependent differences in a pair of noisy images of the same scene taken at two different instances of time. A new approach is proposed that uses Bayesian statistical modeling and simulation based inference. The reconstructed image represented by its posterior distribution is smoothed using using a range of smoothing scales and the smoothed images are analyzed for features that exceed a given level of credibility. The results are summarized in the form of maps that display the features thus found to be statistically significant.

The method can be viewed as a further development of SiZer scale space technology, originally designed for nonparametric curve fitting. A strength of the Bayesian simulation based approach is straightforward inference and modeling flexibility that facilitates the incorporation of domain specific prior information on the images under consideration. The performance of the method is demonstrated using mostly artificial test images. However, our goal is to apply this new approach to satellite based remote sensing and we therefore also include a preliminary analysis of a pair of Landsat images used in satellite based forest inventory.

162 Augmented GARCH sequences: dependence structure and asymptotics

[CS 8,(page 11)]

Siegfried HÖRMANN, *Department of Mathematics, University of Utah, Salt Lake City, USA*

Since the introduction of the ARCH model (autoregressive conditionally heteroscedastic) at the beginning of the 1980s by Robert Engle, many generalizations and modifications have been introduced in the econometrics literature. The main feature of these models is, that their conditional variance is not constant, but changes as a function of past observations – a well known characteristic of financial data. The augmented GARCH model is a unification of numerous extensions of the ARCH process. Besides ordinary (linear) GARCH processes, it contains exponential GARCH, power GARCH, threshold GARCH, asymmetric GARCH, etc. In this talk we will describe the probabilistic structure of augmented GARCH(1,1) sequences and the asymptotic distribution of various functionals of the process occurring in problems of statistical inference. Instead of using the Markov structure of the model and implied mixing properties we utilize independence properties of perturbed GARCH sequences to reduce their asymptotic behavior directly to the case of independent random variables. This method applies for a very large class of functionals and eliminates the fairly restrictive moment and smoothness conditions assumed in the earlier theory. In particular, we derive functional CLTs for powers of the augmented GARCH variables, the error rate in the CLT and obtain asymptotic results of their empirical processes under nearly optimal conditions.

163 Flexible parametric approach for adjusting for the measurement errors in covariates

[CS 7,(page 9)]

Shahadut HOSSAIN, *Post-doctoral Research Fellow, British Columbia Cancer Research Centre, Vancouver, BC, Canada*

Paul GUSTAFSON, *Professor, Department of Statistics, University of British Columbia, Vancouver, Canada*

In most biostatistical and epidemiological investigations the study units are people, the outcome (or the response) variable is a health related event, and the explanatory variables are usually the environmental and/or socio-demographic factors. The fundamental task in such investigations is to quantify the association between the explanatory variables (or the covariates) and the response or outcome variable through a suitable regression model. The accuracy of such quantification depends on how precisely we measure the relevant covariates. In many instances, we can not measure some of the covariates accurately, rather we can measure noisy versions of them. In statistical terminology this is known as measurement errors or errors in variables. Regression analyses based on noisy covariate measurements lead to biased and inaccurate inference about the true underlying response-covariate associations.

In this paper we suggest a flexible parametric approach for adjusting the measurement error bias while estimating the response-covariate relationship through logistic regression model. More specifically, we propose a flexible parametric distribution for modeling the true but unobserved exposure. For inference and computational purpose, we use the Bayesian MCMC techniques. We investigate the performance of the proposed flexible parametric approach in comparison with the other flexible parametric and nonparametric approaches through extensive simulation studies. We also compare the proposed method with a competing flexible parametric method with respect to a real-life data set. Though emphasis is put on the logistic regression model, the proposed method is unified and is applicable to the other members of the generalized linear models, and to the other types of non-linear regression models too.

164 Model-robustly D- and A-optimal designs for mixture experiments

[CS 22,(page 22)]

Hsiang-Ling HSU, *Hsiang-Ling*
Mong-Na Lo HUANG, *National Sun Yat-sen University, Taiwan, R.O.C.*
Chao-Jin CHOU,
Thomas KLEIN,

This paper investigates optimal designs for mixture experiments when there is uncertainty as to whether a polynomial regression model of degree one or two is appropriate. Three groups of novel results are presented: (i) a complete class of designs relative to certain mixed design criteria, (ii) model-robustly D- and A-optimal designs, (iii) D- and A-optimal designs with maximin efficiencies under variation of the design criterion.

165 Bayesian analysis of errors-in-variables growth curves with skewness in models

[CS 59,(page 47)]

Steward HUANG, *Chapman University and University of California at Riverside, USA*

We propose to analyze model data 1) using errors-in-variables (EIV) model and 2) using the assumptions that the error random variables are subject to the influence of skewness through Bayesian approach. The use of EIV in model is necessary and realistic in studying many statistical problems, but their analysis usually mandate many simplifying and restrictive assumptions. Previous studies have shown the superiority of Bayesian approach in dealing with the complexity of these models. In fitting statistical models for the analysis of growth data, many models have been proposed. We selected an extensive list of the most important growth curves and using some of them in our model analysis. Much research using classical approach has clustered on this area. However, the incorporation of EIV into these growth models under Bayesian formulation with skewness models have not yet been considered or studied. A motivating example is presented and in which we expose certain lacunae in the analysis previously done as well as justify, the applicability of the our general approach proposed alone. In addition, auxiliary covariates, both qualitative and quantitative, can be added into our model as an extension. This EIV growth curves with auxiliary covariates in models renders a very general framework for practical application. Another illustrative example is also available to demonstrate how Bayesian approach through MCMC (Metropolis Hastings/slice sampling in Gibbs sampler) techniques as well as Bayesian Information Criterion (BIC) for model selection can be utilized in the analysis of this complex EIV growth curves with skewness in models.

166 Saddlepoint approximation for semi-Markov Processes with application to a cardiovascular randomized

study (LIPID)

[CS 25,(page 25)]

Malcolm HUDSON, *Department of Statistics, Macquarie University, Sydney, Australia*
Serigne N. LÔ, *The George Institute, University of Sydney, Australia*
Stephane HERITIER, *The George Institute, University of Sydney, Australia*

Semi-Markov processes are gaining popularity as models of disease progression in survival analysis. In this paper we extend the approach of Butler and Bronson (2002) to account for censoring and apply it to a cardiovascular trial (LIPID). The technique we propose specifies a likelihood and employs the saddlepoint techniques to fit semi-Markov model parameters. This approach is computationally simple, can model a patient's history of events represented as a flowgraph (or multistage model) and allows accurate estimation of clinically important quantities, e.g. the overall survival, hazard ratio or excess risk.

As a simple illustration, we consider a illness-death model (randomization, recovered stroke, death) for our data. Each individual passage time from one stage to another is fitted by a suitable parametric distribution and the results are then combined to provide i) a likelihood function for the network, ii) a transmittance matrix describing the process, iii) estimation of the relevant parameters. Moreover inference to test the treatment effect is available either by bootstrapping or by a direct derivation of a likelihood ratio test.

The technique is demonstrated to be flexible enough for many applications in clinical trials. The relationship between competing risk models and semi-Markov models is discussed.

167 A local likelihood approach to local dependence

[CS 18,(page 20)]

Karl Ove HUFTHAMMER, *University of Bergen, Department of Mathematics, Johannes Bruns gate 12, 5008 Bergen, Norway*
Dag TJØSTHEIM, *University of Bergen, Department of Mathematics, Johannes Bruns gate 12, 5008 Bergen, Norway*

There exist several commonly used measures of the dependence between two variables. One example is correlation, which measures the degree and direction of linear dependence. Other measures, such as Spearman's ρ and Kendall's τ, can capture nonlinear but monotone dependence, but are still global measures. When we have non-monotone association between two variables, X and Y, we might prefer a measure of dependence that can vary over the support of (X, Y), so that we can quantify the dependence as being, for example, high and positive in one area (e.g., for large X and large Y), and low and negative in a different area (e.g., for small X or small Y). We introduce a family of bivariate distributions, called locally Gaussian distributions. Locally, they have the form of a Gaussian distribution, and globally their flexibility make them appropriate approximations to real data. Using local likelihood methods developed for density estimation, we can estimate *local* parameters, and we obtain a new measure of dependence, a local correlation. The local correlation is intuitively interpretable, and shares some properties with the usual correlation (e.g., a range of $[-1, 1]$), but can vary with both X and Y. Illustrations on real and simulated data will be given.

168 Shewhart-type nonparametric control charts with runs-type signaling rules

[CS 52,(page 43)]

Schalk W. HUMAN, *Department of Statistics, University of Pretoria, Hillcrest, Pretoria, South Africa, 0002.*
Subhabrata CHAKRABORTI, *Department of Information Systems, Statistics and Management Science, University of Alabama, Tuscaloosa, USA.*
Serkan ERYILMAZ, *Department of Mathematics, Izmir University of Economics, Sakarya Caddesi, Turkey.*

New Shewhart-type nonparametric (distribution-free) control charts are proposed for monitoring the location (median) of a process when the process distribution is unknown. The charts are based on order statistics (as plotting statistics) and runs-type signaling rules with control limits given by two specified order statistics from a reference sample. Exact expressions for the run-length distributions and statistical properties (or characteristics), such as the average run-length (ARL), the standard deviation of the run-length (SDRL) and the false alarm rate (FAR), are derived analytically, using conditioning and some results from the theory of runs. Comparisons of the ARL, SDRL and some percentiles show that the new charts have robust in-control performance and are more efficient than their parametric counterparts (such as the Shewhart X-bar chart)

when the underlying distribution is t (symmetric with heavier tails than the normal) or gamma(1,1) (right-skewed). Even for the normal distribution, the new charts are competitive.

169 A mixed single and two-stage testing design for locally more powerful two-sample p test

[CS 63,(page 48)]

Wan-Ping HUNG, *Wan-Ping Hung*
Mong-Na Lo HUANG, *National Sun Yat-sen University, Taiwan, R.O.C.*
Kam-Fai WONG, *Institute of Statistics No.700,Kaohsiung University Rd.,Nan Tzu Dist., 811.Kaohsiung,Taiwan*

A main objective in clinical trials is to find the best treatment in a given finite class of competing treatments and then to show superiority of this treatment against a control treatment. Traditionally, the best treatment is estimated in a phase II trial. Then in an independent phase III trial, superiority of this treatment is to be shown against the control treatment by a level α test. The classical two sample testing methods have been applied frequently in phase III clinical trial studies. Typically, researchers choose a suitable sample size so that the study has enough power to reject the null hypothesis. For the simplest end point which records only whether individuals succeed at the end of the study, researchers then focus on understanding the difference of the success probabilities between the control group and treatment group.

In this work a mixed testing design combing certain single and two-stage testing procedures for a phase III trial is proposed and investigated. An optimal mixed testing procedure of level α and power $1-\beta$ under a specified target alternative with minimal number of individuals is presented. Later through some numerical results the superiority of the mixed testing procedure over the traditional approach is demonstrated.

170 Detectability, multiscale methods, and Statistics

[IS 3,(page 55)]

Xiaoming HUO, *Georgia Institute of Technology*

Detectability problem is to determine the fundamental boundary separate solvable detection problems and unsolvable detection problems. Detectability problems are ubiquitous in engineering. Examples include image detection for microscopic imageries, detection with very noisy sensor data, and so on. It has been shown that multiscale methodology is effective in determining boundaries in several detectability problems. Existing results in this direction will be reviewed. The existing results give the asymptotic rate of the boundaries. More precise distributional results can be derived, and they reveal more accurate properties of statistics that are used; i.e., the statistical properties right at the asymptotic boundary. Detectability problems pose several challenges in statistics and probability.

171 A new approach to Poisson approximation with applications

[IS 13,(page 30)]

H.-K. HWANG, *Institute of Statistical Science, Academia Sinica*
V. ZACHAROVAS, *Institute of Statistical Science, Academia Sinica*

A new approach to Poisson approximation problems is proposed based on properties of Charlier polynomials and Parseval's identity. The approach is simple and can also be applied to some related problems such as de-Poissonization and binomial approximation.

172 Developing a 'Next Generation' Statistical Computing Environment

[IS 1,(page 5)]

Ross IHAKA, *University of Auckland, New Zealand*
Duncan Temple LANG, *University of California, Davis*

While R has been both very successful and popular, its core is based on a computational model is at least 25 years old. And over that period, the nature of data analysis, scientific computing and statistical computing have changed dramatically. We identify some of the deficiencies of the current computational environments for dealing with current and new areas of application, and discuss some of the aspects that a new computational environment for data analysis should provide. These include core support for handling large data, a high-level language for interactive use that can also be "compiled" for efficient execution and in other systems, support for developing high-quality software, and extensibility of the core engine. Optional type specification for parameters and return types is common to all of these and a much needed facility both to improve our own work and also to be able to make statistical methods available for use in other systems. We also outline some efforts to leverage an existing platform (LISP) to build such a new

system.

173 Multi-objective models for portfolio selection

[CS 1,(page 6)]

Usta ILHAN, *Anadolu University*
Memmedaga MEMMEDLI, *Anadolu University*
Yeliz MERT KANTAR, *Anadolu University*

In portfolio management, an important problem is the optimal selection of portfolio with the aim of minimizing risks and maximizing return. Markowitz mean-variance model (MVM) has been accepted as a practical tool in the modern portfolio theory. However, MVM has not been used extensively due to computation burden, concentration of few assets and taking negative values of certain portfolio weights. Thus, to date there has been wide research concerning new methods which can avoid these probable disadvantages. For example, the mean absolute deravation (MAD) as a risk measure and the entropy model (EM) as the objective function are given. Recently, multi-objective models (MOM), which include traditional and new objective functions (entropy, skewness and suchlike), have been used to improve portfolio optimization techniques. In this study, for portfolio selection, we discuss different types of model using data taken from the Istanbul Stock Exchange (ISE 30 index) to evaluate their advantages and disadvantages. Numerical experimentations are conducted to compare the performance of these models. As a result of experimentations, appropriate models are determined for the portfolio selection with the ISE30 index.

174 Inferences on the difference and ratio of the variances of two correlated normal distributions

[CS 23,(page 22)]

Ali Akbar JAFARI, *Department of Statistics, Shiraz University, Shiraz 71454, IRAN*
Javad BEHBOODIAN, *Department of Statistics, Shiraz University, Shiraz 71454, IRAN*

Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be a random sample from a bivariate normal distribution with mean $\mu = (\mu_1, \mu_2)'$ and variance - covariance matrix Σ, i.e. $\mathbf{X}_i \sim N_2(\mu, \Sigma)$. We are interested to inferences on ratio and difference of variances.

Morgan (1939) derived the likelihood ratio test, and Pitman (1939) used these results for constructing confidence interval for the ratio variances. Also, Cacoullos (2001) obtained an F-representation of Pitman - Morgan t-test.

This article concerns inference on the variance - covariance matrix. Inferential procedures based on the concepts of generalized variables and generalized p-values are proposed for elements of Σ. The generalized confidence intervals and generalized p- values are evaluated using a numerical study. The properties of generalized variable approach and other methods are compared using Monte Carlo simulations and we find satisfactory results. The methods are illustrated using a practical example.

175 Identification of stable limits for ARCH(1) processes

[CS 8,(page 11)]

Adam JAKUBOWSKI, *Nicolaus Copernicus University, Torun, Poland*
Katarzyna BARTKIEWICZ, *Nicolaus Copernicus University, Torun, Poland*

Davis and Mikosch (Ann. Statist. 26 (1998) 2049–2080) gave limit theorems for sums and autocovariances of ARCH(1) processes with heavy tails. Applying methods of point processes, they provided a probabilistic representation for the stable limiting laws, without specifying the parameters of the limit.

The present paper contains alternative proofs based on the second author's paper (Stochastic Process. Appl. 68 (1997) 1–20), which allow to identify parameters of the limiting stable laws. The obtained formulae for parameters are given in a form suitable for Monte Carlo calculations.

176 A generalized skew two-piece skew-normal distribution

[CS 41,(page 37)]

Ahad JAMALIZADEH, *Department of Statistics, Shahid Bahonar University, Kerman, Iran*
Alireza ARABPOUR, *Department of Statistics, Shahid Bahonar University, Kerman, Iran*

The univariate skew-normal distribution was presented by Azzalini (1985,1986). A random variable Z_λ is said to have a standard skew-normal with parameter $\lambda \in \mathbf{R}$, denoted by $Z_\lambda \sim SN(\lambda)$, if its probability density function (pdf) is

$$\phi(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbf{R},$$

where $\phi(z)$ and $\Phi(z)$ denote the standard normal pdf and cumulative distribution function (cdf), respectively.

Recently,Kim (2005) presented a symmetric two-piece skew-normal distribution. A random variable X_λ is said to have a two-piece skew-normal distribution with parameter $\lambda \in \mathbf{R}$, denoted by $X_\lambda \sim TPSN(\lambda)$, if its density function is

$$f(x;\lambda) = \frac{2\pi}{\pi + 2\arctan(\lambda)}\phi(x)\Phi(\lambda|x|), \qquad x \in \mathbf{R},$$

In this paper, we present a three parameter generalized skew two-piece skew-normal, through a standard bivariate normal distribution with correlation ρ, which includes the Azzalini skew-normal in (1) and the two-piece skew-normal in (2) as special cases.

We say that a random variable $X_{\lambda_1,\lambda_2,\rho}$ has a generalized skew two-piece skew-normal distribution ,denoted by $X_{\lambda_1,\lambda_2,\rho} \sim GSTPSN(\lambda_1,\lambda_2,\rho)$, with parameters parameters $\lambda_1, \lambda_2 \in \mathbf{R}$ and $|\rho| < 1$, if

$$X_{\lambda_1,\lambda_2,\rho} \overset{d}{=} X \mid (Y_1 < \lambda_1 X, Y_2 < \lambda_2 |X|),$$

where $X \sim N(0,1)$ independently of $(Y_1,Y_2)^T \sim N_2(0,0,1,1,\rho)$. After some simple calculations we can show that the pdf of $X_{\lambda_1,\lambda_2,\rho} \sim GSTPSN(\lambda_1,\lambda_2,\rho)$ is

$$\begin{aligned} & f(x;\lambda_1,\lambda_2,\rho) \\ &= \frac{4\pi}{\cos^{-1}\left[\frac{-(\rho+\lambda_1\lambda_2)}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\right] + \cos^{-1}\left[\frac{-(\rho-\lambda_1\lambda_2)}{\sqrt{1+\lambda_1^2}\sqrt{1+\lambda_2^2}}\right] + 2\tan^{-1}(\lambda_2)} \\ &\quad \times \phi(x)\Phi_2(\lambda_1 x, \lambda_2|x|;\rho). \end{aligned}$$

Here, first we derive the normalizing constant in (4) and the moment generating function of $GSTPSN$ as explicit form. Next, we will discuss about the modes of this disribution and then we will present a probabilistic representation of $GSTPSN$ via a trivariate normal distribution. Finally, we use a numerical example to illustrate the practical usefulness of this family of distributions.

177 Quantum feedback control

[IS 25,(page 51)]

M.R. JAMES, *Australian National University*

The purpose of this talk is to explain the role of quantum probability in providing a useful framework for the engineering of quantum feedback systems. The motivation for the latter arises from recent advances in quantum technology that require the development of a new systems and control theory based on quantum mechanics. We review the basic ideas of quantum probability and feedback control, and discuss some recent theoretical and experimental results concerning optimal measurement feedback quantum control and coherent control (where the controller itself is also a quantum system).

178 Zero-sum semi-Markov ergodic games with weakly continuous transition probabilities

[CS 51,(page 43)]

Anna JASKIEWICZ, *Institute of Mathematics, Polish Academy of Sciences*

Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities and lower semicontinuous, possibly unbounded, payoff functions are studied. The two payoff criteria are considered: the ratio-average and time-average. The main result concerns the existence of a lower semicontinuous solution to the optimality equation and its proof is based on a fixed point argument. Moreover, it is shown that the ratio-average as well as time-average payoff stochastic games have the same value. In addition, one player possesses an ϵ-optimal stationary strategy ($\epsilon > 0$), whereas the other one has an optimal stationary strategy.

179 Estimating the effects of new HIV intervention methods using causal inference terchniques–the MIRA trial

[IS 21,(page 31)]

Michael A. ROSENBLUM, *University of California, Berkeley and San Francisco*
Nicholas P. JEWELL, *University of California, Berkeley*
Mark J. VAN DER LAAN, *University of California, Berkeley*
Stephen SHIBOSKI, *Univeristy of California, San Francisco*
Nancy PADIAN, *RTI, San Francisco*

The MIRA Trial is a randomized trial, designed to evaluate the use of latex diaphragms in reducing the risk of HIV infection. Part of the design of the trial included intensive counseling on the use of, and provision of, condoms in both arms of the the study. We discuss the limitation of ”Intention To Treat (ITT)” estimates of effectiveness, or efficacy, and suggest possible alternative estimation approaches that focus on the ”direct effect” of assignment to the diaphragm arm (thereby removing the ”indirect effect through condom use”). Results will be compared and discussed. Statistical ideas include causal graphs, causal inference, and inverse weighted estimators. The purpose of the talk is to stimulate interest and discussion regarding strict adherence to ITT procedures in complex intervention trials, and suggest possible design modifications.

180 A bivariate chi-square distribution and some of its properties

[CS 76,(page 57)]

Anwar H JOARDER, *Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia*
Mohammed H. OMAR, *Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia*
Abdallah LARADJI, *Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia*

Distributions of the sum, difference, product and ratio of two chi-squares variables are well known if the variables are independent. In this paper, we derive distributions of some of the above quantities when the variables are correlated through a bivariate chi-square distribution and provided graphs of their density functions. The main contribution of the paper is the marginal distribution and closed form expressions for product moments and conditional moments. Results match with the independent case when the variables are uncorrelated.

181 Random matrix universality

[IS 16,(page 55)]

Kurt JOHANSSON, *Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden*

Limit laws and limit processes coming out of random matrix theory have turned out to be natural limit laws that occur in many contexts not all related to spectral problems. Examples are the sine kernel point process and the Tracy-Widom distribution. Within random matrix theory an important aspect of the universality problem is to prove that the local statistics of broad classes of probability measures on symmetric or hermitian matrices is universal and does not depend on the specific choice of probability distribution. The talk will survey some aspects of the random matrix universality problem.

182 Probablistic analysis of an electric supply system

[CS 12,(page 13)]

M.S. KADYAN, *Kurukshetra University,Kurukshetra, India*
Suresh Chander MALIK, *Department of Statistics, M.D.University, Rohtak-124001, India*

This paper deals with a stochastic model of an electric supply system having two non-identical units in which one unit is an electric transformer which fails completely via partial failure and the other unit generator which has a direct failure from normal mode is studied by using semi-Markov process and regenerative point technique. There is a single server which plays the dual role of repair and inspection. The electric transformer gets priority in operation as well as in repair over the generator .The failed generator is first inspected by the server to see the possibility of its repair or replacement by new one in order to avoid the unnecessary expanses of the repair. The repair and inspection times distributions are taken as general with different probability density functions while the failure time distributions of both units follow negative exponential with different parameters. The behaviour of mean time to system failure, availability and profit function of the system have also been studied through graphs.

183 Robustification of the PC-algorithm for directed acyclic graphs

[CS 54,(page 44)]

Markus KALISCH, *Seminar for Statistics ETH Zurich*
Peter BÜHLMANN, *ETH Zurich*

The PC-algorithm (Spirtes, Glymour and Scheines, 2000) was shown to be a powerful method for estimating the equivalence class of a potentially very high-dimensional acyclic directed graph (DAG) with corresponding Gaussian distribution (Kalisch and Buehlmann, 2007). Here we will propose a computationally efficient robustification of the PC-algorithm and prove its consistency. Furthermore, we will compare the robustified and standard version of the PC-algorithm on simulated data using the corresponding R-package pcalg.

184 Inference for stochastic volatility models using time change transformations

[CS 33,(page 32)]

Kostas KALOGEROPOULOS, *University of Cambridge, Engineering Department - Signal Processing Laboratory*
Gareth ROBERTS, *University of Warwick, UK*

Petros DELLAPORTAS, *Athens University of Economics and Business, Statistics Department*

Diffusion processes provide a natural tool for modeling phenomena evolving continuously in time. The talk focuses on diffusion driven stochastic volatility models, used extensively in Econometrics and Finance. The task of inference is not straightforward as we can only observe a finite number of the infinite dimensional diffusion path, and the marginal likelihood for these observations is generally not available in closed form. Furthermore some components of the diffusion process, i.e. the volatility, are entirely unobserved. Consequently, the observed process is not generally Markov. A natural approach to the problem is through data augmentation (Tanner and Wong 1987, JASA 82:528-550), utilizing Markov chain Monte Carlo (MCMC) techniques that impute fine partitions of the unobserved diffusion. However, as noted in (Roberts and Stramer 2001, Biometrika 88:603-621), appropriate reparametrisation is essential to avoid degenerate MCMC algorithms due to the perfect correlation between parameters and imputed diffusion paths.

We introduce a novel reparametrisation of the likelihood through transformations that operate mainly on the time axis of the diffusion rather than its path. The time change transformations may depend on parameters or latent diffusion paths and therefore the adoption of a MCMC algorithm is essential. To deal with the issue of multiple and parameter dependent time scales, we also construct a suitable MCMC scheme equipped with retrospective sampling techniques. The time change reparametrisation ensures that the efficiency of the MCMC algorithm does not depend on the level of augmentation, thus enabling arbitrarily accurate likelihood approximations. The algorithm is fast to implement and the framework covers general stochastic volatility models with state dependent drifts. We illustrate the methodology through simulation based experiments and a real-data application consisting of US Treasury Bill rates.

185 Statistical tests for differential expression in microarray data using generalized p-value technique with shrinkage

[CS 69,(page 52)]

Kanichukattu Korakutty JOSE, *Dept. of Statistics, St. Thomas College, Pala, Mahatma Gandhi University, Kerala, India-686574*

J. Sreekumar JANARDHANAN, *Scientist (Agricultural Statistics), Central Tuber Crops Research Institute,Trivadrum, Kerala, India-695017*

Lishamol Tomy MUTHIRAKALAYIL, *Dept. of Statistics, St. Thomas College, Pala, Mahatma Gandhi University, Kerala, India-686574*

The greatest challenge to microarray technology lies in the analysis of gene expression data to identify which genes are differentially expressed across tissue samples or experimental conditions. The distribution of log transformed values of the gene expression data is usually assumed to be approximately distributed as normal to model microarray data. The sample variances are not homogeneous in expression data and generalized p-value method has been successfully used to provide finite sample solutions for many testing problems in such situations. The current approaches for selection of significant genes involve different t-test approaches with correction designed to control the false discovery rate. In the present study the generalized p-value method for comparison of two lognormal means has been applied to the raw dataset without log transformation to test the differential expression of individual genes. The idea of shrinkage has been applied in generalized p value and the generalized p value methods were compared with ordinary t- test, t-test with unequal variance, moderated t, bootstrapped t-test, and SAM. The expression profiles of acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) samples are compared in the training dataset and the independent dataset from Golub et al., 1999. Numerical results confirmed the superiority of the procedure based on the generalized p-value technique with shrinkage to identify genes with a low level of false discovery rate.

186 Comparison of the power of robust F* and non-parametric Kruskal-Wallis tests in one-way analysis of variance model

[CS 23,(page 22)]

Yeliz Mert KANTAR, *Anadolu University, Science Faculty, Department of Statistics, 26470, Eskisehir, Turkey*

Birdal SENOGLU, *Ankara University, Department of Statistics, 06100 Tandoan, Ankara, Turkey*

Observations made on many agricultural, biological or environmental variables do not often follow a normal distribution. Moreover, outliers may exists

in observed data. In these situations, normal theory tests based on least squares (LS) estimators have low power and are not robust against plausible deviations from the assumed distribution. Therefore, we resort to nonparametric test procedures to analyze the non-normal data. In the context of one-way analysis of variance, the well-known nonparametric Kruskal-Wallis test based on ranks is used to compare the three or more groups of observations. In this paper, we compare the power and the robustness properties of the Kruskal-Wallis test with the test developed by Senoglu and Tiku (2004) when the distribution of error terms are type II censored generalized logistic. Simulation results show that the test is more powerful and robust in general. A real data set from the literature is analyzed by using the test based on modified maximum likelihood (MML) estimators. A Compaq Visual Fortran program is used for the calculations and the executable program is available from the author on request.

187 Computation of the mixing proportions of normal mixture model for QTL mapping in the advanced populations derived from two inbred lines

[CS 9,(page 11)]

Chen-Hung KAO, *Institute of Staistical Science, Academia Sinica, Taiwan, ROC*
Miao-Hui ZENG, *Institute of Staistical Science, Academia Sinica, Taiwan, ROC*

When applying normal mixture model to the estimation of quantitative trait loci (QTL) parameters, the mixing proportions are determined by computing the conditional probabilities of the putative QTL genotypes given their flanking marker genotypes. Therefore, the quality of QTL detection will rely heavily on the correct computation of the onditional probabilities. When the sample is from the backcross or F_2 population, the computation of the conditional probabilities is straightforward as the population genome structure has the Markovian property. However, when the sample is from the progeny populations of the F_2 population, obtaining the conditional probabilities is not simple, and further considerations are needed in derivation as the genomes are no longer Markovian. This talk focuses on the issues of deriving the conditional distribution of the putative QTL under the framework of interval mapping procedure.

188 Agreement analysis method in case of continuous variable

[PS 3,(page 29)]

Kulwant Singh KAPOOR, *All India Institute Of Medical Sciences Ansari Nagar New Delhi 110029 , INDIA*

In clinical and epidemiological studies research are very much interested to know the inter observer variation in a continuous variable or two measurement techniques .

Example . Measurement of blood pressure with pulse oximetry and ausculatory method or measurement of PEFR respiratory diseases by wright peak flow meter and mini wright meter in other case pulse rate of patient measure by two nurse or doctor

The conventional Statistical method applied for studying the agreement between two method of measuring a continuous variable is computing the Correlation Coefficient (r) , but many times this is misleading for this purpose. A change of scale of measurement does not alter r but affect the agreement . In order to overcome this difficulty we will apply five test and in case three will come out to be true we can say that there is good agreement exist between two rater or techniques

189 Estimation of engel curve by partially linear model

[CS 68,(page 51)]

Sri Haryatmi KARTIKO, *Gadjah Mada University, Yogyakarta, Indonesia*

Most papers investigate consumer behaviour in a nonparametric context that are used as a way of modelling the form of Engel curve. Those focused on the unidimensional nonparametric effects of log total expenditures on budget expenditures. Since some independent variables have parametric while others have nonparametric relationship with the response, an additive partially linear model is used to estimate semiparametrically the relationship in the context of Engel curve. Empirical results obtained from the application of additive partially linear model is presented, while the result of additive specification as well as the linearity of the nonparametric component is also proposed.

190 On empirical Bayes tests for continuous distributions

[CS 31,(page 28)]

R.J. KARUNAMUNI, *University of Alberta*

We study the empirical Bayes two-action problem

under linear loss function. Upper bounds on the regret of empirical Bayes testing rules are investigated. Previous results on this problem construct empirical Bayes tests using kernel type estimates of nonparametric functionals. Further, they have assumed specific forms, such as the continuous one-parameter exponential family for the family of distributions of the observations. In this paper, we present a new unified approach of establishing upper bounds (in terms of rate of convergence) of empirical Bayes tests for this problem. Our results are given for any family of continuous distributions and apply to empirical Bayes tests based on any type of nonparametric method of functional estimation. We show that our bounds are very sharp in the sense that they reduce to existing optimal rates of convergence when applied to specific families of distributions.

191 A random effects formulation of high-dimensional Bayesian covariance selection

[CS 11,(page 12)]

Jessica KASZA, *School of Mathematical Sciences, University of Adelaide*
Gary GLONEK, *School of Mathematical Sciences, University of Adelaide*
Patty SOLOMON, *School of Mathematical Sciences, University of Adelaide*

In a microarray experiment, it is expected that there will be correlations between the expression levels of different genes under study. These correlation structures are of great interest from both biological and statistical points of view. From a biological perspective, the correlation structures can lead to an understanding of genetic pathways involving several genes, while the statistical interest lies in the development of statistical methods to identify such structures. However, the data arising from microarray studies is typically very high-dimensional, with an order of magnitude more genes being analysed than there are slides in a typical study. This leads to difficulties in the estimation of the dependence structure of all genes under study. Bayesian graphical models can be used in such a situation, providing a flexible framework in which restricted dependence structures can be considered.

Dobra *et al* (2004) utilise such models in the analysis of the dependence structure of microarray data, using their technique “High-dimensional Bayesian Covariance Selection”, or HdBCS. While this technique allows for the analysis of independent, identically distributed gene expression levels, often the data available will have a complex mean structure and additional components of variance. For example, we may have gene expression data for genes from several different geographical sites. Our inclusion of site effects in the HdBCS formulation allows such data to be combined, and the dependence structure of the genes estimated using all of the data available for each gene, instead of examining the data from each site individually. This is essential in order to obtain unbiased estimates of the dependence structure. This site effects formulation can be easily extended to include more general random effects, so that any covariates of interest can be included in the analysis of dependence structure.

References

1. A. Dobra *et al.* Sparse Graphical Models for Exploring Gene Expression Data. Journal of Multivariate Analysis 90 (2004) 196-212.

192 A family of asymmetric distributions on the circle with links to Möbius transformation

[CS 17,(page 19)]

Shogo KATO, *Institute of Statistical Mathematics*
M. C. JONES, *The Open University*

We propose a family of four-parameter asymmetric distributions on the circle that contains the von Mises and wrapped Cauchy distributions as special cases. The family can be derived by transforming the von Mises distribution via Möbius transformation which maps the unit circle onto itself. Some properties of the proposed model are obtainable by applying the theory of the Möbius transformation. The density of the family represents a symmetric or asymmetric, unimodal or bimodal shape, depending on the choice of the parameters. Conditions for unimodality or symmetry are given. The proposed family is used to model an asymmetrically distributed dataset and discussion on the goodness of fit is briefly made.

193 Functional ANOVA modeling of regional climate model experiments

[IS 26,(page 36)]

Cari KAUFMAN, *University of California, Berkeley*
Stephan SAIN, *National Center for Atmospheric Research, Boulder, CO, USA*

We present functional ANOVA models for at-

tributing sources of variability in climate models, specifically regional climate models (RCMs). RCMs address smaller spatial regions than do global climate models (GCMs), but their higher resolution better captures the impact of local features. GCM output is often used to provide boundary conditions for RCMs, and it is an open scientific question how much variability in the RCM output is attributable to the RCM itself, and how much is due simply to large-scale forcing from the GCM. Illustrating with data from the Prudence Project, in which RCMs were crossed with GCM forcings in a designed experiment, we will present a framework for Bayesian functional ANOVA modeling using Gaussian process prior distributions. In this framework, we obtain functional and fully Bayesian versions of the usual ANOVA decompositions, which can be used to create useful graphical displays summarizing the contributions of each factor across space.

194 Greeks formulae for an asset price dynamics model with gamma processes

[CS 68,(page 51)]

Reiichiro KAWAI, *Center for the Study of Finance and Insurance, Osaka University, Japan*

Atsushi TAKEUCHI, *Department of Mathematics, Osaka City University, Japan*

Following the well known Girsanov transform approach of Bismut (1983) with a parameter separation property of gamma processes on the Esscher transform, we derive Greeks formulae of delta, rho, vega and gamma in closed form for an asset price dynamics model formulated with the gamma process and with the Brownian motion time-changed by an independent gamma process. Our results differ from those of Davis and Johansson (2006) and of Cass and Friz (2007) in the sense that our model can be of a pure-type type, while improving those of El-Khatib and Privault (2004) in that our model is formulated with Lévy processes of a more realistic infinite activity type.

195 Semi-supervised logistic discrimination via regularized basis expansions

[CS 40,(page 37)]

Shuichi KAWANO, *Graduate School of Mathematics, Kyushu University*

Sadonori KONISHI, *Faculty of Mathematics, Kyushu University*

The classification or discrimination technique is one of the most useful statistical tools in various fields of research, including engineering, artificial intelligence and bioinformatics. In practical situations such as medical diagnosis, labeling data sets may require expensive tests or human efforts and only small labeled data sets may be available, whereas unlabeled data sets can be easily obtained. Recently, a classification method that combines both labeled and unlabeled samples, called as semi-supervised learning, has received considerable attention in the statistical and machine learning literature.

Various model approaches have been taken to exploit information from the sets of labeled and unlabeled data; e.g., a mixture model approach, a logistic discriminant model approach, a graphical model approach and so on. A logistic discriminant model approach constructs models by extending linear logistic discriminant models to cope with additional unlabeled data, and unknown parameters in the model are estimated by the maximum likelihood method. This method, however, has some drawbacks. First, the estimated models cannot capture complex structures with the nonlinear decision boundaries as the models produce only the linear decision boundaries. Second, a large number of predictors lead to unstable or infinite maximum likelihood parameter estimates and, consequently, may result in incorrect classification results.

To overcome these problems, we present semi-supervised nonlinear logistic models based on Gaussian basis expansions. We use Gaussian basis functions with hyper-parameter, which provide a clear improvement for classification results. In order to avoid the ill-posed problem, the unknown parameters are estimated by the regularization method along with the technique of EM algorithm. The crucial points for model building process are the choice of the number of basis functions and of the values of the regularization parameter and hyper-parameter including in Gaussian basis functions. To choose the adjusted parameters we introduce a Bayesian type criteria for evaluating models estimated by the method of regularization. The numerical examples are conducted to investigate the effectiveness of our modeling strategies.

196 The power of the allele-based N-test in linkage analysis

[CS 21,(page 21)]

Sajjad Ahmad KHAN, *Department of Human Genetics, University of Pittsburgh, PA, USA. and Department*

of Statistics, University of Peshawar, Peshawar, NWFP, Pakistan.
Shuhrat SHAH, *Department of Statistics, University of Peshawar, Peshawar, NWFP, Pakistan.*
Daniel E WEEKS, *Department of Human Genetics, University of Pittsburgh, PA, USA. and Department of Biostatistics, University of Pittsburgh, PA, USA*

There are many tests of inheritance based upon sibling information for diseases that have late onset. The N-test (Green et al. 1983) is one of these tests, which utilizes information from affected siblings. The N-test is the count in affected siblings of the most frequently occurring haplotype from the father plus the analogous count from the mother. When applied to haplotypes, the N-test excludes recombinant families from the analysis. In this study we modified the N-test to be based on alleles instead of haplotypes. This modified allele-based N-test can include all families (recombinant as well as non-recombinant). We carried out a simulation study to compare the power of the allele-based N-test with the powers of the Sall and Spairs non-parametric statistics as computed by Merlin. The powers of the allele-based N-test, Sall and Spairs statistics are identical to each other for affected sibships of size 2 and 3. For affected sibships of larger sizes, the powers of the Sall and Spairs statistics are larger than the power of allele-based N-test. These simulation-based results are consistent with earlier results based on analytical computations.

197 Statistical Study of the socio-economic factors of Tuberculsis in Kohat district (Pakistan)

[CS 79,(page 58)]

Salahuddin KHAN, *Department of Statistics University of Peshawar, Pakistan*
M. Karim KHATTACK, *Deptt: of Statistics, Government Post Graduate College Kohat, Pakistan*

Abstract:- Tuberculosis is a chronic or acute bacterial infection that primarily attacks the lungs, but it may also affect any part of the body, particularly the kidneys, bones, lymph nodes, and brain. It is caused by Mycobacterium tubercle (a rod- shaped bacterium). In this study an effort has been made to determine the socio-economic factors of Tuberculosis, popularly known as TB, in Kohat District, Pakistan. For this purpose, a random sample of 935 Patients collected from various hospitals of Kohat district was investigated to determine the socio-economic factors of tuberculosis. Since all factors were categorical, therefore log-linear analysis was the appropriate statistical technique. Log-linear modeling is essentially a discrete multivariate technique that is designed specially for analysis of data when both the independent and dependent variables are categorical or nominal. The factors used for tuberculosis, included in the log-linear analysis were residence, age, sex, population and economic status. The best-fitted log-linear model, consisting of one three-factor interaction, two four-factor interactions, and one five-factor interaction, was selected by means of backward elimination procedure. In order to make easy the interpretation of the interactions included in the model, a statistical technique of binary logistic regression was applied. Tuberculosis, population and economic status were first three-factor association. The interaction between population and economic status was significantly related to the log odds of tuberculosis positive. Tuberculosis, age and sex, and population interaction was the first four-factor interaction of selected log-linear model. This interaction when analyzed at different levels of age, sex, and population by applying logistic regression showed that the interaction among age, population and sex, that is, overcrowded population of male sex of age 15-54 years, was significantly related to the log odds of tuberculosis positive. Tuberculosis, residence and age, sex and economic status interaction was the only five factor interaction of the best-selected log-linear model. Fit of the model shows that the interaction among residence

198 BEC in statistical mechanics

[CS 72,(page 53)]

Abidin KILIC, *Anadolu University, Turkey*
Secil ORAL,

The generalized Bose-Einstein distribution, within the dilute gas assumption, in Tsallis statistics is worked without approximation for the Bose-Einstein condensation (BEC). In order to promote a complete analysis for the BEC in the statistics we also find exact expression within the normalized constraints in a harmonic trap. In this last decade, we have witnessed a growing interest in the Tsallis statistics. The starting point of Tsallis statistics is based on the nonextensive entropy It has been applied in many situations such as, Euler turbulence, self-gravitating and correlated systems politrop, and among others. In particular, the Bose-Einstein and Fermi-Dirac distributions have been extensively ana-

lyzed in the Tsallis framework. For the most of these cases the analyses involving these generalized Bose-Einstein and Fermi-Dirac distribution have been restricted for free systems, i.e., the interaction between the particles is absent.

199 A note on the central limit theorem for bipower variation of general functions

[CS 74,(page 54)]

Silja KINNEBROCK, *Oxford-Man Institute of Quantitative Finance, University of Oxford*
Mark PODOLSLIJ, *University of Aarhus*

In this paper we present a central limit theorem for general functions of the increments of Brownian semimartingales. This provides a natural extension of the results derived in Barndorff-Nielsen, Graversen, Jacod, Podolskij & Shephard (2006), who showed the central limit theorem for even functions. We prove an infeasible central limit theorem for general functions and state some assumptions under which a feasible version of our results can be obtained. Finally, we present some examples from the literature to which our theory can be applied.

200 Model selection with multiply-imputed data

[CS 45,(page 39)]

S. K. KINNEY, *National Institute of Statistical Sciences*
J. P. REITER, *Duke University*
J. O. BERGER, *Duke University*

Several statistical agencies use, or are considering the use of, multiple imputation to limit the risk of disclosing respondents' identities or sensitive attributes in public use data files. For example, agencies can release partially synthetic datasets, comprising the units originally surveyed with some values, such as sensitive values at high risk of disclosure or values of key identifiers, replaced with multiple imputations. Methods for obtaining inferences for scalar and multivariate estimands have been developed for partially synthetic datasets, as well as several other types of multiply-imputed datasets. Methods for conducting model selection with multiply-imputed data are very limited, and this paper moves toward filling this need. As a first step, a simple case is considered in the context of partially synthetic data. It is assumed that the analyst has some knowledge about the imputation procedure. In this scenario, a Bayesian approach to implementing model selection with partially synthetic datasets and a Bayes factor approximation similar to the BIC are derived and illustrated. We also consider how these procedures can be generalized beyond the simple case and extended to multiple imputation for missing data.

201 Ruin analysis in the constant elasticity of variance model

[CS 50,(page 43)]

Fima KLEBANER, *Monash University*
Robert LIPTSER, *Tel Aviv University*

We give results on the probability of absorption at zero of the diffusion process with non-Lipschitz diffusion coefficient

$$dX_t = \mu X_t dt + \sigma X_t^\gamma dB_t,$$

with $X_0 = K$, and $1/2 \leq \gamma < 1$. Let τ be time to ruin $\tau = \inf\{t : X_t = 0\}$. We give the probability of ultimate ruin, and establish asymptotics

$$\lim_{K\to\infty} \frac{1}{K^{2(1-\gamma)}} \log P(\tau \leq T)$$

We also find an approximation to the most likely paths to ruin for large K. The asymptotics in K is obtained by proving the Large Deviations Principle (LDP) and solving a control problem.

202 Minimaxity of Stein-type Bayesian prediction for normal regression problem

[CS 42,(page 38)]

Kei KOBAYASHI, *Institute of Statistical Mathematics*
Fumiyasu KOMAKI, *University of Tokyo*

We consider Bayesian shrinkage predictions for the Normal regression problem under the frequentist Kullback-Leibler risk function. The result is an extension of Komaki (2001, Biometrika) and George (2006, Annals. Stat.).

Firstly, we consider the multivariate Normal model with an unknown mean and a known covariance. The covariance matrix can be changed after the first sampling. We assume rotation invariant priors of the covariance matrix and the future covariance matrix and show that the shrinkage predictive density with the rescaled rotation invariant superharmonic priors is minimax under the Kullback-Leibler risk. Moreover, if the prior is not constant, Bayesian predictive density based on the prior dominates the one with the uniform prior.

In this case, the rescaled priors are independent of the covariance matrix of future samples. Therefore, we can calculate the posterior distribution and the mean of the predictive distribution (i.e. the posterior mean and the Bayesian estimate for quadratic loss) based on some of the rescaled Stein priors without knowledge of future covariance. Since the predictive density with the uniform prior is minimax, the one with each rescaled Stein prior is also minimax.

Next we consider Bayesian predictions whose prior can depend on the future covariance. In this case, we prove that the Bayesian prediction based on a rescaled superharmonic prior dominates the one with the uniform prior without assuming the rotation invariance.

Applying these results to the prediction of response variables in the Normal regression model, we show that there exists the prior distribution such that the corresponding Bayesian predictive density dominates that based on the uniform prior. Since the prior distribution depends on the future explanatory variables, both the posterior distribution and the mean of the predictive distribution may depend on the future explanatory variables.

The Stein effect has robustness in the sense that it depends on the loss function rather than the true distribution of the observations. Our result shows that the Stein effect has robustness with respect to the covariance of the true distribution of the future observations.

As the dimension of the model becomes large, the risk improvement by the shrinkage with the rescaled Stein prior increases. An important example of the high dimensional model is the reproducing kernel Hilbert space. Therefore Bayesian prediction based on shrinkage priors could be efficient for kernel methods.

203 An application of the fractional probability weighted moments for estimating the maximum entropy quantile function

[CS 38,(page 34)]

S. KORDNOURIE, *Department of statistics ,The Islamic Azad University,North Tehran Branch.Iran*
H. MOSTAFAEI, *Department of statistics ,The Islamic Azad University,North Tehran Branch.Iran*

Abstract In this article we will show an efficient method for extreme quantile estimation by applying a Maximum entropy principle from a small sample of data.We will used a probability weighted moments (PWMs) of integral orders instead of product moments that mostly used in Maximum entropy principle.For improving the approximation of distribution tail we will use the fractional PWMs over integral orders PWMs or IPWMs.We combine the optimization algorithms and monte carlo simulations to estimate the fractional probability weighted moments(FPWMs).Finally we will compare the accuracy of FPWMs according to the quantile function with usual IPWMs by the numerical example.

204 On reducing variance by smoothing influence function of statistical functionals

[CS 49,(page 42)]

Andrzej KOZEK, *Macquarie University, DEFS, Department of Statistics*

Let $\mathcal{K}$ be a class of Hadamard-differentiable statistical functionals with the influence function being twice continuously differentiable with an exception of at most of a finite number of points. This class includes all statistical functionals useful in practice. We characterize in terms of the influence function the cases where functionals with kernel-smoothed influence function have lower asymptotic variance than the variance of the corresponding non-smoothed functionals. This result is related to the partial results obtained in Fernholz (1997), even though the concepts of *smoothing* in both papers differ slightly: we consider functionals corresponding to smoothed influence functions while in Fernholz (1997) the original functionals are evaluated at a smoothed empirical cumulative distribution function. We show in particular that smoothing of the influence function can be beneficial when the influence function or its derivative have jumps. As the smoothing may cause some asymptotic bias we suggest a simple regression based method reducing the asymptotic bias effect.

205 On efficiency and alarm system in reinsurance contracts.

[CS 1,(page 6)]

Marie KRATZ, *ESSEC Business School Paris and MAP5 Univ. Paris Descartes.*
Shubhabrata DAS, *Indian Institute Of Management Bangalore (IIMB)*

Insurance companies protect themselves from large claims by entering into reinsurance contracts in

exchange for sharing part of the premiums. One popular criterion for selecting appropriate form of reinsurance contract is the benefit in the survival probability of the primary insurer (cedent) through entering such contracts. Recent literature (Ignatov et al. (2004), Kaishev et al. (2006)) has studied the problem by looking at the cedent and reinsurers perspective simultaneously; however such research is limited to one to one relationship between cedent and reinsurer. In practice, a reinsurer has reinsurance contract with multiple cedent companies. Thus the reinsurer may survive a lean period from a particular contract thanks to the financial status in the other reinsurance contracts. One goal of the current work is to exhibit this phenomenon through a model involving single reinsurer and multiple cedents. While we focus on Excess of Loss contracts, we plan to cover other reinsurance schemes and compare their efficiencies. We consider two alternative formulations of the efficiency measures of the reinsurance system, depending on whether the contracts are identical across all the cedents or not. A second motivation of the study is to explore the effectiveness of having multiple layers of reinsurance contracts in the system. Towards this we propose a modified version of the efficiency measure(s) and study its behaviour. The efficiency measures help in selecting one among the possible reinsurance schemes as well as specific choice of optimal parameter, like retention level in Excess of Loss contract, or number of reinsurance layers. An additional way of risk management for the (re-) insurance company is to develop an early and appropriate alarm system before the possible ruin. In that case, the problem boils down to the determination of a suitable level for the risk process which corresponds to a minimum pre-specified high probability of ruin within a given timeframe after the alarm. The formulation may be generalized from covering a single risk process to multiple ones, extending the concept of alarm system to reinsurance contracts.

206 Conformal geometry of random triangulations

[IS 18,(page 19)]

Maxim KRIKUN, *Institut Élie Cartan, Université Nancy 1*
Oded SCHRAMM, *Microsoft Research*

Given a planar triangulation, endow it with a complex structure by declaring every triangle to be equilateral. By the uniformization theorem the resulting Riemann surface possesses a unique (up to Mobius transformation) conformal map to the sphere. If the original triangulation was taken uniformly at random within some class, this randomness is captured by the image of the Lebesgue measure under the conformal map. The resulting random measure can then be probed using the planar brownian motion.

207 Statistics for climate prediction: uncertainty and biases

[IS 26,(page 36)]

Hans R. KUENSCH, *Seminar fur Statistik ETH Zurich Switzerland*
Christoph M. BUSER, *Seminar for Statistics, ETH Zurich, Switzerland*
Christoph SCHAER, *Institute for Atmospheric and Climate Sciene, ETH Zurich, Switzerland*

The basis of climate predictions are global circulation models (GCMs) which are deterministic. In order to obtain a finer resolution, Regional climate models (RCMs) which take the output of a GCM for boundary conditions are used. Several GCMs and RCMs are available, and they all are run under a number of different emission scenarios. As a result, one obtains a large number of predictions which differ considerably. Bayesian methods provide an attractive framework for combining these predictions and assessing uncertainty. All models predict also the current climate in a control run. When comparing these predictions with observations on the current climate, one sees that most models have considerable additive and multiplicative biases. Often it is assumed that these biases are unchanged when the future climate under a given emission scenario is predicted, but this is debatable. In this talk, we show how we can replace the assumption of no bias change by an informative prior on the size of the bias change. This is possible because in contrast to previous studies we consider not only the mean values, but also the interannual variation. There are however two plausible methods for extrapolating the multiplicative bias of the control run into the future which lead in some cases to substantially different results.

208 Wavelet regression in random design for long memory processes

[CS 60,(page 47)]

Rafal KULIK, *Department of Mathematics and Statistics University of Ottawa*
Marc RAIMONDO, *University of Sydney*

We investigate global performance of non-linear wavelet estimation in random-design regression models with long memory errors. The setting is as follows. We observe $Y_i = f(X_i) + \epsilon_i$, $i = 1, \ldots, n$, where $X_i, i \geq 1$, are (observed) independent identically distributed (i.i.d.) random variables with a distribution function G, $\epsilon_i, i \geq 1$ is a stationary Gaussian dependent sequence with a covariance function $\rho(m) \sim m^{-\alpha}$, $\alpha \in (0, 1)$.

For nonlinear wavelet estimator $\hat{f}_n$ we prove that the expected L_p risk is

$$E\|f - \hat{f}_n\|_p^p \leq Cn^{-\gamma(s)}(\log n)^{2\gamma(s)},$$

where

$$\gamma(s) = \begin{cases} \frac{ps}{2s+1}, & \text{if } s > \frac{p-\pi}{2\pi} \text{ and } \alpha > \alpha_0, \\ p\frac{s-\left(\frac{1}{\pi}-\frac{1}{p}\right)}{1+2\left(s-\frac{1}{\pi}\right)}, & \text{if } s < \frac{p-\pi}{2\pi} \text{ and } \alpha > \alpha_1. \end{cases}$$

Above, α_0 and α_1 are values in $(0, 1)$ which depend on a Besov class parameters s, π and on p in the sparse case. In particular, the estimator achieves the same rates as in i.i.d. case. For the other set of parameters (s, α),

$$E\|f - \hat{f}_n\|_p^p \leq Cn^{-p\alpha/2}.$$

Furthermore, we construct an estimator for $f - \int f$. It always achieves i.i.d. rates, i.e. its expected L_p risk is bounded by $Cn^{-\gamma(s)}(\log n)^{2\gamma(s)}$.

Our obtained rates of convergence agree (up to the log term) with the minimax rates of Yang 2001. From a probabilistic point of view the main new ingredient of our proof is a large deviation result for long memory sequences.

209 Information theoretic models for dependence analysis and missing data estimation

[CS 45,(page 39)]

Parmil KUMAR, *Dept. of Statistics,University of Jammu, Jammu,India*

D. S. HOODA, *Jaypee Institute of Engineering and Technology, Guna, India*

In the present paper we have derived a new information theoretic model for testing and measurement of dependence among attributes in a contingency table. A relationship between information theoretic measure and chi-square statistic has also been established and discussed with numerical illustrations. A new generalized information theoretic measure is also defined and studied. Maximum entropy model for estimation of missing data in design of experiment is also discussed.

210 Michaelis-Menten and reduction of stochastic models for enzyme reactions

[IS 14,(page 5)]

Thomas G. KURTZ, *University of Wisconsin-Madison*

David F. ANDERSON,

Gheorghe CRACIUN,

The classical determinisitic Michaelis-Menten equation for enzyme kinetics can be obtained from the stochastic model for a simple enzyme reaction network by a simple averaging argument. More complex models of enzyme reaction networks present more complex averaging problems. We address some of these complexities by exploiting results that show many of the fast subnetworks have product-form stationary distributions. That fact allows explicit calculation of the averaged model.

211 Asymmetric Laplace processes

[CS 44,(page 39)]

Jayakumar KUTTAN PILLAI, *Associate Professor, Department of Statistics, University of Calicut, Kerala-673 635, India*

The Laplace distribution is symmetric, and there were several asymmetric extensions in generalizing the Laplace distribution. The class asymmetric Laplace distributions with characteristic function arise as limiting distribution of a random (geometric) sum of independent and identically distributed random variables with finite second moments. Hence the class of asymmetric Laplace distributions forms a subclass of geometric stable distributions where the geometric stable distributions, similarly to stable laws, have tail behavior governed by the index of stability. The asymmetric Laplace distribution plays an analogues role among geometric stable laws as Gaussian distributions do among the stable laws. Asymmetric Laplace distribution has found applications in the fields of biology, financial mathematics, environmental science etc. In this paper, we introduce a two-parameter time series model that is free from the drawback 'zero defects' using asymmetric Laplace marginal distribution. The innovation sequence of the process is obtained as sum of two independent random variables of which one is a convex mixture of exponential random variables with different parameters and other is an asymmetric Laplace random variable. The properties of the process are studied. Also we develop a three-parameter autoregressive model using Laplace, asymmetric Laplace and semi -Laplace variables. If the marginal distribution of the stationary three- parameter autoregressive process is Laplace then we obtain the solution of innovation sequence as a convex mixture of Laplace random variables. Further we establish the condition for existence of stationary solution of the three-parameter autoregressive process when the marginal distribution is semi -Laplace distribution. Also the condition for existence of solution with negative weights is given. The autocorrelation function and joint characteristic function and sample path behavior of the process are obtained. A second order autoregressive process with Laplace marginal distribution is introduced and the distribution of the innovative sequence in this case is obtained. The first order moving average process with Laplace marginal distribution is also developed.

212 Modelling healthcare associated infections: a Bayesian approach.

[CS 81,(page 59)]

Theodore KYPRAIOS, *University of Nottingham, United Kingom*

Philip D. O'NEILL, *University of Nottingham, United Kingom*

Ben COOPER, *Health Protection Agency, United Kingdom*

There are large knowledge gaps in both the epidemiology and population biology of major nosocomial pathogens such as methicillin-resistant Staphylococcus aureus (MRSA) and glycopeptide-resistant enterococci (GRE). We are interested in answering questions such as: what value do specific control measures have? how is transmission within a ward related with colonisation pressure? what effects do different antibiotics play? Is it of material benefit to increase or decrease the frequency of the swab tests? what enables some strain to spread more rapidly than others?

Most approaches in the literature to answering questions such those listed above are based on coarse aggregations of the data. Although using aggregated data is not necessarily inappropriate, the limitations of such an approach have been well-documented in the literature. In addition, when individual-level data are available, at present most authors simply assume outcomes to be independent. First, such independence assumptions can rarely be justified; moreover, it has been shown that failing to account for such dependencies in the data will result in incorrect inferences and lead to major errors in interpretation.

Our approach is to construct biologically meaningful stochastic epidemic models to overcome unrealistic assumptions of methods which have been previously used in the literature, include real-life features and provide a better understanding of the the dynamics of the spread of such major nosocomial pathogens within hospital wards. We implement Markov Chain Monte Carlo (MCMC) methods to efficiently draw inference for the model parameters which govern transmission. Moreover, the extent to which the data support specific scientific hypotheses is investigated by considering different models. Trans-dimensional MCMC algorithms are employed for Bayesian model choice. The developed methodology is illustrated by analysing highly detailed individual-level data from a hospital in Boston.

213 Refracted and reflected Levy processes and de Finetti's control problem.

[IS 8,(page 50)]

Andreas KYPRIANOU, *Department of Mathematical Sciences, University of Bath*

I will give a review of a collection of results spanning four different papers concerning mathematical aspects of de Finetti's actuarial control problem when the underlying source of randomness is a spectrally negative Lévy process. As well as considering solutions to the latter, we shall examine issues concerning associated fluctuation identities of refracted and reflected Lévy processes as well as presenting new results on the theory of scale functions necessitated by the de Finetti problem.

214 Functional sparsity

[IS 23,(page 35)]

John LAFFERTY, *Carnegie Mellon University*

Han LIU, *Carnegie Mellon University*

Pradeep RAVIKUMAR, *University of California, Berkeley*

Larry WASSERMAN, *Carnegie Mellon University*

Substantial progress has recently been made on understanding the behavior of sparse linear models in the high-dimensional setting, where the number the variables can greatly exceed the number of samples. This problem has attracted the interest of multiple communities, including applied mathematics, signal processing, statistics and machine learning. But linear models often rely on unrealistically strong assumptions, made mainly for convenience. Going beyond parametric models, can we understand the properties of high-dimensional functions that enable them to be estimated accurately from sparse data? We present some progress on this problem, showing that many of the recent results for sparse linear models can be extended to the infinite-dimensional setting of nonparametric function estimation. In particular, we present some theory for estimating sparse additive models, together with algorithms that are scalable to high dimensions. We illustrate these ideas with an application to functional sparse coding of natural images.

215 Empirical Bayes for ARCH models

[CS 31,(page 28)]

Fazlollah LAK, *Persian Gulf University, Bushehr*

Abstract: One important input in the derivation of the empirical Bayes estimator is the estimation of hyperparameters from the marginal distribution. In this paper, the SAME method is used to estimate these hyperparameters. This algorithm has been introduced by Doucet et al. (2002) to obtain the maximum of marginal posterior distribution and it is used in hidden Markov models by Doucet and Robert (2000). The basic idea of this algorithm is a form of simulated annealing. Simulated annealing considers simulating from increasing powers of $m(y|\theta_1)$. The motivation is that m^γ gets more and more concentrated on the set of maxima of m when γ increases. Simulation results show that the Bayes estimator, under a noninformative prior distribution, dominates the maximum likelihood estimator with respect to MSE criteria. As in other settings, hierarchical Bayes models come as a natural competitor model for the empirical Bayes method. A one stage Bayesian hierarchical analysis has been undertaken for the same data set. The results compared to empirical Bayes. The results are mixed, some times empirical Bayes outperform, the hierarchical Bayes and vice versa.

Key words and phrases: ARCH process, Bayesian inference, Gibbs sampler, Markov chain Monte Carlo, Metropolis-Hasting algorithm, SAME method.

216 Projection properties of non regular fractional factorial designs

[CS 43,(page 38)]

M.H. LAKHO, *University of Peshawar*
Harvey Xianggui QU, *Oakland University Michigan USA*
Muhammad Fazli QADIR, *University of peshawar*

(M,S)- Optimality criterion is used for classification of nonregular factorial designs. Some properties of trace ( ) and trace ( ) were used in selecting and projecting nonregular designs. (M,S) criterion proposed is easier to compute and it is also independent of the choice of orthonormal contrasts. It can be applied to two-level as well as multi-level symmetrical and asymmetrical designs. The criterion is also applied to study the projective properties of nonregular designs when only 24 runs are used for 23 factors. It is applied to examine designs constructed from 60 Hadamard matrices of order 24 and obtain lists of designs that attain the maximum or high Estimation Capacity (EC) for various dimensions. Keywords: Fractional factorial designs, Minimum aberration, (M,S)-optimality, Hadamard matrices.

217 Sequential designs in computer experiments for response surface model fit

[CS 43,(page 38)]

Chen Quin LAM, *The Ohio State University*
William NOTZ, *The Ohio State University*

Computer simulations have become increasingly popular for representing the complex physical reality in a set of computer codes. Due to limited computing capabilities to carry out these simulations at very fine grids, computer experiments have often been performed to approximate the unknown response surface given sparse observations. While space-filling designs, such as maximin distance and Latin hypercube designs, are useful for initial exploratory purposes, they do not allow the addition of designs points iteratively while still maintaining the space-filling property. Other complex designs based on certain optimality criteria and cross validation approaches have been proposed and will be reviewed.

In this talk, we will present some new criteria for the cross validation approach and a modified expected improvement criterion, which is originally proposed for global optimization, for global fit of response surfaces. Results from empirical studies reveal that sequential (adaptive) designs are potentially superior. While many sequential designs have been proposed, it is not clear how the performance of these methods might be affected by the type of response surface, choice of the correlation function, size of initial starting designs etc. We will address some of these issues in our discussion.

218 Estimation of large covariance matrices through block penalization

[CS 26,(page 25)]

Clifford LAM, *Princeton University*
Jianqing FAN, *Princeton University*

Estimation of covariance or inverse covariance (precision) matrix has always been a fundamental problem in statistical inferences and multivariate analyses, from risk management and portfolio allocation in finance, to input in various statistical procedures like the Principal Component Analysis (PCA) or the Linear/Quadratic Discriminant Analysis (LDA/QDA) in classification, among many others. Also, with the advance of technologies, data with number of variables comparable to or even larger than the sample size is common nowadays. In this talk, we focus on the estimation of inverse covariance matrices for data with a natural ordering or a notion of distance between variables. Through the introduction of a block penalty on the blocks of off-diagonals of the modified Cholesky factor for the covariance matrix of interest, we show a notion of sign consistency for the resulting estimator when number of variables is much larger than the sample size, where blocks of zero off-diagonals are estimated to be zero, and blocks of non-zero off-diagonals are estimated to be non-zero, all with probability goes to one. We also prove a rate of convergence in the matrix operator norm, which links to the non-sparsity rate of the true inverse covariance matrix, for the resulting estimator under certain tail assumptions of the variables. Hence the data is not required to be normally distributed for the results to hold. Simulations and real data analysis demonstrate the effectiveness and flexibility of the method over established tools such as the LASSO and banding.

219 Statistical significance of probability models in volatility forecasting

[CS 8,(page 11)]

K.P. LAM, *Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong*

Probability models have played a fundamental role in forecasting conditional volatility and variance from past historical data. In the celebrated GARCH and ARCH formulations, an additive error model using an independent identical distribution (iid) process with a normal probability density function (pdf) is often assumed. Recent research in high frequency finance led successfully to the development of Autoregressive Conditional Duration (ACD), which uses a multiplicative error model (MEM) for a non-negative stochastic iid process. Several probability density functions for non-negative variable including the Weibull, gamma, exponential, and half-gaussian distributions, among others, have been proposed. A comparative study of using these probability models for MEM, especially for their merits under different specifications, is necessary.

Using simulated data generated from specified Weibull, gamma, and exponential distributions, verification tests on the respective MEM models were performed. The estimated GARCH parameters are compared with respect to statistical measures and mean-square errors from the specifications. As expected from using a maximum likelihood procedure, more accurate parameter estimates are obtained if prior knowledge of a true probability model is used. For real market data where the probability model is unknown, we study two practical cases using 10-minutes intra-daily NASDAQ and Standard and Poor 500 indexes for conditional volatility prediction. Using the realized volatility derived as proxy for the underlying latent process, we compare the different MEM models for Weibull, gamma, exponential, and half-gaussian pdfs. Statistical significance is evaluated using R2 of Mincer-Zarnowitz regression and t-statistics of Diebold-Mariano-West test.

220 The needlets bispectrum

[CS 72,(page 53)]

Xiaohong LAN, *Institute of Mathematics, Chinese Academy of Sciences and Department of Mathematics, University of Rome Tor Vergata*
Domenico MARINUCCI, *Department of Mathematics, University of Rome Tor Vergata*

The purpose of this paper is to join two different threads of the recent literature on random fields on the sphere, namely the statistical analysis of higher order angular power spectra on one hand, and the construction of second-generation wavelets on the sphere on the other. To this aim, we introduce the needlets bispectrum and we derive a number of convergence results. Here, the limit theory is developed in the high resolution sense. The leading motivation of these results is the need for statistical procedures for searching non-Gaussianity in Cosmic Microwave

Background radiation. In recent years, this issue has drawn an enormous amount of attention in the physical literature.

References

1. Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2006), Asymptotics for Spherical Needlets, Annals of Statistics, in press, arxiv:math/0606599
2. Dodelson, S. (2003), Modern Cosmology, Academic Press
3. Lan, X. and Marinucci, D. (2008), The Needlets Bispectrum, submitted for publication, arxiv:math.st/0802.4020
4. Marinucci, D. (2006), High-Resolution Asymptotics for the Angular Bispectrum of Spherical Random Fields, Annals of Statistics 34, 1–41, arxiv:math/0502434
5. Narcowich, F.J., Petrushev, P. and Ward, J.D. (2006) Localized Tight Frames on Spheres, SIAM Journal of Mathematical Analysis 38, 2, 574–594

221 Forgetting of the initial law in non-ergodic case

[CS 58,(page 47)]

B. LANDELLE, *Université Paris-Sud, Orsay, France; Thales Optronique, Élancourt, France*

E. MOULINES, *École Nationale Supérieure des Télécommunications, Paris, France*

E. GASSIAT, *Université Paris-Sud, Orsay, France*

A Hidden Markov Model (HMM) is a doubly stochastic process $(X_k, Y_k)_{k>0}$ with an underlying Markov chain that is not directly observable. The distribution of the current hidden state X_n conditionally on the observations $Y_1, ...Y_n$ is called the optimal filter. A problem of interest is to understand whether two filters are close, in some sense, for large values of n and two different choices of the initial distribution. This problem is known as the forgetting property of the optimal filter.

The case of ergodic HMM has been extensively studied and is now well-understood [1], [2]. Our contribution [3] focuses on discrete non-ergodic Hidden Markov Models: a new set of conditions is proposed to establish the forgetting property of the filter at a geometric rate. Both a pathwise-type convergence of the total variation distance of the filter started from two different initial distributions, and a convergence in expectation are considered. The results are shown to hold under rather weak conditions on the observation process $(Y_k)_{k>0}$ which do not necessarily entail that the observations are from an HMM. The results are illustrated using generic models of non-ergodic HMM and extend all the results known so far [4], [5].

References

1. R. Douc, G. Fort, E. Moulines, P. Priouret, Forgetting of the initial distribution for Hidden Markov Models, submitted.
2. M.L. Kleptsyna, A.Y. Veretennikov, On discrete time ergodic filters with wrong initial conditions, C. R. Acad. Sci. Paris, Ser. I 344, 2007.
3. B. Landelle, E. Moulines, E. Gassiat, Forgetting of the initial law in non-ergodic case, submitted.
4. A. Budhiraja, D. Ocone, Exponential Stability in Discrete Time Filtering for Non-Ergodic Signals, Stoch. Proc. Appl., 82(2):245–257, 1999.
5. Nadia Oudjane, Sylvain Rubenthaler, Stability and Uniform Particle Approximation of Nonlinear Filters in Case of Non Ergodic Signals, Stoch. Anal. Appl., 23(3):421–448, 2005.

222 Parameters of simple stochastic neuronal models

[IS 32,(page 36)]

Petr LANSKY, *Institute of Physiology, Academy of Sciences of the Czech Republic*

Stochastic approach to the problems of computational neuroscience is common due to the apparent randomness of neuronal behavior. Therefore many stochastic models of neurons have been proposed and deeply studied. They range from simple statistical descriptors to sophisticated and realistic biophysical models. On their basis, properties of neuronal information transfer are deduced. The present talk aims to contribute to this effort.

The basic assumptions made on the spiking activity permit us to consider spike trains as realizations of a stochastic point process. Then, it is required to characterize it in a most compact way. Thus parameters characterizing firing rate or firing variability are calculated. Further, having the experimental data, the spike trains or membrane depolarization trajectories, we may ask what was the signal stimulating the neuron producing this sequence of action potentials. In this way, the parameters of the models have to be determined. The recent results achieved in both these directions and extending our previous effort [1-7] are summarized.

References

1. Greenwood, P.E.and Lansky, P. (2007) Information content in threshold data with non-Gaussian

noiseOptimal signal estimation in neuronal models. Fluctuation and Noise Letters, 7: 79-89.

2. Hampel, D. and Lansky, P.(2008) On the estimation of the refractory period. J. Neurosci. Meth. (in press).
3. Kostal, L., Lansky, P. and Rospars, J.P. (2007) Neuronal coding and spiking randomness European J. Neurosci. 26: 2693-2701, 2007.
4. Lansky, P. (2005) Greenwood P.E., Optimal signal estimation in neuronal models. Neural Computation 17: 2240-2257.
5. Lansky, P. (2004) Rodriguez R., Sacerdote L., Mean instantaneous firing frequency is always higher than the firing rate. Neural Comput. 16: 477-489.
6. Lansky, P., Sanda, P. and He J.F. (2006) The parameters of the stochastic leaky integrate-and-fire neuronal model. J. Comput. Neurosci. 21: 211-223.
7. Pokora, O. and Lansky, P.(2008) Statistical approach in search for optimal signal in simple olfactory neuronal models, Math. Biosci. (in press).

223 Inference for partially observed multitype branching processes and Ecological applications

[CS 66,(page 49)]

Catherine LARÉDO, *Mathématiques et Informatique Appliquées, INRA, Jouy-en-Josas, FRANCE et LPMA, Universités Paris 6 et Paris 7, Paris*

Olivier DAVID, *MIA, INRA, Jouy-en-Josas, FRANCE*

Aurélie GARNIER, *Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, FRANCE*

In Ecology, demographic matrix models (Caswell, 2001) are widely used to analyse the population dynamics of age or stage-structured populations. These models are mainly deterministic, with noise added to take into account some variability: they cannot describe the stochasticity originating from the dynamics. Here, we use multitype branching processes with immigration in one type to model these dynamics. It corresponds to a description of annual plants, where mature plants die just after seed production, and where immigration of new seeds occurs because of spillage during seed transport (Garnier et al., 2006). Our aim is to estimate the individual demographic parameters underlying the population dynamics from data consisting in the populations counts in each stage. A central theme in Ecology is that these parameters may not be fully identifiable knowing only the population dynamics. This also occurs in parametric inference for multitype branching processes (Maaouia and Touati, 2005). In practice, Bayesian methods and E-M type algorithms are used to circumvent this problem but cannot address it. We first study the parametric inference in this context when all the types are observed. We prove identifiability and build consistent and asymptotically Gaussian estimators of all the parameters.

However, for many ecological data, some types cannot be observed in practice; this often holds for the seeds. Thus, we address the problem of unobserved types. This results in an "Incomplete Data model" (Cappé, Moulines, Ryden, 2005) where estimation is linked to the statistical inference for state-space models (Genon-Catalot and Larédo, 2006). The process associated to these observations is no longer Markovian. We obtain that the model with Poisson distributions provides a useful example with explicit computations. In that case, we characterize the dependence of present observations with respect to past ones. We also prove that identifiability holds only for a subset of the parameters and derive their asymptotic properties. Finally, we study numerically the behaviour of the estimation procedures when the model is not Poisson. We obtain satisfying results for a large class of models with distributions whose mean and variance are of the same magnitude.

224 Graphical models with edge and vertex symmetries

[CS 54,(page 44)]

Steffen LAURITZEN, *University of Oxford, United Kingdom*

Søren HØJSGAARD, *Aarhus University, Denmark*

Graphical Gaussian models with symmetry restrictions on the concentration or correlation matrix are studied. The models can be represented by coloured graphs, where parameters associated with edges or vertices of same colour are restricted to being identical. The properties of such models are studied, including conditions for restrictions on the concentration and correlation matrices being equivalent. This is for example the case when symmetries are generated by permutation of variable labels. More detailed results are derived in the latter case and the types of models are classified.

225 Pseudo-likelihood estimation for non-hereditary Gibbs point processes

[CS 2,(page 6)]

Frédéric LAVANCIER, *Laboratoire Jean Leray, Uni-*

versité de Nantes, France
David DEREUDRE, *LAMAV, Université de Valenciennes et du Hainaut-Cambrésis, France.*

The aim of this talk is to present a parametric estimation of the potential interaction in non-hereditary Gibbs point processes.

An interaction is hereditary if for every forbidden point configuration γ in R^d then, for every x in R^d, the configuration $\gamma+\delta_x$ remains forbidden. In the domain of stochastic geometry, it seems natural to meet some non-hereditary interaction : as an example, a model of geometric objects submitted to a hardcore interaction will be presented.

The main problem, in the non-hereditary case, is that the energy of a point x in a configuration γ is not always defined. We then introduce the set $\mathcal{R}(\gamma)$ of removable points in γ : x in γ is said to be removable if the energy of $\gamma - \delta_x$ is locally finite. This concept allows us to extend the definition of the pseudo-likelihood contrast function to the non-hereditary case. Denoting h the local energy of x in γ, we take

$$\begin{aligned} PLL_{\Lambda_n}(\gamma,\alpha,\theta) = \frac{1}{|\Lambda_n|}\Big[\int_{\Lambda_n} exp\left(-h^{\alpha,\theta}(x,\gamma)\right) dx \\ + \sum_{x\in\mathcal{R}^{\alpha}(\gamma)\cap\Lambda_n} h^{\alpha,\theta}(x,\gamma-\delta_x)\Big], \end{aligned}$$

where α is the hardcore parameter, θ parameterizes the classical smooth interaction and where Λ_n denotes the domain of observation.

We first estimate the hardcore parameter α by choosing the one associated to the smaller support containing the realisation γ. Then, we estimate θ by maximizing $PLL_{\Lambda_n}(\gamma,\hat{\alpha},\theta)$ where $\hat{\alpha}$ is the estimator of α.

To prove the consistency of our estimators, we need to extend the Campbell equilibrium equation, initially proposed by Nguyen and Zessin, to the non-hereditary setting : let μ be a Gibbs measure and $\mathcal{C}^!_\mu$ its Campbell measure, we prove that

$$\mathbb{1}_{x\in\mathcal{R}(\gamma+\delta_x)}\mathcal{C}^!_\mu(dx,d\gamma) = e^{-h(x,\gamma)}\lambda\otimes\mu(dx,d\gamma),$$

where λ is the Lebesgue measure on R^d.

We establish the consistency of both estimators of α and θ resulting from this procedure.

226 Markov chain Monte Carlo simulations of biomolecular reactions with randomly fluctuating kinetics

[CS 62,(page 48)]

Chia Ying LEE, *Division of Applied Mathematics, Brown University*
Jiawei CHIU, *A*STAR Institute of High Performance Computing, Singapore*
K.-H. CHIAM, *A*STAR Institute of High Performance Computing, Singapore*

Traditional models of chemical reactions model the evolution of the chemical system by assuming a fixed occurrence rate of a reaction, namely the rate constant. This model works adequately well for large chemical systems, due to averaging effects. More recent developments suggest that on a molecular level, the rate constant is seldom constant, but instead exhibits random fluctuations over time, due to environmental influences, conformational changes of the molecule, etc. Hence, in small systems with low copy numbers of reactants, such as in the intracellular environment, the fluctuating rate constants produce significant deviation from predictions of the traditional model. The original Gillespies Stochastic Simulation Algorithm [1] was designed as a way to simulate trajectories of the evolution of small chemical systems, but is based on the assumption that the rate constant is constant.

In light of the recent findings, we propose a modification of the algorithm to allow for the simulation of systems whose randomly fluctuating rate constant comes from a given distribution. Constraints on the inter-reaction fluctuations of the rate constant, such as slowly inter-converting conformers, are implemented by Markov Chain Monte Carlo (MCMC) techniques to preserve the rate constants distribution. In particular, we apply the modified algorithm and MCMC model to the classical Michaelis-Menten reaction to yield numerical results that corroborate experimental results obtained on single molecule enzyme kinetics [2]. This suggests the validity of our algorithm, and allows us to discuss scenarios in intracellular signaling where our algorithm may be applicable.

References

1. Gillespie, D. T.(1977) J. Phys. Chem. 81, 2340.
2. English, B.P. et al. (2006) Nature Chem. Biol. 2, 87.

227 Nonparametric tests for distributional treatment effect for randomly censored responses

[CS 18,(page 20)]

Myoung-jae LEE, *Korea University*

With treatment and control groups available, Koul and Schick (1997) provided a nonparametric test for no distributional treatment effect when the response variable is fully observed and the distribution of covariate X is the same across the two groups. This test is, however, not applicable to censored responses, nor to non-experimental (i.e., observational) studies that entail different distributions of X across the two groups. In this paper, we propose 'X-matched' nonparametric tests generalizing the test of Koul and Schick (1997) following the idea in Gehan (1965). Our tests are applicable to non-experimental data with randomly censored responses. In addition to these motivations, the tests have a number of advantages. First, they have the intuitive appeal of comparing all available pairs across the treatment and control groups, instead of selecting a number of matched controls/treated in the usual pair or multiple matching. Second, whereas most matching estimators/tests have a non-overlapping support (of X) problem across the two groups, our tests have a built-in protection against the problem. Third, Gehan's (1965) idea allows the tests to make a good use of censored observations. A simulation study is conducted, and an empirical illustration for a job-training effect on unemployment duration is provided.

228 A Bahadur representation of the linear support vector machine

[CS 70,(page 53)]

Yoonkyung LEE, *Department of Statistics, The Ohio State University, U.S.A.*
Ja-Yong KOO, *Department of Statistics, Korea University, Korea*
Changyi PARK, *Department of Statistics, University of Seoul, Korea*
Yuwon KIM, *Data Mining Team, NHN Inc., Korea*

The support vector machine has been successful in a variety of applications. Also on the theoretical front, statistical properties of the support vector machine have been studied quite extensively with a particular attention to its Bayes risk consistency under some conditions. In this talk, basic statistical properties of the support vector machine are investigated, namely the asymptotic behavior of the coefficients of the linear support vector machine. A Bahadur type representation of the coefficients is established under appropriate conditions, and their asymptotic normality and statistical variability are derived on the basis of the representation. These asymptotic results do not only help further our understanding of the support vector machine, but also they can be useful for related statistical inferences.

229 On general adaptive sparse principal component analysis

[CS 26,(page 25)]

Chenlei LENG, *National University of Singapore*
Hansheng WANG, *Peking University*

The method of sparse principal component analysis (S-PCA) proposed by Zou, Hastie and Tibshirani (2006) is an attractive approach to obtain sparse loadings in principal component analysis (PCA). S-PCA was motivated by reformulating PCA as a least squares problem so that a lasso penalty on the loading coefficients can be applied. In this article, we propose new estimates to improve S-PCA on the following two aspects. Firstly, we propose a method of simple adaptive sparse principal component analysis (SAS-PCA), which uses the adaptive lasso penalty instead of the lasso penalty in S-PCA. Secondly, we replace the least squares objective function in S-PCA by a general least squares objective function. This formulation allows us to study many related sparse PCA estimators under a unified theoretical framework and leads to the method of general adaptive sparse principal component analysis (GAS-PCA). Compared with SAS-PCA, GAS-PCA enjoys much further improved finite sample performance. In addition to that, we show that when a BIC-type criterion is used for selecting the tuning parameters, the resulting estimates are consistent in variable selection. Numerical studies are conducted to compare the finite sample performance of various competing methods.

230 Wavelet covariance and correlation properties of spurious and cointegrated regressions: a Monte Carlo study

[CS 67,(page 49)]

Chee Kian LEONG, *Nanyang Technological University*

This paper examines the wavelet covariance and correlation properties of spurious and cointegrated bivariate time series using Monte Carlo simulations. In the case of the spurious regression, the null hypotheses of zero wavelet covariance and correlation for these series across the scales fail to be rejected. Conversely, these null hypotheses across the scales

are rejected for the cointegrated bivariate time series. These non residual based tests are then applied to analyze if any relationship exists between the extraterrestial phenomenon of sunspots and the earthly economic time series of oil prices. Conventional residual based tests appear sensitive to the specifications in both the cointegrating regression and the lag order in the Augmented Dickey-Fuller tests on the residuals. In contrast, the wavelet tests, with bootstrap t-statistics and confidence intervals, dispel any doubts about the spuriousness of this relationship.

231 A nonparametric test for convexity of regression functions against local concavity.

[CS 60,(page 47)]

Samantha LEORATO, *University of Rome Tor Vergata*

We consider the nonparametric regression model $Y = \mu(X) + \varepsilon$. We propose a new nonparametric test for the hypothesis of convexity of the regression function $\mu(x)$. The methodology relies on a characterization of convex functions based on a refined version of Jensen ineqality. We define a sequential procedure that consists in minimizing differences of conditional expectations over a particular class of convex sets. The test statistic we propose is a Kolmogorov type test and is suited to test for convexity of μ against alternatives of local concavity of the regression function around some fixed point x_0. We don't need to impose restrictions on the derivatives of the conditional mean function (cmf), thus allowing it to be non-differentiable. Consistency and asymptotic normality of the test statistic (under the null) are proved. A simulation study concludes the paper.

232 Stochastic relations of random variables and processes

[CS 58,(page 46)]

Lasse LESKELÄ, *Helsinki University of Technology, Department of Mathematics and Systems Analysis, PO Box 1100, 02015 TKK, Finland*

Stochastic monotonicity is a key methodology for approximating random variables and processes whose distributions are hard to compute explicitly [1,2,3]. For example, the stochastic ordering of two stationary random dynamical systems can be established without the explicit knowledge of the stationary distributions, by verifying that the generators of the systems preserve the order [4,5]. In this talk I will sketch a new definition of a stochastic relation, which extends the notion of stochastic order to a relation between probability measures over arbitrary measurable spaces. This definition is motivated by the observation that for the stochastic ordering of two stationary random dynamical systems, it suffices to show that the generators of the systems preserve some, not necessarily reflexive or transitive, subrelation of the order. The main contributions of the talk are: (i) functional and coupling characterization of stochastic relations based on Strassen's coupling theorem [6], (ii) necessary and sufficient conditions for a pair of probability kernels to preserve a stochastic relation, and (iii) an iterative algorithm for computing the maximal subrelation of a given relation that is preserved by a pair of probability kernels. The theory is illustrated with applications to multidimensional random walks, population processes, and queueing systems [7]. The main part of the work presented has been carried out at Centrum voor Wiskunde en Informatica and Eindhoven University of Technology, the Netherlands.

References

1. Kamae, T., Krengel, U. and O'Brien, G. (1977) Stochastic inequalities on partially ordered spaces, Ann. Probab. 5, 899-912.
2. Müller, A. and Stoyan, D. (2002) Comparison Methods for Stochastic Models and Risks, Wiley.
3. Shaked, M. and Shanthikumar, J. G. (2006) Stochastic Orders, Springer.
4. Whitt, W. (1986) Stochastic comparisons for non-Markov processes, Math. Oper. Res. 11, 608-618.
5. Massey, W. A. (1987) Stochastic orderings for Markov processes on partially ordered spaces, Math. Oper. Res. 12, 350-367.
6. Strassen, V. (1965) The existence of probability measures with given marginals, Ann. Math. Statist. 36, 423-439.
7. Jonckheere, M., and Leskelä, L. (2007) Stochastic bounds for two-layer loss systems, Preprint: arXiv:0708.1927.

233 A note on the existence and uniqueness of a bounded mean-reverting process

[CS 20,(page 21)]

Dharma LESMONO, *Department of Mathematics Parahyangan Catholic University, Bandung INDONESIA*

Phil POLLETT, *Department of Mathematics The University of Queensland AUSTRALIA*
Elliot TONKES, *Energy Edge Pty Ltd., Brisbane, AUSTRALIA*
Kevin BURRAGE, *Institute for Molecular Bioscience, The University of Queensland, AUSTRALIA*

We study a stochastic differential equation (SDE) a class of mean-reverting diffusions on a bounded interval. The drift coefficient is not continuous near the boundaries. Nor does it satisfy either of the usual Lipschitz or linear growth conditions. We characterize the boundary behaviour, identifying two possibilities: entrance boundary and regular boundary. In the case of an entrance boundary we establish existence and uniqueness of the solution to the SDE.

234 Weighted area under the receiver operating characteristic curve and its application to gene selection

[CS 69,(page 52)]

Jialiang LI, *National University of Singapore*
Jason FINE, *Department of Biostatistics, University of North Carolina at Chapel Hill*

Partial area under the ROC curve (PAUC) has been proposed for gene selection in Pepe et al. (2003) and thereafter applied in real data analysis. It was noticed from empirical studies that this measure has several key weaknesses, such as an inability to reflect nonuniform weighting of different decision thresholds, resulting in large numbers of ties. We propose the weighted area under the ROC curve (WAUC) in this paper to address the problems associated with PAUC. Our proposed measure enjoys a greater flexibility to describe the discrimination accuracy of genes. Nonparametric and parametric estimation methods are introduced, including PAUC as a special case, along with theoretical properties of the estimators. We also provide a simple variance formula, yielding a novel variance estimator for nonparametric estimation of PAUC, which has proven challenging in previous work. Simulations and re-analysis of two well-known microarray datasets illustrate the practical utility of WAUC.

235 Bayesian inference for normal tempered stable stochastic volatility models

[CS 24,(page 24)]

Junye LI, *PhD Program in Economics Bocconi University*

This paper introduces Normal Tempered Stable Lévy process, which is constructed via subordinating Brownian motion with drift using Tempered Stable subordinator. The tempered stable distribution is obtained by exponentially tilting the positive α-stable distribution and formally introduced in Hougaard (1986) in survival analysis. The subordinate Normal Tempered Stable process is an infinite activity Lévy process which can be either of bounded variation or of infinite variation depending on different values of its stable index. Two of special cases are Variance Gamma process (Madan et al., 1998) and Normal Inverse Gaussian process (Barndorff-Nielsen, 1998). Stock price dynamics is then modeled through taking into account both jump and stochastic volatility. Jump is modeled with normal tempered stable process and stochastic volatility is realized by adopting time-changing approach. I propose Bayesian Markov Chain Monte Carlo methods for inferences of the models. The Brownian subordination property provides us extra convenience to apply Bayesian methods. Since the probability density function of the tempered stable distribution is known only in series representation whereas its Characteristic function is in simple form, I adopt Saddlepoint approach to approximate the density function. I investigate both single stochastic volatility model where the stochastic volatility is only from time-changing the diffusion part and double stochastic volatility model in which the stochastic volatility is from time-changing both the diffusion and jump parts. The empirical study indicates that (i) the normal tempered stable process takes on finite variation in stock price dynamic modeling; (ii) it is very flexible and can capture most of characteristics of asset returns; and (iii) the double stochastic volatility model only provides a marginal modeling improvement.

236 Semiparametric estimation for a class of time-inhomogeneous diffusion processes

[CS 2,(page 7)]

Min LI, *California State University, Sacramento*
Yan YU, *University of Cincinnati*
Keming YU, *CARISMA, Brunel University,UK*
Hua WANG, *Yahoo Inc.*

We develop two likelihood based approaches to semiparametrically estimate a class of time-

inhomogeneous diffusion process: log penalized splines (P-splines) and the local log-linear method. Positive volatility is naturally embedded and this positivity is not guaranteed in most existing diffusion models. We investigate different smoothing parameter selection. Separate bandwidths are used for drift and volatility estimation. In the log P-splines approach, different smoothness for different time varying coefficients is feasible by assigning different penalty parameters. We also provide accompanying theorems for both approaches and report statistical inference results. Finally, we present a case study using the weekly three-month Treasury bill data from 1954 to 2004. We find that the log P-splines approach seems to capture the volatility dip in mid-1960s the best. We also present an application of calculating a financial market risk measure called Value at Risk (VaR) using statistical estimates from log P-splines.

237 A nested mixture model for protein identification using mass spectrometry

[CS 9,(page 11)]

Qunhua LI, *Dept of Statistics Univ of Washington*
Michael MACCOSS, *Dept of Genome Sciences*
Matthew STEPHENS, *Dept of Statistics and Human Genetics*

Protein identification using mass spectrometry is a high-throughput way to identify proteins in biological samples. The identification procedure consists of two steps. It first identifies the peptides from mass spectra, then determines if proteins assembled from the putative peptide identifications, many of which will be incorrect, are present in the samples. In this study, we develop an unsupervised protein identification approach based on a nested mixture model, which assesses the evidence for presence of proteins constructed from putative peptide identificaitons. Our approach is essentially a model-based clustering method, which simultaneously identifies which proteins are present and which peptides are correctly identified. In this fashion, our model incorporates the evidence feedback between proteins and their substring peptides. Using a yeast dataset, we show that our method has a competitive performance on protein identification and a superior performance on peptide validation over leading products.

238 Bayesian nonlinear principal component analysis using random fields

[PS 3,(page 29)]

Heng LIAN, *Nanyang Technological University*

Principal component analysis (PCA) is an old statistical technique for unsupervised dimension reduction. It is often used for exploratory data analysis with the objective of understanding the structure of the data. PCA aims to represent the high dimensional data points with low dimensional representers commonly called latent variables, which can be used for visualization, data compression etc.

PCA is a linear procedure since the reconstruction is based on a linear combination of the principal components. We propose a novel Bayesian approach to nonlinear PCA. The model is based on the probabilistic formulation of PCA, but with two differences. First, the linear transformation is defined on the stiefel manifold (the manifold of orthnormal matrices). Second, the linear transformation in our model is dependent on the corresponding latent variable. The linear transformations in different parts of the latent space are related by putting a Markov random field prior over the space of orthonormal matrices which makes the model identifiable. The model is estimated by Gibbs sampling which explores the posterior distribution of both the latent space and the transformation space. The computational burden for each iteration of Gibbs sampling is square in the number of data points.

239 Semifolding in the 2^{n-p} plan

[CS 43,(page 38)]

Pen-Hwang LIAU, *Professor; Dept. of Mathematics, National Kaohsiung Normal University, Taiwan, R.O.C.*
Pi-Hsiang HUANG, *Dept. of Mathematics, National Kaohsiung Normal University, Taiwan, R.O.C.*
Chien-Shiang LIAU, *Dept. of Mathematics, National Kaohsiung Normal University, Taiwan, R.O.C.*
Chien-Te LEE, *Dept. of Mathematics, National Kaohsiung Normal University, Taiwan, R.O.C.*

According to the initial analysis or prior information, some factors may be more important than others. In this situation, one may apply the technique of fold-over to isolate these main factors and each of their two-factor interactions in the follow-up experiment. However, the technique requires the same size as the original experiment. This can sometimes be quite wasteful. In fact, the technique of semifolding may be considered by the experiment to the next plan to isolate these factors. In this article, the criterion of clear effect may be applied to perform the experiment. Furthermore, the optimal semifolding design is also considered for a 2^{n-p} plan and we will use the computer to search the corresponding optimal semifolding design for the given 2^{n-p} designs that are tabulated in Chen, Sun, and Wu's (1993) paper.

240 On means of random probability measures

[CS 19,(page 20)]

Antonio LIJOI, *University of Pavia (Italy)*
Lancelot F. JAMES, *Hong Kong University of Science and Technology*
Igor PRUENSTER, *University of Turin (Italy)*

Linear functionals of random probability measures are of interest in a variety of research areas ranging from Bayesian nonparametric statistics to the theory of random processes. While many results are known by now for random Dirichlet means, little is known for means of random probabilities different from the Dirichlet process. The present talk illustrates a few recent results which lead to exact expressions of the probability distribution of a linear functional of the two parameter Poisson-Dirichlet process. Moreover, attention will be focussed on an application to Bayesian statistics where, under the usual assumption of exchangeability of the sequence of obsevations, a normalized completely random measure P is used to define a nonparametric prior. In this setting, we will derive a representation of the posterior distribution of a mean of P.

241 Analogue to Poisson's equation for a jump process

[CS 32,(page 31)]

Adrian LIM, *Vanderbilt University*

Suppose we have a bounded domain $D \subseteq R^d$. Given the Laplacian Δ on R^d and a compactly supported continuous function f in D, Poisson's Equation involves solving

$$\Delta u = -f,$$

with $u \equiv 0$ at the boundary. If $f \in C^{k+\beta}$, then the unique solution is given by $E_x \int_0^{\tau_D} f(B_t)dt$ and the solution is $C^{k+2+\beta}$. B_t is a d-dimensional Brownian motion.

We can also replace Δ by an elliptic operator in non-divergence form in the Poisson's Equation and obtain a similar expression, in terms of a process X_t associated to the elliptic operator. Again, if $f \in C^{k+\beta}$ and the coefficients of the elliptic operator are $C^{k+\beta}$, then $E_x \int_0^{\tau_D} f(X_t)dt$ solves Poisson's Equation on a domain D and is $C^{k+2+\beta}$.

It would be nice to consider an analogue of this problem for pure jump process. Consider a strong Markov process taking values in the space of right continuous with left limits paths in R^d, generated by the following integral operator, acting on C_b^2 functions,

$$Lf(x) = \int \left[f(x+h) - f(x) - 1_{(|h|\leq M)} \nabla f(x) \cdot h\right] \frac{A(x,h)}{|h|^{d+\alpha}} dh.$$

Here, $0 < \alpha < 2$, and M is any fixed constant, $0 < c_l \leq A(x,h) \leq c_u$ for any $x, h \in R^d$. It would seem that $E_x \int_0^{\tau_D} f(X_t)dt$ will solve Poisson's Equation on a bounded domain D using the integral operator L, vanishing outside D.

Unfortunately, such a solution does not exist. Hence we modify the problem to, given $f \in C^{k+\beta}$ with all its derivatives up to the kth order in L^2, consider

$$Lu = -f, \lim_{|x|\to\infty} u(x) = 0.$$

The main result is that under some assumptions on $A(\cdot,\cdot)$, including $C^{k+\beta}$ on its first variable and some continuity on its second variable, there is a unique solution, and this solution is $C^{k+2+\epsilon}$, where $\epsilon = \alpha+\beta-2$ if $\alpha > 1$ and $\epsilon = \alpha + \beta - 1$ if $\alpha \leq 1$.

242 Optimal designs for estimating control values

[CS 22,(page 22)]

Chun-Sui LIN, *Cheng Shiu University, Taiwan, R.O.C.*
Mong-Na Lo HUANG, *National Sun Yat-sen University, Taiwan, R.O.C.*

In this talk, optimal design problems for estimating control values are discussed. Based on a multiresponse regression model, this is of interest to obtain a suitable estimation of the corresponding control value for a given target on the expected responses. Note that there is a subtle difference between estimating the optimal control value corresponding to a given target on the responses and the usual calibration problem of estimating the true control value corresponding to measured responses. The objective of this study is to find an optimal design for estimating the adequate control value, which minimizes the mean squared error of the estimator of the control value.

243 Prediction intervals for general balanced linear random models

[PS 1,(page 4)]

Tsai-Yu LIN, *Department of Applied Mathematics, Feng Chia University, Taichung 40724, Taiwan*
Chen-Tuo LIAO, *Division of Biometry, Institute of Agronomy, National Taiwan University, Taipei 10617, Taiwan*

The main interest of prediction intervals lies in the results of a future sample from a previously sampled population. In this article, we develop procedures for the prediction intervals which contain all of a fixed number of future observations for general balanced linear random models. Two methods based on the concept of a generalized pivotal quantity (GPQ) and one based on ANOVA estimators are presented. A simulation study using the balanced one-way random model is conducted to evaluate the proposed methods. It is shown that one of the two GPQ-based and the ANOVA-based methods are computationally more efficient and they also successfully maintain the simulated coverage probabilities close to the nominal confidence level. Hence, they are recommended for practical use. In addition, one example is given to illustrate the applicability of the recommended methods.

244 On maxima of periodograms of stationary processes

[CS 30,(page 27)]

Zhengyan LIN, *Zhejiang University*

We consider the limit distribution of maxima for periodograms of a wide of stationary processes. our method is based on m-dependent approximation for stationary processes and a moderate deviation result.

245 Analytical approaches to stochastic neural systems

[IS 32,(page 36)]

Benjamin LINDNER, *Max Planck Institute for the Physics of Complex Systems*

Fluctuations and noise are important aspects of neural systems. A faithful stochastic modeling and an analytical treatment of the resulting dynamics is needed in order to better understand how noisy neurons transmit and process information. The analysis is complicated by different features of neural systems: (i) the nonlinearity (all-or-none response) of neurons; (ii) the non-Gaussian and non-additive nature of noise sources (e.g. multiplicative shot-noise resulting from synaptic inputs); (iii) the presence of delayed feedback in various neural systems; (iv) slow degrees of freedom (adaptation) in neural dynamics. In my talk I will review several analytical approaches to some of these problems and discuss implications of the results for the neural signal transmission.

246 Large deviations for stochastic evolution equations

[CS 50,(page 43)]

Wei LIU, *Department of Mathematics, University of Bielefeld, Germany*

The Freidlin-Wentzell type large deviation principle is established for the distributions of stochastic evolution equations with general monotone drift and small multiplicative noise. Roughly speaking, besides the standard assumptions for exsitence and uniqueness of strong solution, one only need to assume some additional assumptions on diffusion coefficient in order to establish large deviation principle. As applications we can apply the main result to different type examples of SPDEs (e.g. stochastic reaction-diffusion equation, stochastic porous media and fast diffusion equations, stochastic p-Laplacian equation) in Hilbert space, which also improved some known results in earlier works. The weak convergence approach is employed to verify the Laplace principle, which is equivalent to large deviation principle in our framework.

247 Robust large-margin classifiers

[IS 9,(page 18)]

Yufeng LIU, *University of North Carolina at Chapel Hill*

Classification has become increasingly important as a means for information extraction. Among many different classification methods, margin-based techniques are generally expected to yield good performance, especially for high dimensional data. Typical large margin classifiers include the Support Vector Machine (SVM) and the Penalized Logistic Regression (PLR). Despite successes, there are still some drawbacks in certain situations. In particular, large-margin classifiers with unbounded loss functions such as the SVM and the PLR can be sensitive to outliers in the training sample. In this talk, I will present a group of new robust large-margin classifiers. The proposed techniques are shown to be less sensitive to outliers and deliver more accurate classification boundaries than the original classifiers. Asymptotic results such as Fisher consistency as well as the computational algorithms will be discussed. Numerical examples will be used to demonstrate the performance of the proposed robust classifiers.

248 Calibrating chemical transport models for inferring policy related background levels of ozone

[IS 6,(page 30)]

Zhong LIU, *Department of Statistics, U. of British Columbia*

Nhu LE, *British Columbia Cancer Research Centre and Department of Statistics, U. of British Columbia*

James V. ZIDEK, *Department of Statistics, U. of British Columbia*

Ozone, an air pollutant thought to be harmful to human health, is regulated in many countries. Those regulations which attempt to control the levels of ozone from anthropogenic sources, need to set above the policy related background (PRB) level, i.e., the level produced by non-anthropogenic sources. However, the pervasiveness of ozone means no area in many developed countries is pristine, making the PRB a non-measurable quantity. Hence it is sometimes inferred from the output of deterministic chemical transport models (CTMs) with anthropogenic sources ẗurned off:· But are these outputs meaningful? This talk will explore three approaches to calibrating CTMs using measured pollutant concentrations when there are anthropogenic sources. The first, Bayesian melding is a spatial model while the latter two are regression approaches with bias parameters that calibrate the CTM outputs to the measured ozone levels. Finally the results are compared in a case study involving an ozone field over the central and eastern US. Conclusions of the comparison are also given.

249 The uniform asymptotic normality of the poverty measures

[CS 37,(page 33)]

Gane Samb LO, *Université Gaston Berger de Saint-Louis du Senegal*

Cheikh Tidiane SECK, *Université Paris VI, LSTA*

Serigne Touba SALL, *Ecole Normale Superieure, UCAD, Dakar et Université Paris VI*

We consider a generalized form of the most used poverty measure, that is the Foster, Greer and Thorbecke one, of the form

$$J_n(,Y,\gamma)=\frac{1}{n}\sum_{j=1}^{q}\gamma(\frac{Z-Y_{j,n}}{Z}),$$

computed from the ordered sampled incomes or the expenditures $\mathrm{Y}_{i,n}\leq ...\leq Y_{n,n}$, a poverty line Z, the number q of poor in the sample and a positive and monotone function γ with $\gamma(0)=0$. In order to follow up the poverty evolution, we study the uniform asymptotic behavior of the poverty index when the income or the expenditure is viewed as a continuous time series Y=$\{Y(t),t\geq 0\}$. We establish the asymptotic normality of the process

$$\{J_n(,Y(t),\gamma),\gamma\in\mathcal{M},t\geq 0\}$$

indexed by the time $t\geq 0$ and a suitable class of positive motonone functions. The results are used next for the available poverty databases non rich countries.

250 An ecology conjecture approached by means of a dispersion criterion

[CS 75,(page 56)]

Miguel LOPEZ-DIAZ, *Dpto. Estadistica e Investigacion Operativa. Universidad de Oviedo*

Carlos CARLEOS, *Dpto. Estadistica e Investigacion Operativa. Universidad de Oviedo*

In many animals the sexes show different dispersion tendencies. So females benefit from dispersion, which allows them to choose among males and also avoid inbreeding. Dispersion tends to be female-biased in birds.

Lek mating systems are common in several grouse species, like for instance the capercaillie. Leks are clusters of territories where males congregate and display themselves in order to attract mates. The study of leks is of central importance in behavioral ecology.

Ecologists consider that in the case of capercaillies, leks are loose territories, roughly circular, with a radius depending on the characteristics of the habitat. This radius usually varies from 500 m to 1000 m

It has been assumed that during the mating season, male capercaillies show a behavior different from that of females, in the sense that the latter tend to leave and go into leks, and move around them, more frequently than males, which show more static positions with respect to leks.

This dispersion behavior of the capercaillie has received considerable attention in biological literature, where it is normally considered female-biased. However, and contrary to this traditional view that the

female is the dispersing sex, recently some works suggest a similar dispersion in both sexes in the case of Scandinavian forests.

In this communication we introduce a mathematical model to approach the problem of the dispersion behavior of female and male capercaillies with respect to leks during the mating season. Main properties of the model are obtained. Such a problem is discussed with real data of positions of capercaillies and leks taken by ecologists specialized in the ecosystem of this animal.

The idea for mathematically formalizing the enounced problem is the following. We want to consider the lek as a whole, without distinguishing the position of an animal when it is in such an area. Thus we associate with each animal, the set given by the union of the lek and the singleton which describes its position.

So, we have associated a set-valued mapping with each capercaillie. We are comparing the dispersions of the positions of male and female capercaillies with respect to a lek, by means of the dispersions of their associated set-valued mappings.

The idea for doing this is the following: given the set-valued mappings of two capercaillies of the same sex, we can consider the distance between such mappings by means of an appropriate distance. Intuitively, small values for such distances will show less dispersion than large ones. The above idea will be mathematically modelled by means of an indexed multivariate dispersion ordering.

251 Adaptive estimation of convex functions

[Medallion Lecture 1,(page 15)]

Mark G LOW, *Statistic Department University of Pennsylvania*

Theory is developed for the optimal adaptive estimation of a convex function. Lower bounds are given in terms of a modulus of continuity which depends on the unknown convex function. An estimator is constructed with mean squared error performance which adapts to the unknown function.

252 Accurate likelihood inference in mixed effects models

[CS 35,(page 32)]

Claudia LOZADA-CAN,
Anthony C. DAVISON, *Swiss Federal Institute of Technology Lausanne*

Statistical models with random effects are widely used in applications, under a wide variety of names. One particularly important class of such models is the class of generalized linear mixed models, which include logistic regression and log-linear models with normal random effects. Frequentist inference in such cases is generally based on large-sample likelihood results, so-called first order asymptotics, which are accurate to first order. Recent developments on likelihood inferences result in third order accuracy for continuous data, and second order accuracy for discrete outcomes. In this paper, we discuss higher order methods for parameters of such models.

253 An index of similarity and its applications

[CS 65,(page 49)]

I-Li LU, *The Boeing Company*
Ranjan PAUL, *The Boeing Company*

In this paper, we extend the concept of a similarity measure on strings of binary responses to strings of polychotomous responses. The index generalizes the development of a modified version of the Jaccard-Tanimoto coefficient. Asymptotic theory is employed together with the stability principle to derive the index weights. Maximum likelihood estimators of the probabilities of occurrence for categories are used to estimate the index weights. When the structure of probabilities of occurrence is specified by the Dirichlet priors, estimates based on the admissible minimax principle are computed and compared with those estimated by the maximum likelihood procedures. Asymptotic distributions of these indices are derived. Large sample properties are evaluated through Monte Carlo experiments. Results of simulation are presented. Applications of the proposed indices and their Relations to the Bayesian probabilistic information retrieval are discussed.

254 Sliced inverse factor analysis

[CS 55,(page 44)]

Ronghua LUO, *Guanghua School of Management, Peking University*
Hansheng WANG, *Peking University*
Chih-Ling TSAI, *Graduate School of Management, University of California Davis*

We employ the sliced inverse approach in conjunction with the factor analysis model to obtain a

sufficient dimension reduction method, named sliced inverse factor analysis (SIFA). This method is particularly useful when the number of predictors is substantially larger than the sample size. We show theoretically that SIFA is able to recover the latent factor accurately. In addition, a consistent BIC-type selection criterion is developed to select the structure dimension. Both simulations and a microarray example are presented to illustrate the usefulness of the new method.

255 Semi-parametric mixture distribution and its application

[CS 6,(page 8)]

Jun MA, *Statistics Department, Macquarie University*
Sibba GUDLAUGSDOTTIR,
Graham WOOD,

In practice, we sometimes face the situation where independent observations are obtained from a two-component mixture distribution, where one component possesses a known probability density function but the other do not. We call this a semi-parametric mixture distribution and our aim is to estimate the parameters of this mixture model. We first define the maximum penalized likelihood (MPL) estimates of the mixture model parameters and then develop a generalized EM (GEM) iterative scheme to compute the MPL estimates. A biology example will be given in the talk.

256 Probability and statistics in internet information retrieval

[Medallion Lecture 2,(page 28)]

Zhi-Ming MA, *Inst.Appl.Math, Academy of Math and Systems Science, CAS*

In this talk I shall briefly review some of our recent joint work (in collaboration with Microsoft Research Asia) concerning Internet Information Retrival. Our work reveals that Probability and Statistics is becoming more and more important in the area of Internet information retrieval. It reflects in turn that Internet information retrieval has been a rich source providing plenty of interesting and challenging problems in Probability and Statistics.

257 High dimensional Bayesian classifiers

[IS 9,(page 18)]

David MADIGAN, *Columbia University*

Supervised learning applications in text categorization, authorship attribution, hospital profiling, and many other areas frequently involve training data with more predictors than examples. Regularized logistic models often prove useful in such applications and I will present some experimental results. A Bayesian interpretation of regularization offers advantages. In applications with small numbers of training examples, incorporation of external knowledge via informative priors proves highly effective. Sequential learning algorithms also emerge naturally in the Bayesian approach. Finally I will discuss some recent ideas concerning structured supervised learning problems and connections with social network models.

258 Information-theoretic bounds for compound Poisson approximation

[CS 38,(page 34)]

Mokshay MADIMAN, *Department of Statistics, Yale University*
Andrew BARBOUR, *The Institute of Mathematics, University of Zurich*
Oliver JOHNSON, *Department of Mathematics, University of Bristol*
Ioannis KONTOYIANNIS, *Department of Informatics, Athens University of Economics and Business*

Information-theoretic ideas and techniques are applied to the problem of approximating the distribution of a sum of independent, discrete random variables by an appropriate compound Poisson (CP) distribution. The main goal is to provide nonasymptotic, computable bounds, in a variety of scenarios. Specifically, we derive bounds for the relative entropy distance (or Kullback-Leibler divergence) between the distribution of the sum and an appropriate CP measure, which are, in a certain sense, optimal within the domain of their applicability. Several new bounds are also developed in terms of total variation distance, and it is shown that they are of optimal order in some parameter regimes. Most of the results are obtained via a novel approach that relies on the use of a "local information quantity." These are functionals that play a role analogous to that of the Fisher information in normal approximation. Several local information quantities are introduced, and their utility for CP approximation is explored. In particular, it is shown that they can be employed in connection with some simple techniques related to Stein's method, to obtain strong new bounds. Finally, several monotonicity properties are demonstrated for the convergence of sums to Poisson and compound Poisson distributions; these are natural analogs of the recently established monotonicity of Fisher information in the central limit theorem.

259 The limits of nested subclasses of several classes of infintely divisible distributions are identical

[CS 37,(page 33)]

Makoto MAEJIMA, *Keio University, Yokohama*
Ken-iti SATO,

It is shown that the limits of the nested subclasses of five classes of infintely divisible distributions on $\mathbf{R}^d$, which are the Jurek class, the Goldie-Steutel-Bondesson class, the class of selfdecomposable distributions, the Thorin class and the class of generalized type G distributions, are identical with the closure of the class of stable distributions, where the closure is taken under convolution and weak convergence.

260 Edgeworth expansion for the kernel quantile estimator

[CS 25,(page 24)]

Yoshihiko MAESONO, *Faculty of Mathematics, Kyushu University, Fukuoka, Japan*
Spiridon PENEV, *Department of Statistics, School of Mathematics and Statistics, The University of New South Wales, Sydney, Australia*

Using the kernel estimator of the p-th quantile of a distribution brings about an improvement in comparison to the sample quantile estimator. The size and order of this improvement is revealed when studying the Edgeworth expansion of the kernel estimator. Using one more term beyond the normal approximation significantly improves the accuracy for small to moderate samples. The investigation is non-standard since the influence function of the resulting L-statistic explicitly depends on the sample size. We obtain the expansion, justify its validity and demonstrate the numerical gains in using it.

261 Second order approximation to the risk of the sequential procedure in estimation of a function of the Poisson parameter

[CS 5,(page 8)]

Eisa MAHMOUDI, *Department of Statistics, Yazd University, Yazd, Iran*
Hamzeh TORABI, *Department of Statistics, Yazd University, Yazd, Iran*

Sequential point estimation of the Poisson parameter, subject to the loss function given as a sum of the squared log error and a linear cost is considered. For a fully sequential sampling scheme, we present a sufficient condition to get a second order approximation to the risk of the sequential procedure, as the cost per observation tends to zero. In the end, to justify the theoretical results we shall give brief simulation, using Monte-Carlo method.

262 A novel estimator of additive measurement error in linear models

[CS 7,(page 9)]

Anirban MAJUMDAR, *Graduate Business School, University of Chicago*
Indrajit MITRA, *MSCI Barra*

In this talk we present a novel method to estimate the variance of additive measurement errors. The estimator is manifestly unbiased up to quadratic order in estimation error and outperforms a standard benchmark in a mean squared error sense. An important application of this technique is in single factor linear models which appear in many areas of finance. It is well known that additive measurement errors in the indicator variable lead to biased, attenuated estimates of the sensitivity (”beta”) of the dependent variable to the indicator. Good estimates of the variance of the measurement errors are very useful in improving estimates of beta.

263 Stochastic Modeling of 2-out-of-3 redundant system subject to degradation

[CS 12,(page 13)]

Suresh Chander MALIK, *Reader, Department of Statistics, M.D.University, Rohtak-124001, India*
M.S. KADYAN, *Kurukshetra University, Kurukshetra, India*

The aim of this paper is to develop a stochastic model for 2-out-of-3 redundant system in which unit becomes degraded after repair. There is a single server who plays the dual role of inspection and repair. The system is considered in up-state if any of two original/degraded or both units are operative. Server inspects the degraded unit at its failure to see the feasibility of repair. If repair of the degraded unit is not feasible, it is replaced by new one. The original (or new) unit gets priority in operation over

the degraded unit. The distributions of failure time of the units follow negative exponential while that of inspection and repair times are taken as arbitrary with different probability density functions. Various reliability and economic measures are obtained by using semi-Markov process and regenerative point technique. Graphs are drawn to depict the behaviour of MTSF, availability and profit of the model for a particular case.

264 Nonparametric regression on unobserved latent covariates with application to semiparametric GARCH-in-mean Models

[IS 19,(page 35)]

Enno MAMMEN, *University of Mannheim*
Christian CONRAD, *ETH Zurich*

We consider time series models where the conditional mean of the time series given the past depends on unobserved latent covariates. We assume that the covariate can be estimated consistently and use an iterative nonparametric kernel smoothing procedure for estimating the dependence of the observation on the unobserved covariate. The covariates are assumed to depend parametrically on past values of the covariates and of the residuals. Our procedure is based on iterative fits of the covariates and nonparametric kernel smoothing of the conditional mean function. In the paper we develop an asymptotic theory for the resulting kernel estimator and we use this estimator for testing parametric specifications of the mean function. Our leading example is a semiparametric class of GARCH-in-Mean models. In this setup our procedure provides a formal framework for testing economic theories which postulate functional relations between macroeconomic or financial variables and their conditional second moments. We illustrate the usefulness of the methodology by testing the linear risk-return relation predicted by the ICAPM.

265 Outliers in uniform distribution

[PS 3,(page 29)]

Sadaf MANZOOR, *Ph.D Student, Martin-Luther-University, Germany*

Observations which deviate strongy from the main part of the data, usually lebeled as 'outliers' are troublesome and may cause completely misleading results. Therefore, the solitariness of outliers is important for quality assurance. It is desirable to derive such approaches in a systematic manner from general principles and guidlines rather than human decision making or simply plotting the data.

A test for outliers of normally distributed data has been developed by J. W. Dixon (1950). The present study uses the same idea by arranging the data set in ascending order for the particular case where samples are comming out of Uniform distribution. Percentage points are tabulated for testing hypothesis and constructing confidence intervals for different significance levels.

266 Recursive Monte Carlo methods for nonlinear tracking and fixed-parameter estimation in a chaotic CO_2 laser

[CS 57,(page 46)]

Ines P. MARINO, *Universidad Rey Juan Carlos*
Joaquin MIGUEZ, *Universidad Carlos III de Madrid*
Riccardo MEUCCI, *Istituto Nazionale di Ottica Applicata*

The tracking of nonlinear dynamical systems is needed in many important problems in science and engineering. It usually involves the estimation of a set of latent (hidden) variables in a dynamical model from the observation of a noisy time series. During the last years, sequential Monte Carlo (SMC) methods, also known as "particle filters", have emerged as powerful tools for this class of problems.

In this work, we address the estimation of the latent physical variables that govern the dynamics of a CO_2 laser with modulated losses in a chaotic regime. Full details of the experimental setup from which the data are collected can be found in (Mariño et al, PRE 70(036208), 2004). Only the one-dimensional time series representing the output laser intensity is directly observable, but the complete numerical model of the system consists of five dynamical state variables and five nonlinear differential equations that depend on a set of fixed parameters. Both the state variables and the static parameters have sharply defined, and relevant, physical meanings, hence their accurate estimation is very important for the study of the system. Unfortunately, the model, in its standard form, reproduces *the type of* chaotic behavior displayed by the system, but its state paths *cannot follow* the actual time series obtained from the experiment.

To tackle this problem, we have modified the model by adding dynamical noise to the latent state variables. An adequate selection of the noise distribution ensures that the random model exhibits the same dynamical features as its deterministic counterpart (i.e., the resulting phase-space plots are just slightly perturbed versions of the deterministic ones) but it provides the means to apply powerful SMC methods to the estimation of the system state. We have shown that some standard techniques, including the sequential importance resampling algorithm and the auxiliary particle filter (APF), can yield estimates of the system latent variables and, therefore, provide a better insight of the physical system.

We have also addressed the estimation of the model fixed parameters. However, the joint estimation of (random) static parameters and dynamical variables using SMC methods is still an open problem in the literature. In this case, we have applied both APF-based algorithms and a novel SMC optimization technique. The results yield numerical approximations of several static parameters which differ from those obtained using "traditional" deterministic tools, but are in good agreement with physical interpretations of the system.

267 Asymptotics for spherical needlets

[IS 4,(page 55)]

Domenico MARINUCCI, *Department of Mathematics, University of Rome Tor Vergata*
Paolo BALDI, *Department of Mathematics, University of Rome Tor Vergata*
Gerard KERKYACHARIAN, *LPMA, Jussieu, Paris*
Dominique PICARD, *LPMA, Jussieu, Paris*

In recent years an enormous amount of attention has been driven by the statistical analysis of cosmological data, in particular in connection with Cosmic Microwave Background radiation datasets (CMB). In this talk, we shall propose a new approach to the construction of spherical wavelets and discuss their applications to CMB data analysis. More precisely, we investigate invariant random fields on the sphere using so-called needlets. These are compactly supported in frequency and enjoy excellent localization properties in real space, with quasi-exponentially decaying tails. We show that, for random fields on the sphere, the needlet coefficients are asymptotically uncorrelated for any fixed angular distance. This property is used to derive CLT and functional CLT convergence results for polynomial functionals of the needlet coefficients: here the asymptotic theory is considered in the high frequency sense. We then discuss applications to angular power spectrum estimation, testing for Gaussianity and testing for cross-correlation between CMB and Large Scale Structure data on galaxy distributions.

References

1. Baldi, P., Kerkyacharian, G., Marinucci, D., Picard, D. (2008) "Asymptotics for Spherical Needlets", Annals of Statistics, in press.

268 On the extremes of nonlinear time series models applied to river flows

[CS 14,(page 14)]

László MÁRKUS, *Eötvös Loránd University, Budapest, Hungary*

Our aim is the evaluation of inundation risks of Danube and Tisza Rivers in Hungary, on the basis of daily river discharge data for the entire last century. Elek and Markus (JTSA 2008) suggests a light tailed conditionally heteroscedastic model for the deseasonalised river flow series X_t:

$$\begin{aligned} X_t &= \sum_{i=1}^{p} a_i(X_{t-i}) + \sum_{i=1}^{q} b_i\epsilon_{t-i} \\ \epsilon_t &= \sigma(X_{t-1})Z_t, \\ \sigma(x) &= (\alpha_0 + \alpha_1(x-m)_+)^{1/2}. \end{aligned}$$

with positive constants $a_i, b_i, \alpha_0, \alpha_1, m$, innovation ϵ_t and noise Z_t. The model differs from conventional ARMA-GARCH ones as the variance of *innovations* is conditioned on the lagged values of the *generated process* and is asymptotically *proportional* to its past values.

Alternatively, Vasas et al. (JSPI 2007) propose a switching autoregression, with altering coefficients and innovation distributions along two-regimes (called "ascending" and "descending"), and a non-Markovian hidden regime indicator process. The ascending regime is a random walk generated by an i.i.d. Gamma noise, the descending one is a Gaussian AR(1) series. The duration of the ascending regime is distributed as negative binomial, whereas the duration of the descending regime is geometrically distributed. Aside of classical criteria, model fit is evaluated by the extremal behavior of empirical and fitted series. Properties of the extremal characteristics of the models, such as quantiles (return levels), maxima, extremal index (clustering of high flows), time spent over threshold (flood duration) and aggregate excesses (flood volume) will be reported in the talk.

269 Student t-statistic distribution for non-Gaussian populations

[CS 41,(page 37)]

João Paulo MARTINS, *CEAUL e ESTG – Instituto Politécnico de Leiria*

$t_{(n-1)} = \sqrt{n}\frac{\overline{X}_n-\mu}{S_n}$, when the parent population is Gau (μ, σ), is easily derived since $\overline{X}_n$ and S_n are independent. However this is an exceptional situation and for whatever else regular parent $\overline{X}_n$ and S_n are dependent. It is this dependence structure that difficults the computation of the exact distribution of $T_{n-1} = \sqrt{n}\frac{\overline{Y}_n-\mu}{S_n}$ for non-Gaussian Y parent.

Our aim has been to investigate, for general parent Y with known a- symmetry and kurtosis, whether there in the Pearson system of distributions one type that provides a better approximation to $T_{n-1} = \sqrt{n}\frac{\overline{Y}-\mu}{S_n}$, in the specific sense that the higher percentiles of $X_{\beta,\mu,\sigma}$ provide better approximations than $t_{(n-1)}$ to the corresponding percentiles of T_{n-1}. In fact, we could establish that the pre-asymptotic behaviour of T_{n-1} distribution is very close to the type IV distributions. This is to a certain extent surprising, since t_n is not of Pearson type IV. Observe that so far Pearson type IV family hasn't been widely used in statistical modeling.

When $\frac{\overline{X}-\mu}{\sigma} \approx Z \frown \text{Gau}\,(0,1)$ and $\dfrac{S_n}{\sigma} \underset{n\to\infty}{\to} 1$, there are some grounds to expect that $T_{n-1} \approx t_{(n-1)}$; this pre-asymptotic approximation is also investigated.

270 Multiclass functional discriminant analysis and its application to gesture recognition

[CS 64,(page 48)]

Hidetoshi MATSUI, *Graduate School of Mathematics, Kyushu University*
Takamitsu ARAKI, *TOME R&D Inc.*
Sadanori KONISHI, *Faculty of Mathematics, Kyushu University*

Functional data analysis provides a useful tool for analyzing a data set observed at possibly differing time points for each individual, and its effectiveness has been reported in various science fields such as bioscience, ergonomics and signal processing. The basic idea behind functional data analysis is to express discrete data as a smooth function and then draw information from the collection of functional data.

We introduce a multiclass functional discriminant procedure for classifying multiple functionalized data; discrete observations are transformed to a set of functions, using Gaussian basis functions with help of regularization, and then a discriminant rule is constructed by applying logistic modeling. Advantages of our modeling strategy are that it provides a flexible instrument for transforming discrete observations into functional form and that it can also be applied to analyze a set of surface fitting data. Furthermore, the proposed functional logistic discriminant procedure enables us to consider a suggestion for misclassified data by using the posterior probabilities.

A crucial issue in functional logistic discrimination is the choice of smoothing parameters involved in the regularized maximum likelihood procedure. We derive a model selection criterion from a Bayesian viewpoint for evaluating models estimated by the regularization method in the context of functional discriminant analysis.

The proposed modeling strategy is applied to the recognition of handwriting characters. Handwriting characters are written in the air with one of our fingers and captured by a video camera in order to record the trajectories of moving fingers. The trajectories corresponding to the XY coordinate values are transformed to smooth functional data, and classify them into several patterns. The results show that our functional discriminant procedure provides a useful tool for classifying functions or curves.

271 Quantifying field transmission of Tasmanian devil facial tumour disease

[IS 21,(page 31)]

Hamish MCCALLUM, *School of Zoology The University of Tasmania*

Tasmanian Devil facial tumour disease is an infectious cancer that is threatening to cause the extinction of the largest surviving marsupial carnivore. The tumour, which emerged in the mid-1990s is thought to be spread by biting and has spread across most of the range of the devil, causing an overall population decline of 50 percent (up to 90 percent in affected areas). It is therefore critical to develop strategies to manage this disease threat. To evaluate alternative management strategies, we need to estimate the rate of disease transmission in natural populations and to determine whether and how this rate depends on host density. Estimating the transmission rate of any wildlife disease under field conditions is difficult. We have extensive individual-based mark recapture data from one site (Freycinet National Park) that has been regularly monitored from three years before disease arrival until the present (five years after disease arrival). We also have similar data at two other sites from the time of disease arrival and information from three other sites where disease was established when surveys commenced. To estimate transmission, we have modelled capture histories of individuals. Estimated recapture probabilities did not vary between infected and uninfected devils, providing confidence that disease prevalence in captured animals is an unbiased estimate of population prevalence. Prevalence did not differ between male and female devils, although there were strong age class effects. The rate of increase in prevalence with time differed somewhat between the three sites that had been monitored from the time of disease arrival, but was not related to host density. Prevalence remained high in areas where disease was well-established, despite major decreases in devil density, suggesting little association between host density and transmission. At the Freycinet site we have used multistate models to estimate transition rates from healthy to diseased states, which provides an estimate of the force of infection. The main gap in our current knowledge, which inhibits our ability to estimate the crucial parameter R_0, is that we have a poor understanding of the incubation and latent periods: we do not know the interval between acquiring disease and either showing clinical signs or becoming infectious. Nevertheless, using available evidence, we have estimated a plausible range for R_0. This suggests that it may be possible to suppress disease by removing at least 50 percent of diseased animals every three months. A trial on a semi-isolated peninsula provides evidence that this strategy might be successful.

272 Efficiently estimating personal network size

[CS 54,(page 44)]

Tyler H. MCCORMICK, *Department of Statistics, Columbia University*
Tian ZHENG, *Department of Statistics, Columbia University*
Matthew J. SALGANIK, *Department of Sociology, Princeton University*

In this paper we develop a method to estimate both individual social network size (i.e., degree) and the distribution of network sizes in a population by asking respondents how many people they know in specific subpopulations (e.g., people named Kevin). Building on the scale-up method of Killworth et al. (1998) and other previous attempts to estimate individual network size, we first propose a latent non-random mixing model which resolves three known problems with previous approaches. As a byproduct, our method also provides estimates of the rate of social mixing between population groups. We demonstrate the model using a sample of 1,370 adults originally collected by McCarty et al. (2001). Based on insights developed during the statistical modeling, we conclude by offering practical guidelines for the design of future surveys in this area. Most importantly, we show that if the specific subpopulations are chosen wisely, complex statistical procedures are no longer required for estimation.

273 Representations of the moments of the Dickey-Fuller and related distributions

[CS 30,(page 27)]

J. Roderick MCCRORIE, *School of Economics and Finance, University of St Andrews*

This paper offers a Mellin transform approach to the open problem of characterizing the laws and moments of test statistics pertaining to an autoregressive model under the unit root hypothesis. We find that the moments of the Dickey-Fuller and the related t-type distribution are expressible as series involving the moments of two related distributions associated with the hyperbolic functions cosh and tanh, the Mellin transforms of which relate to a generalized Hurwitz zeta function which is interpolated at the integers by a generalized hypergeometric function of specific form. This result, which can be given a probabilistic interpretation, offers expressions for the moments and, when suitably generalized, has implications in terms of describing the closed-form summation of certain Dirichlet L-functions and their generalizations.

274 Statistical models with autogenerated units

[Neyman Lecture 1,(page 41)]

Peter MCCULLAGH, *University of Chicago*

In the formal theory of stochastic processes, which includes conventional regression models, the index set of potential experimental or observational units is fixed, and usually infinite. The response distribution, which is specified in a consistent manner by the regression model $p_{\mathbf{x}}(\mathbf{y})$ for each fixed finite sample of units, depends on the sample configuration $\mathbf{x} = (x(u_1), \ldots, x(u_n))$ of covariate values. Although the definition of a fixed sample is unambiguous mathematically, the meaning is not at all clear in the majority of applications. Random samples of units are hard to avoid in biological work because the population units are typically unlabelled. Sequential recruitment of units is standard practice in clinical work, ecological studies and market research: labels affixed to the sample units after recruitment tend to obscure the sampling scheme. It is by no means obvious that standard models with distributions determined for *fixed* samples are suitable for applications in which the units are unlabelled and samples are random. I propose an alternative process in which the formal concept of a fixed set of statistical units is absent. Instead, the process itself generates a stream of 'units' in time, each unit being identified with its (x, y, t) value. Samples are automatically random because the units themselves are random. The relation between the conditional distribution $p(\mathbf{y} \mid \mathbf{x})$ for a sequential samples, and the stratum distribution $p_{\mathbf{x}}(\mathbf{y})$ for fixed quota $\mathbf{x}$, will be discussed in the context of random-effects models. This analysis reveals that that the phenomenon of parameter attenuation in logistic models is a statistical illusion caused by sampling bias.

275 Genetic association studies with known and unknown population structure

[IS 24,(page 9)]

Mary Sara MCPEEK, *University of Chicago Department of Statistics*
Mark ABNEY, *University of Chicago Department of Human Genetics*

Common diseases such as asthma, diabetes, and hypertension, which currently account for a large portion of the health care burden, are complex in the sense that they are influenced by many factors, both environmental and genetic. One fundamental problem of interest is to understand what the genetic risk factors are that predispose some people to get a particular complex disease. Technological advances have made it feasible to perform case-control association studies on a genome-wide basis. The observations in these studies can have several sources of dependence, including population structure and relatedness among the sampled individuals, where some of this structure may be known and some unknown. Other characteristics of the data include missing information, and the need to analyze hundreds of thousands or millions of markers in a single study, which puts a premium on computational speed of the methods. We describe a combined approach to these problems which incorporates quasi-likelihood methods for known structure with principal components analysis for unknown structure.

276 High-dimensional l_1 penalized regression: fast algorithms and extensions

[CS 26,(page 25)]

Lukas MEIER, *ETH Zurich*
Peter BÜHLMANN, *ETH Zurich*

l_1 penalization approaches like the (adaptive) Lasso are very successful for analyzing high-dimensional data-sets, both from a theoretical and a practical point of view. Their success crucially depends on the availability of fast algorithms to solve the corresponding convex optimization problem. More recently, the focus has come (back) to coordinatewise (one at a time) minimization approaches which are suitable for very high-dimensional problems where the dimensionality of the predictor can be in the thousands. We give an overview of current approaches and show how they can be extended to more general models and penalties, e.g. for the Group Lasso penalty in generalized linear models. Moreover, we show Lasso extensions which lead to more interpretable models and better estimation performance.

277 Stability and sparsity

[IS 11,(page 23)]

Nicolai MEINSHAUSEN, *University of Oxford*

The properties of L1-penalized regression have been examined in detail in recent years. I will review some of the developments for sparse high-dimensional data, where the number of variables p is potentially very much larger than sample size n. The necessary conditions for convergence are less restrictive if looking for convergence in L2-norm than if looking for convergence in L0-quasi-norm. I will discuss some implications of these results. These promising theoretical developments notwithstanding, it is unfortunately often observed in practice that solutions are highly unstable. If running the same model selection procedure on a new set of samples, or indeed a subsample, results can change drastically. The choice of the proper regularization parameter is also not obvious in practice, especially if one is primarily interested in structure estimation and only secondarily in prediction. Some preliminary results suggest, though, that the stability or instability of results is informative when looking for suitable data-adaptive

regularization.

278 On Jensen inequality for medians and centers of distributions

[CS 10,(page 12)]

Milan MERKLE, *University of Belgrade, Belgrade, Serbia*

The standard Jensen inequality that has found numerous applications in Statistics and Probability theory, states that $f(m) \leq M$ where f is a convex function on some interval D, m is the expectation of a distribution on D, and M is the expectation of the set $f(D)$. It turns out that the same inequality holds when means are replaced by medians, even with a class of functions that is strictly larger than the class of convex functions. This result can be also extended to multidimensional setup, with various choices for medians (or, more appropriately, centers of distributions).

279 On the mean of a stochastic integral with non-Gaussian α-stable noise

[CS 77,(page 57)]

Zbigniew MICHNA, *Department of Mathematics and Cybernetics Wroclaw University of Economics Wroclaw, Poland*

In this paper we consider a Lévy process under condition $\Gamma_1 = x$ where $\{\Gamma_k\}$ is a sequence of arrivals of a Poisson process with unit arrival rate. We show that under condition $\Gamma_1 = x$ Lévy process can be decomposed into a simple process and a Lévy process. These two processes are independent. As an application of this decomposition we consider α-stable Lévy processes. We give a closed form of the process $X(t) = \int_0^t Z(s-)\,dZ(s)$ and evaluate its expected value where Z is an α-stable Lévy process with $0 < \alpha < 2$. We show that $EX(t) = 0$ for $1 < \alpha < 2$ and this expectation is equal infinity for $\alpha \leq 1$.

280 The assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints

[CS 63,(page 48)]

M. MIELKE, *Institute for Mathematical Stochastics, Georg-August-University of Goettingen*
A. MUNK, *Institute for Mathematical Stochastics, Georg-August-University of Goettingen*

The objective of this paper is to develop statistical methodology for noninferiority hypotheses to censored, exponentially distributed time to event endpoints. Motivated by a recent clinical trial in depression we consider a gold standard design where a test group is compared to an active reference and to a placebo group. The test problem is formulated in terms of a retention of effect hypothesis. Thus, the proposed test procedure assures that the effect of the test group is better than a pre-specified proportion *Delta* of the treatment effect of the reference group compared to the placebo group. A sample size allocation ratio to the three groups to achieve optimal power is presented, which only depends on the pre-specified *Delta*. In addition, a pretest is presented for either the reference or the test group to ensure assay sensitivity in the complete test procedure. The actual type I error and the sample size formula of the proposed tests is explored asymptotically and by means of a simulation study showing good small sample characteristics. To illustrate the procedure a randomized, double blind clinical trial in depression is evaluated. This paper is a joint work with A. Munk (University of Goettingen) and A. Schacht (Lilly Deutschland GmbH).

281 Some geometric aspects of large random maps

[IS 18,(page 19)]

Grégory MIERMONT, *CNRS and Fondation des Sciences Mathématiques de Paris*

Random maps arise in the Physics literature as a way to discretize certain ill-defined integrals on the space of all surfaces. On a mathematical level, this gives rise to a number of problems about the convergence of suitably rescaled random maps, considered as random metric spaces, towards a universal limiting random surface. A complete characterization of this limiting space is one of the most fundamental issues that are to be faced. At present, only partial information is available, essentially for surfaces of genus 0, such as the typical distance between two randomly chosen points on the space (Chassaing and Schaeffer), the universal aspects of the latter result (Marckert, Miermont, Weill), or the fact that the limit has a.s. the topology of the sphere (Le Gall and Paulin).

In this talk, we will present some geometric properties that are satisfied by the limiting space, such as the uniqueness of the typical geodesic linking two points. This relies on generalizations of Schaeffer's bijection between maps and labeled trees that are valid for any genus, and allow to keep track of the metric structure of the map locally around an arbitrary number of points.

282 A sequential Monte Carlo optimization method for state-space random dynamical models

[CS 57,(page 46)]

Joaquin MIGUEZ, *Universidad Carlos III de Madrid*

Consider a discrete-time random dynamical system in state-space form and the problem of tracking the system state from the series of associated observations. When the system is linear and Gaussian the Kalman filter yields an exact solutions. More general nonlinear and/or non-Gaussian systems, however, demand the use of numerical approximation techniques. Sequential Monte Carlo (SMC) methods, also known as particle filters, are a family of simulation-based algorithms that recursively approximate the probability density function of the state given the available observations.

In practice, the particle filter is derived for a specific model of the system of interest. If there is some significant statistical discrepancy between the assumed model and the observed time series, one can expect a degradation of the algorithm performance. Discrepancies can affect the state dynamics, the observations or both. E.g., in a target tracking problem we may choose a linear motion model, but the target exhibit highly maneuvering dynamics. Also, observational noise in digital communication receivers is often assumed Gaussian, but impulsive processes appear in many environments.

Recently, a SMC methodology specifically aimed at estimation and prediction in nonlinear dynamical systems for which a reliable state-space model is not available has been proposed (Miguez et al, EURASIP JASP, 2004(15):2278-2294). The key of the new approach is to identify a cost function whose minima provide valid estimates of the system state at successive time instants. This function is optimized using an algorithmic procedure analogous to conventional SMC techniques, hence we terme the method as sequential Monte Carlo minimization (SMCM). The main advantage of the methodology is its flexibility, which makes very simple for practitioners to derive, code and evaluate new algorithms.

In the talk, we will describe an extension of the original SMCM algorithm and an asymptotic convergence analysis of the resulting method. Specifically, we will define a broader class of cost functions that can be optimized, including those that involve the state dynamics explicitly, and extend the methodology to handle them. Then, we will provide sufficient conditions for the resulting algorithms to yield system-state estimates that converge, in probability, to a sequence of cost minimizers. A discussion on the implications of these conditions leads to a comparison with standard SMC algorithms in terms of practical design. Finally, we consider two application examples to numerically illustrate how SMCM methods converge in the way predicted by the theory and their robustness when the actual observations and the assumed model are discrepant.

283 Asymptotic behavior and rate of convergence of the parameter estimators for the Ornstein-Uhlenbeck and Feller diffusions. Application to neuronal models.

[CS 32,(page 31)]

Rosa Maria MININNI, *Departement of Mathematics, University of Bari, Via Orabona 4, 70125 Bari, Italy*
Maria Teresa GIRAUDO, *Departement of Mathematics, University of Torino, Via Carlo Alberto 10, 10123 Torino, Italy*
Laura SACERDOTE, *Departement of Mathematics, University of Torino, Via Carlo Alberto 10, 10123 Torino, Italy*

We consider a sample of i.i.d. times and we interpret each item as the first passage time (FPT) of a diffusion process $X = \{X_t;\, t \geq 0\}$ through a constant boundary S. This representation models a variety of applications in reliability theory, engineering, neurobiology and so on.

We are interested in the estimation of the parameters characterizing the underlying diffusion process through the knowledge of the observed FPT

$$T = inf\,\{t > 0 :\ X_t \geq S;\ X_0 = x_0\}\,,$$

where x_0 is considered non-random.

The experimentally observable data are then the FPT's of the underlying stochastic process. Despite the conceptual simplicity of this description, its practical use is made difficult by the impossibility to get closed form expressions for the FPT probability density (pdf) except for some special cases (cf. [4]). We consider as underlying processes the Ornstein-Uhlenbeck and the Feller processes that play a relevant role in different areas like neurosciences, survival analysis and mathematical finance. Both of them are fully described by five parameters. Their FPT pdf is known only through numerical and simulation techniques or asymptotically (cf. [3]). Hence, standard statistical inference such as maximum likelihood or Bayes estimation cannot be applied. Recently, within the neurobiological context in [1] and [2] moment type estimators of two model parameters have been proposed. Closed expressions have been derived, suggesting their qualitative and asymptotic behavior by means of numerical computations on simulated data.

Here as our primary goal we study the asymptotic properties (consistency and asymptotic normality) of the estimators obtained in [1] and [2]. Further, we establish upper bounds for the rate of convergence of the empirical distribution of each estimator to the normal density. Applications to the neurobiological context are also considered. The accuracy of the moment type estimators and the goodness of analytical approximations to the normal density are discussed by means of simulated experiments.

References

1. S. Ditlevsen and P. Lánský, Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model, *Phys. Rev. E*, **71**, Art. No. 011907 (2005).
2. S. Ditlevsen and P. Lánský, Estimation of the input parameters in the Feller neuronal model, *Phys. Rev. E*, **73**, Art. No. 061910 (2006).
3. A.G. Nobile, L.M. Ricciardi and L. Sacerdote, Exponential trends for a class of diffusion processes with steady state distribution, *J. Appl. Prob.* **22**: 611-618 (1985).
4. L.M. Ricciardi, A. Di Crescenzo, V. Giorno and A.G. Nobile, An outline of theoretical and algorithmic approaches to first passage time problems with application to biological modeling, *Math. Japonica* **50**, 2: 247–322 (1999).

284 Optimal oracle inequalities for model selection

[CS 46,(page 41)]

Charles MITCHELL, *ETH, Zurich*
Sara VAN DE GEER, *ETH, Zurich*

While statistics has been very successful in constructing an ever-increasing array of estimators tailored to specific conditions and optimal under the right assumptions, this development has also increased the need for effective model selection and aggregation techniques, and for precise knowledge about their effectiveness. The effectiveness of a model selection or aggregation procedure can be expressed by an oracle inequality, which compares the performance of the selected or aggregated procedure to the performance of the best possible choice.

Most of the work about model selection and aggregation is, however, specific to a particular problem, such as fixed-design regression or density estimation, and moreover restrictive conditions such as boundedness are usually employed. Our approach is a more general one: we describe model selection problems using general loss functions, and investigate what conditions on these loss functions suffice to produce good oracle inequalities. This we do for model selection that uses empirical risk minimization. Describing excess losses using empirical processes, we then work with margin and envelope conditions and use concentration inequalities to derive oracle risk bounds. We thus present a very general oracle inequality for model selection among fixed loss functions, and show that it has an optimal rate. Here the conditions that lead to the oracle inequality are clearly recognizable and can be tracked when it is applied to regression, machine learning and density estimation. We also look ahead to the two-stage estimation procedure involving data splitting, and sketch what restrictions need to be imposed in order to obtain oracle inequalities there.

285 On moment-type estimation in some statistical inverse problems

[CS 5,(page 8)]

R. M. MNATSAKANOV, *Department of Statistics, West Virginia University, Morgantown, WV 26506, USA*

Some fundamental properties of moment-recovered distributions and its probability density functions have been investigated in Mnatsakanov (2008a, 2008b). In this talk we apply these moment type reconstructions in certain statistical inverse problems, e.g., those based on mixtures, convolutions, biased sampling, and multiplicative censoring models. Namely, given the sample from a distribution, say G, which is related in some specific way to the target (unknown) distribution F, the problem of recovering (estimating) the distribution F, its density function f, or corresponding quantile function Q is studied. The convergence rates of proposed estimators are derived.

References

1. Mnatsakanov, R.M. (2008a). Hausdorff moment problem: Reconstruction of distributions.*J. Statist. Probab. Lett.*, Doi: 10.1016/j.spl.2008.01.011.

2. Mnatsakanov, R.M. (2008b). Hausdorff moment problem: Reconstruction of probability density functions. *J. Statist. Probab. Lett.*, Doi: 10.1016/j.spl.2008.01.054.

286 Integro-local theorems on the whole semiaxis for sums of random variables with regular distributions

[CS 37,(page 33)]

Anatolii A. MOGULSKII, *Sobolev Institute of Mathematics, Novosibirsk, RUSSIA*

Let $S_n = \xi_1 + ... + \xi_n$ be partial sums of i.i.d. random variables. Integro-local theorems on asymptotics of the probabilities

$$\mathbf{P}(S_n \in [x, x + \Delta))$$

were obtained on the whole semiaxis in the cases:

1. The right tail of distribution of the summand $\xi := \xi_1$ has the form

$$\mathbf{P}(\xi \geq t) = t^{-\beta} L(t), \quad \beta > 2,$$

where $L(t)$ is a slowly varying function as $t \to \infty$ ([1], [2]).

2. The distribution of the summand ξ is semiexponential; i.e.,

$$\mathbf{P}(\xi \geq t) = e^{-t^{\beta} L(t)}, \quad \beta \in (0,\ 1),$$

where $L(t)$ is a slowly varying function as $t \to \infty$ possessing some smoothness properties ([3]).

3. The Cramér condition holds

$$\mathbf{E}e^{\delta\xi} < \infty \quad \text{for some} \quad \delta > 0,$$

and one of the following conditions is valid:

a) the relative (scaled) deviations x/n remain in the analyticity domain $\mathcal{A}'$ of the large deviations rate function $\Lambda(\alpha)$ for the summands ([4],[5],[6]);

b) the relative deviations x/n lie on the boundary of or outside the analyticity domain $\mathcal{A}'$ ([7]).

In all the considerations integro-local theorems for sums of random vectors (variables), forming a triangle array, in the domain of normal deviations ([8]) play the important role.

References

1. *Borovkov A.A. and Borovkov K.A.* Aymptotic Analysis of Random Walks. Part I: Slowly Decreasing Jumps (in Russian), Nauka, Moscow (to appear).
2. *Mogulskii A.A.* Integro-local theorem on the whole semiaxis for sums random variables with regular distributions, Siberian Math.J. (to appear).
3. *Borovkov A.A. and Mogulskii A.A.* Integro-local and integral theorems for sums of random variables with semiexponential distributions, Siberian Math.J., V.47, N6, 990—1026 (2006).
4. *Borovkov A.A. and Mogulskii A.A.* Large deviations and super large deviations of the sums of independent random vectors with the Cramér condition. I. Teor. Veroyatnost. i Primenen., V. 51, N2, 260—294 (2006).
5. *Borovkov A.A. and Mogulskii A.A.* Large deviations and super large deviations of the sums of independent random vectors with the condition. II. Teor. Veroyatnost. i Primenen., V. 51, N4, 641-673 (2006).
6. *Mogulskii A.A. and Pagma Ch.* Super large deviations of the sums of random variables with common arithmetic distribution (to appear).
7. *Borovkov A.A. and Mogulskii A.A.* Large deviation probabilities for the sums of random vectors on the boundary and outside of the Cramér zone, Teor. Veroyatnost. i Primenen.(to appear).
8. *Borovkov A.A. and Mogulskii A.A.* Integro-local theorems for sums of random vectors in a series scheme. Mat. Zametki, V.79, N4, 468-482 (2006).

287 Some contributions to the class of two sex branching processes in random environment

[CS 66,(page 49)]

Manuel MOLINA, *Department of Mathematics. University of Extremadura. Badajoz, Spain*

Shixia MA, *Department of Applied Mathematics. Hebei University of Technology. Tianjin, China*

Branching process theory provides stochastic models for description of populations where an individual exists for a time and then may be replaced by others of a similar or different type. Nowadays, it is an active research area with theoretical interest and practical applications in several fields. In particular, in order to describe the evolution of two-sex populations several classes of stochastic processes have been introduced, we refer the reader to Haccou et al. (2005) or Hull (2003) for surveys about some of these classes of processes. However, the range of two-sex stochastic models considered until now is not large enough to get an optimum mathematical modelling in certain populations with sexual reproduction. For instance, it could be advisable to assume that the limiting evolution of some two-sex populations is governed by an environmental process. This question has been studied in asexual branching populations, but it has not been investigated for populations with sexual reproduction. In this work, in an attempt to contribute some solution to this problem, we will introduce the class of two-sex branching processes with offspring probability distribution and mating function depending on an environmental process. We will consider that the environmental process is formed by independent but not necessarily identically distributed random variables and we will prove that, conditionally to the environmental process, the stochastic sequences corresponding to the number of females and males, and the number of couples, in the population are Markov chains. For such a class of bisexual branching proceses, several theoretical results will be provided. In particular, some relations among the probability generating functions involved in the probability model and expressions for its main moments will be determined, and results concerning its extinction probability and limiting evolution will be established. As illustration, some simulated examples will be given.

References

1. P. Haccou, P. Jagers and V. Vatutin. (2005). Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press.
2. D. M. Hull. (2003). A survey of the literature associated with the bisexual Galton–Watson branching process. Extracta Mathematicae, 18, 321-343.

288 Coherent forecasting for integer-valued autoregressive processes with periodic structure

[PS 3,(page 29)]

Magda MONTEIRO, *Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, Portugal*
Manuel SCOTTO, *Departamento de Matemática, Universidade de Aveiro, Portugal*
Isabel PEREIRA, *Departamento de Matemática, Universidade de Aveiro, Portugal*

The aim of this talk is to discuss forecasting methods in the context of the periodic integer-valued autoregressive process of order one with period T, defined by the recursive equation

$$X_t = \phi_t \circ X_{t-1} + Z_t,\ t \geq 1,$$

being $(Z_t)_{t\in N}$ a periodic sequence of independent Poisson-distributed random variables with mean $\upsilon_t = \lambda_j$ for $t = j + kT, (j = 1, \dots, T,\ k \in N_0)$, which are assumed to be independent of X_{t-1} and $\phi_t \circ X_{t-1}$, and $\phi_t = \alpha_j \in (0, 1)$ for $t = j+kT, (j = 1, \dots, T,\ k \in N_0)$, where the *thinning* operator $\circ$ is defined as

$$\phi_t \circ X_{t-1} \overset{d}{=} \sum_{i=1}^{X_{t-1}} U_{i,t}(\phi_t),$$

being $(U_{i,t}(\phi_t))$, for $i = 1, 2, \dots$, a periodic sequence of independent Bernoulli random variables with success probability $P(U_{i,t}(\phi_t) = 1) = \phi_t$.

Methods for generating coherent predictions are discussed in detail.

289 Estimating an earthquake occurrence probabilities by the semi-Markov model in Zagros Fold-Thrust Belt, Iran

[PS 2,(page 17)]

H. MOSTAFAEI, *Department of Statistics, The Islamic Azad University, North Tehran Branch.*
S. KORDNOURIE, *Department of statistics ,The Islamic Azad University,North Tehran Branch.Iran*

Abstract A Semi-Markov model is a stochastic model that can be used for estimating an earthquake occurrence probability. By this model, we can examine the great earthquake occurrences in three dimensions of space, time and magnitude. We can apply this model in a discontinuity that has a same structure and assume that the successive earthquakes are dependent events that are influenced by the elapsed time interval between them. The Zagros fold-thrust belt, in which more than fifty percent of Iran

290 GARCH and COGARCH: On convergence and statistical equivalence

[CS 8,(page 11)]

Gernot MUELLER, *Munich University of Technology, Germany*
Boris BUCHMANN, *Monash University, Melbourne, Australia*
Ross MALLER, *Australian National University, Canberra, Australia*
Alex SZIMAYER, *Fraunhofer ITWM, Kaiserslautern, Germany*

Nelson (1990) extended the discrete-time GARCH model to a continuous-time version, which turned out to be a diffusion limit driven by two independent Brownian motions. Klüppelberg, Lindner and Maller (2004) introduced a continuous-time version of the GARCH model, called COGARCH, which is constructed directly from a single background driving Lévy process.

We show that there exists a sequence of discrete-time GARCH models which converges to the COGARCH model in a strong sense (in probability, in the Skorohod metric). One can use this result, for example, to construct a pseudo-maximum-likelihood procedure for the estimation of the COGARCH parameters. Finally, we investigate whether GARCH and COGARCH are statistically equivalent in terms of Le Cam's deficiency distance.

References

1. Buchmann, B., and Müller, G. (2008). On the limit experiments of randomly thinned GARCH(1,1) in deficiency. Preprint. Monash University Melbourne and Munich University of Technology.
2. Klüppelberg, C., Lindner, A., and Maller, R. (2004). A continuous-time GARCH process driven by a Lévy process: Stationarity and second-order behaviour. *J. Appl. Probab.* **41** 601-622.
3. Maller, R., Müller, G., and Szimayer, A. (2008). GARCH Modelling in Continuous Time for Irregularly Spaced Time Series Data. *Bernoulli*, to appear.
4. Nelson, D.B. (1990). ARCH models as diffusion approximations. *J. Econometrics* **45** 7-38.

291 Partially linear model selection by the bootstrap

[CS 46,(page 41)]

Samuel MUELLER, *School of Mathematics and Statistics F07, University of Sydney, Australia*
Celine VIAL, *IRMAR, ENSAI, CNRS, UEB, Campus de Ker Lann, Bruz, France*

The purpose of model selection is to choose one or more models α from $\mathcal{A}$ with specified desirable properties, where $\mathcal{A}$ denotes a set of partially linear regression models for the relationship between a response vector $\mathbf{y}$ and a design matrix $\mathbf{X}$; that is $\mathcal{A} = \{\alpha \subseteq \{1, \ldots, p\}$, such that $\alpha = \{\alpha_1, \alpha_2\}, \alpha_1 \cap \alpha_2 = \emptyset\}$ where the functional relationship is

$$y_i = \theta_{\alpha_2}^T \, x_{\alpha_2,i} + g\left(\mathbf{x}_{\alpha_1,i}^T\right) + \epsilon_i, \quad i = 1, \ldots, n.$$

θ_{α_2} denotes an unknown p_{α_2}-vector of parameters, g is an unknown function from $\mathcal{R}^{p_{\alpha_1}}$ to $\mathcal{R}$, the design matrix $\mathbf{X}_\alpha$ and the errors $\epsilon_\alpha = (\epsilon_{1\alpha}, \ldots, \epsilon_{n\alpha})^T$ are independent, and the $\epsilon_{i\alpha}$ have location zero and variance σ^2. We use least squares estimation for estimating the two unknowns θ_{α_2} and g in the partially linear model and propose a new approach to the selection of partially linear models which in the spirit of Shao (1996; JASA) is based on the conditional expected prediction square loss function which is estimated using the bootstrap. Due to the different speeds of convergence of the linear and the nonlinear parts, a key idea is to select each part separately. In the first step we select the nonlinear components using a 'm-out-of-n' residual bootstrap which ensures good properties for the nonparametric bootstrap estimator. The second step selects the linear components among the remaining explanatory variables and the non-zero parameters are selected based on a two level residual bootstrap. We show that the model selection procedure is consistent under some conditions and simulations show that it performs well in comparison to other selection criteria.

292 Efficient estimators for linear functionals in nonlinear regression with responses missing at random

[CS 45,(page 39)]

Ursula MUELLER-HARKNETT, *Texas A&M University*

We consider regression models with responses that are allowed to be missing at random. The models are semiparametric in the following sense: we assume a parametric (linear or nonlinear) model for the regression function but no parametric form for the distributions of the variables; we only assume that the errors have mean zero and are independent of the covariates. For estimating general expectations of functions of covariate and response we use an easy to implement weighted imputation estimator. The estimator is efficient in the sense of Hajek and Le Cam since it uses all model information.

More precisely, we use an empirical estimator based on appropriate estimators of conditional expectations given the covariate. The independence of covariates and errors is exploited by writing the conditional expectations as unconditional expectations, which can themselves now be estimated by empirical estimators, with estimators of the regression parameters plugged in. The mean zero constraint on the error distribution is exploited by adding suitable residual-based weights (adapting empirical likelihood ideas).

Our results give rise to new efficient estimators of smooth transformations of expectations such as covariances and the response variance. Estimation of the mean response, which is usually considered in the literature, is discussed as a special (degenerate) case. The results are illustrated with computer simulations.

293 Quantum quadratic stochastic processes and related Markov processes

[CS 27,(page 26)]

Farrukh MUKHAMEDOV, *Faculty of Science, International Islamic University Malaysia, P.O. Box, 141, 25710, Kuantan, Malaysia*

It is known that the theory of Markov processes is a rapidly developing field with numerous applications to many branches of mathematics and physics. However, there are physical systems that are not described by Markov processes. One of such systems is described by quantum quadratic stochastic processes (q.q.s.p.). Such processes model systems with interacting, competing species and received considerable attention in the fields of biology, ecology, mathematics. In [1] we have defined a Markov processes associated with q.q.s.p. But there is a natural question can any Markov process define a q.q.s.p. so that the associated Markov process with q.q.s.p. will be the given one? In this report we present a condition for Markov processes to define the given q.q.s.p.

Reference

1. Mukhamedov F.M. On decomposition of quantum quadratic stochastic processes into layer-Markov processes defined on von Neumann algebras, *Izvestiya Math.* **68**(2004), 1009-1024

294 Optimization in a multivariate generalized linear model situation

[CS 35,(page 33)]

Siuli MUKHOPADHYAY, *Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076 India*

Andre I. KHURI, *Department of Statistics, University of Florida Gainesville, Florida 32611 USA*

One of the primary objectives in response surface methodology is the determination of operating conditions on a set of control variables that result in an optimum response. Optimization is more complex in a multiresponse situation as it requires finding the settings of the control variables that yield optimal, or near optimal, values for all the responses considered simultaneously. Several approaches dealing with multiresponse optimization in the case of linear models, where the responses are assumed to be continuous with uncorrelated errors and homogeneous error variances, are available in the literature. However, clinical or epidemiological data, for example, quite frequently do not satisfy these assumptions. For example, data on human responses tend to be more variable than is expected under the homogeneous error variances, biological data are correlated due to genetic relationships, and dose-response experiments yield discrete data. In such situations, analysis using generalized linear models (GLMs) is quite effective. The purpose of this talk is to discuss optimization in multivariate GLMs. Since optimal conditions for one mean response may be far from optimal or be even physically impractical for the others, we resort to finding compromise conditions on the input variables that are favorable to all the mean responses in a GLM environment. The deviation from the ideal optimum is measured by a distance function expressed in terms of the estimated mean responses along with their variance-covariance matrix. By minimizing such a distance function we arrive at a set of conditions for a compromise simultaneous optimum". An application of the proposed methodology is presented in the special case of a bivariate binary distribution resulting from a drug testing experiment concerning two responses, namely, efficacy and toxicity of a particular drug combination. This optimization procedure is used to find the dose levels of two drugs that simultaneously maximize their therapeutic effect and minimize their toxic effect.

295 On bivariate Mittag-Leffler distribution

[CS 41,(page 37)]

Davis Antony MUNDASSERY, *Christ college*

DR. K. JAYAKUMAR, *University of Calicut*

Due to the memory less property and constant hazard rate, exponential distribution has been received considerable importance in renewal theory and reliability contexts, especially in life testing. However, there are numerous situations where we encounter more heavy tailed data in which the exponential distribution is unfit. Pillai (1990) introduced Mittag-Leffler distribution as a generalization to exponential distribution. A random variable X is said to follow Mittag-Leffler distribution if its distribution function is

$$F_\alpha(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1} x^{\alpha k}}{\Gamma(1+\alpha k)}, \quad 0 < \alpha \leq 1, \ \ x \geq 0.$$

We get the exponential distribution function when $\alpha = 1$. The Mittag-Leffler distribution, being heavy tailed, is an adequate tool to model most of the data in economics and finance. Jayakumar (2003) used Mittag-Leffler distribution to model the rate of flow of water in Kallada river in Kerala, India.

Probability distribution of random sums of independently and identically distributed random variables especially geometric sums is an intensive area of research during the last decade (see, Gnedenko and Korolev (1996), Kozubowski and Panorska (1999)). The Mittag-Leffler distribution is closed under geometric summation. Kozubowski and Rachev (1994) used geometric sums to model the foreign currency exchange rate data. The present work is focused on introducing bivariate Mittag-Leffler forms of many important bivariate exponential distributions, using geometric summation. In this connection, we consider four candidates from the family of bivariate exponential distributions.

Moran (1967) introduced a bivariate exponential distribution which was later popularized by Downton (1970) as a model to describe the failure time of a two component system. A random vector(X, Y) is said to follow Moran's bivariate exponential distribution if the joint density function is

$$f(x,y) = \frac{\mu_1\mu_2}{1-\theta} I_0\left(\frac{2\sqrt{(\mu_1\mu_2\theta xy)}}{1-\theta}\right) exp-\left(\frac{\mu_1 x+\mu_2 y}{1-\theta}\right),$$

where $\mu_1, \mu_2 > 0$; $x, y > 0$; $0 \leq \theta \leq 1$ and $I_0(z) = \sum_{j=0}^{\infty}\left(\frac{z}{2j!}\right)^{2j}$ is the modified Bessel function of the first kind of order zero. In this paper special attention is paid to study the distributional properties of bivariate Mittag-Leffler form of Moran's bivariate exponential distribution. It is also devoted to introduce the bivariate Mittag-Leffler forms of Marshall-Olkin's (1967) bivariate exponential distribution, Hawkes' (1972) bivariate exponential distribution and Paulson's (1973)bivariate exponential distribution.

References

1. Downton, F. (1970) Bivariate exponential distributions in reliability theory. *Journal of Royal Statistical Society* **B 32**, 408-417.
2. Gnedenko, B. V. and Korolev, V (1996) *Random summation: Limit theorems and applications.* CRC Press, New York.
3. Hawkes, A. G. (1972) A bivariate exponential distribution with applications to reliability. *Journal of Royal Statistical Society* **B 34**, 129-131.
4. Jayakumar, K. (2003) Mittag-Leffler Processes. *Mathematical and Computer Modeling*, **37**, 1427-1434.
5. Kozubowski, T. J. and Panorska, A. K. (1999) Simulation of geometric stable and other limiting multivariate distributions arising in random summation scheme. *Mathematical and Computer Modelling*, **29**, 255-262.
6. Kozubowski, T. J. and Rachev, S. T. (1994) The theory of geometric stable laws and its use in modeling financial data. *European Journal of Operations Research*, **74**, 310-324.
7. Marshall, A. W. and Olkin, I. (1967) A Multivariate exponential distribution. *Journal of American Statistical Association*, **62**, 30-44.
8. Moran, P. A. P. (1967) Testing for correlation between non negative variates. *Biometrika*, **54**, 385-394.
9. Paulson, A. S. (1973) A characterization of the exponential distribution and a bivariate exponential distribution. *Sankhya* **A 35**, 69-78.
10. Pillai, R.N. (1990) On Mittag-Leffler functions and related distributions. *Annals of theInstitute of Statistical Mathematics*, **42**, 157-161.

296 Suitable hyper-parameters are applied to penalized maximum likelihood estimation in generalized extreme value distribution

[PS 1,(page 4)]

Md. Sharwar MURSHED, *Ph.D. Student, Department of Statistics, Chonnam National University, Gwangju 500-757, South Korea.*

Jeong Soo PARK, *Professor, Department of Statistics, Chonnam National University, Gwangju 500-757, South Korea*

To characterize the extreme behavior of a process we introduce generalized extreme value distribution as a model in this study. Maximum likelihood method has appeared as a significant and flexible modeling tool up to now in extreme value analysis but the disadvantage of this method is to make an agreement with shorter sample and causes large bias. This limitation has already been solved by imposing penalty function to the maximum likelihood equation. Data adaptive estimation procedure is exposed here to select a couple of appropriate hyper-parameter for each of the given data sets. We showed here that instead of selecting an overall hyper-parameter to the penalty function it is superior to go for a suitable chosen hyper-parameter for each of the given data sets and in addition comparing the performance of the penalty functions as well as with L-moment estimation method. The above method is verified in a simulation study especially with the bootstrap approach and in application to some of the rainfall data sets of South Korea.

297 Central limit results with application to regression with slowly varying regressors

[CS 37,(page 33)]

Kairat T. MYNBAEV, *Kazakh-British Technical University, Almaty, Kazakhstan*

In a recent paper by P.C.B. Phillips, some central limit results for weighted sums of linear processes have been developed to accommodate linear and nonlinear regressions with slowly varying regressors. Here we show that standardized slowly varying regressors are a special case of Lp-approximable sequences introduced in 2001 by K.T. Mynbaev. This fact allows us to generalize the central limit theorems due to Phillips in two directions: the requirements to the regressors are relaxed and the class of linear processes is widened to encompass short-memory processes. Further, it is shown that in case of a linear regression with two slowly varying regressors the asymptotic behavior of the OLS estimator is more complex than envisioned by Phillips.

298 Pathwise uniqueness for stochastic heat equations with Hölder continuous coefficients

[IS 28,(page 36)]

Leonid MYTNIK, *Technion — Israel Institute of Technology*
Edwin PERKINS, *The University of British Columbia*

We describe results of joint work with Edwin Perkins on pathwise uniqueness for solutions of stochastic heat equation

$$\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\Delta u(t,x)dt + \sigma(u(t,x))\dot{W}(x,t), \quad t \geq 0, \ x \in R,$$

where $\dot{W}$ is space-time white noise on $R \times R_+$. We show that the pathwise uniqueness holds for solutions of the above equation if σ is Hölder continuous of index $\gamma > 3/4$.

299 Stochastic search variable selection method with anatomical prior to detect brain activation from functional MRI data

[CS 47,(page 42)]

Rajesh Ranjan NANDY, *Departments of Biostatistics and Psychology, University of California, Los Angeles*

Functional MRI (fMRI) has been the most important non-invasive tool to analyze functions of human brain in the last decade. The conventional approach to the analysis of fMRI data is characterized by a two step process. The first step applies a statistical model (usually univariate) to the temporal response from each voxel from the stimulus and produces a test statistic to summarize the effect. In the second step test statistics are thresholded based on some pre-assigned significance level corrected for multiple comparison, and the voxels that meet or exceed the threshold are declared active. This formulation, though informative, performs these steps as independent processes. In other words, the statistical model is run independent of the thresholding process. We propose to unite these steps under a single coherent framework. This goal is achieved by exploiting characteristics of the Bayesian variable selection technique known as Stochastic Search Variable Selection (SSVS). The primary appeal of this framework is the possibility of inclusion of prior information on hypothesized effect. This approach also avoids sensitivity and specificity issues introduced by the thresholding. By incorporating the threshold within the model itself, activation maps are obtained that provide sharp contrast between active and non-active regions without an explicit thresholding based on significance values. Furthermore, with this approach two kinds of anatomical information may be incorporated in the prior which greatly reduces or eliminate false activation. First, since brain activation can only occur in gray matter voxels, the probability of a particular voxel being gray matter can be easily be included in the prior which will essentially eliminate false activations from voxels in white matter and CSF. Second, in many situations, the investigator has prior knowledge (possibly from other anatomic studies) about expected region of activation. It is a common practice in these situations to put a mask on the region of interest and restrict the analysis to that region. However, the proposed method permits us to simply put a stronger prior on the region of interest rather than excluding voxels outside the region of interest. Unlike conventional methods, this approach not only detects activation in the region of interest, it also can detect activation outside the region of interest provided the activation is strong enough. Finally, a detailed comparison of the proposed method with existing methods such as Statistical Parametric Mapping (SPM) will be provided.

300 On profiles of random search trees

[IS 13,(page 30)]

Ralph NEININGER, *J.W. Goethe University Frankfurt a.M.*

The profile of a search tree is the vector of the numbers of nodes with distance $0, 1, 2, \ldots$ from the root of the tree. The profile is a shape parameter that yields fine information on the complexities of various algorithms on search trees. In such performance studies of search trees the trees grow as random data are successively inserted into the tree, hence the trees are then random and the profile becomes a stochastic process. Recent results on the asymptotic behavior of profiles of various random search trees are discussed in this talk, including functional limit laws.

301 Statistics and data mining

[CS 40,(page 37)]

Hamid NILSAZ DEZFOULI, *Faculty Member of Islamic Azad University, Mahshahr , Iran*

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. The first and simplest analytical step in data mining is to describe the data summarize its statistical attributes(such as means and standard deviations), visually review it using charts and graphs, and look for potentially meaningful links among variables. The main activities of data mining are description and visualization, classification, estimation, prediction, clustering, and affinity grouping. So the field of data mining, like statistics, concerns itself with l̈earning from data.: In this article we will look at the connection between data mining and statistics and some of the major tools used in data mining.

302 A local-time correspondence for stochastic partial differential equations

[IS 28,(page 36)]

Eulalia NUALART, *University of Paris 13*
Davar KHOSHNEVISAN, *University of Utah*
Mohammud FOONDUN, *University of Utah*

It is frequently the case that a white-noise-driven parabolic and/or hyperbolic stochastic partial differential equation (SPDE) can have random-field solutions only in spatial dimension one. Here we show that in many cases, where the "spatial operator" is the L^2-generator of a Lévy process X, a linear SPDE has a random-field solution if and only if the symmetrization of X possesses local times. This result gives a probabilistic reason for the lack of existence of random-field solutions in dimensions strictly bigger than one. In addition, we prove that the solution to the SPDE is [Hölder] continuous in its spatial variable if and only if the said local time is [Hölder] continuous in its spatial variable. We also produce examples where the random-field solution exists, but is almost surely unbounded in every open subset of space-time. Our results are based on first establishing a quasi-isometry between the linear L^2-space of the weak solutions of a family of linear SPDEs, on one hand, and the Dirichlet space generated by the symmetrization of X, on the other hand. We study mainly linear equations in order to present the local-time correspondence at a modest technical level. However, some of our work has consequences for nonlinear SPDEs as well. We demonstrate this assertion by studying a family of parabolic SPDEs that have additive nonlinearities. For those equations we prove that if the linearized problem has a random-field solution, then so does the nonlinear SPDE. Moreover, the solution to the linearized equation is [Hölder] continuous if and only if the solution to the nonlinear equation is. And the solutions are bounded and unbounded together as well. Finally, we prove that in the cases that the solutions are unbounded, they almost surely blow up at exactly the same points.

303 Modeling distribution extreme value on multivariate binary response using mle and quasi mle

[PS 1,(page 4)]

Jaka NUGRAHA, *Islamic University of Indonesia Yogyakarta, Indonesia*
Suryo GURITNO, *Gadjah Mada University, Indonesia*
Sri Haryatmi KARTIKO, *Gadjah Mada University, Yogyakarta, Indonesia*

In this paper, we discuss binary multivariate response modeling based on extreme value distribution. Independent variables used in these models are some attributes of the alternative (labeled Zijt) and some attributes of the decision maker (labeled Xi). We assumed that n the decision maker observed with T response. Yit is tnd response variables from decision maker i and value Yit is binary. Response of decision maker i can be expressed as Yi = (Yi1,....,YiT), in which random variable Yit = 1 at decision maker i response variables t choosing alternative 1 and Yit =0 at decision maker i response variables t choosing alternative 0. In each of the decision maker, we have data (Yi, Xi, Zi). Models are derived by the assumption that maximum random utility which the decision maker i choose one of the alternatives having greatest utility. Methods of parameter estimation are Maximum Likelihood Estimator (MLE) method and quasi MLE method (that is Generalized Estimating Equation (GEE)). First discussion in this study is the estimation by MLE with independent assumption among response and then the MLE estimation using joint distribution by Bahadur

304 Climate past, climate present and climate future: a tale from a statistician.

[Public Lecture,(page 40)]

Douglas NYCHKA, *National Center for Atmospheric Research*

A grand scientific challenge for this century is to understand the complex interrelationships among the physical processes and human activities that define the Earth's climate. One specific concern is the warming of our climate brought about by the increase of greenhouse gases, such as carbon dioxide, being released into the atmosphere. What do we know about the Earth's past climate? Is global warming over the last century real? What is a climate model and how is it used to understand changes in our future climate? In answering each of these each of these questions statistical science can play a role in quantifying the uncertainty in scientific conclusions, for combining different kinds of information and summarizing complex data.

305 Gaussian Process Emulation for Dynamic Computer Models

[IS 34,(page 10)]

Jeremy E. OAKLEY, *School of Mathematics and Statistics, The University of Sheffield, UK.*

We consider the problem of constructing fast statistical approximations of computationally expensive dynamic computer models. By dynamic computer model, we mean a model that iteratively applies the same function at each time step, using the outputs of the function at one time step as the inputs to the function at the next time step, to produce a time series output. Gaussian process emulators are used to provide a complete probability distribution of the model outputs given initial conditions and any external forcing inputs. We compare two approaches to constructing the emulator: emulating the single time-step function, and using dimension reduction methods on the model output.

306 Bayesian model choice for high dimensional data

[IS 11,(page 23)]

Anthony O'HAGAN, *University of Sheffield*

This presentation will review ideas of model choice from a Bayesian perspective, and will discuss their application in the context of high-dimensional data.

307 Random fields with singularities in spectrum

[CS 20,(page 21)]

Andriy OLENKO, *La Trobe University, Vic, 3086, Australia*

Let $\xi(x),\ x \in R^n$ be a real, measurable, mean-square continuous, homogeneous isotropic Gaussian random field with $E\xi(x) = 0,\ E\xi^2(x) = 1$ and isotropic spectral function $\Phi(\lambda),\ \lambda \geq 0$.

Let us denote

$$\tilde{b}^a(r) = D\left[\int_{R^n} f_{r,a}(|t|)\xi(t)dt\right]\ ;$$

$$f_{r,a}(|t|)$$

$$=\frac{1}{|t|^{\frac{n}{2}-1}}\int_0^\infty \lambda^{n/2}\frac{J_{\frac{n}{2}}(r(\lambda-a))}{(r(\lambda-a))^{n/2}}J_{\frac{n}{2}-1}(|t|\lambda)d\lambda, \quad |t| \neq r.$$

Abelian and Tauberian theorems linking the local behavior of the spectral function $\Phi(x)$ in arbitrary point $x = a$ and weighted integral functionals $\tilde{b}_a(r)$ of random fields are presented. The asymptotic behavior is described in terms of functions of the class OR. Representations of weight functions $f_{r,a}(|t|)$ as series are obtained and investigated. Examples are given.

The results generalize some properties of long-memory random fields which can be obtained if one chooses $a = 0$.

308 Functional mathematical skills for accountancy students: basis for en-

hanced curriculum

[CS 4,(page 8)]

Melanie Joyno ORIG, *Faculty Member College of Arts and Sciences University of Mindanao, Davao City, Philippines*

The study was conducted to determine the functional mathematical skills needed by the accountancy students. It also sought to determine the differences among the functional mathematical skills needed by the accountancy students as viewed by the students, teachers and the practitioners, to determine in what mathematical skills are the accountancy students proficient and weak and to determine the significant difference in the perception of students and teachers in the proficiency and weaknesses in mathematical skills of the accountancy students and to prepare a functional curriculum in mathematics for the accountancy students. A list of mathematical skills was submitted to two practicing accountants and a mathematician for validation and was refined into a questionnaire. The questionnaires were sent to the 120 senior accountancy students, 15 full time accounting professors currently teaching at the College of Accountancy of the University of Mindanao and the 93 Certified Public Accountants. Tables were prepared showing the ranking of the functional mathematical skills useful to the accounting students as perceived by the students, teachers and practitioners. The most useful skills cited by the respondents are the basic mathematical concepts and operations and the least useful are the topics in trigonometry. Based on the findings, accountancy students need 7 subjects in mathematics namely preparatory

309 Maximum equality parameter estimation for regression models

[CS 5,(page 8)]

Ozer OZDEMIR, *Anadolu University*
Senay ASMA, *Anadolu University*

The aim of this study is to determine the parameter estimations of various regression models by using maximum equality (ME) principle. The ME estimators are summarized in a table which can be used as a successful guide. Furthermore, the obtained parameters are applied on a real life economical data.

310 Markov regenerative stochastic Petri net (MRSPN) representations of the M/G/1 retrial queueing system with a finite population and with server

[CS 13,(page 13)]

Lakshmi P, *Lecturer Kalasalingam University, India*
Kasturi RAMANATH, *Reader, School of Mathematics, Madurai Kamaraj University, India*

In this paper we have obtained Markov regenerative stochastic Petri net of an M/G/1 retrial queueing system with a finite population and with server vacations. We have derived closed form analytical expressions for the local and global kernels of the system. We have illustrated with the help of simple numerical examples the method by which the above representations can be used to obtained useful performance measures of the queueing system considered.

311 Modifications on Rescaling bootstrap in case of rare and illusive population

[CS 79,(page 58)]

Sanghamitra PAL, *River Research Institute, West-Bengal, India*

In estimating the nonlinear statistics we may employ Bootstrap technique. Rao and Wu (1988) have given their Rescaling bootstrap technique to find bootstrap estimators for nonlinear functions of finite population totals of several real variables. But their Rescaling bootstrap technique cannot be used to construct confidence intervals of non-linear statistics if the sample is chosen by varying probability sampling scheme for which the sample size varies.

In this paper to deal with rare population, we employ the technique of Adaptive sampling defining appropriate (1)'neighbourhoods' and (2) 'networks'. One may refer to Thompson (1992), Thompson and Seber (1996) and Chaudhuri (2000), Chaudhuri and Pal (2002) for a discussion on adaptive sampling technique. In this technique in order to capture more units beyond the 'initial sample' accommodating the rare commodities we consider adaptive sampling procedure. We modify rescaling bootstrapmethod to cover situations of rare populations.

The resulting relative performances of the alternative estimators noted above based on 'initial' and 'adaptive' samples are numerically examined through a simulation exercise utilizing known values.

312 Exit problem of a two-dimensional risk process from the quadrant: exact and asymptotic results

[CS 20,(page 21)]

Zbigniew PALMOWSKI, *Wrocław University, Poland*
Florin AVRAM, *Universite de Pau, France*
Martijn PISTORIUS, *King's College London, UK*

Consider two insurance companies (or two branches of the same company) that divide between them both claims and premia in some specified proportions. We model the occurrence of claims according to a renewal process. One ruin problem considered is that of the corresponding two-dimensional risk process first leaving the positive quadrant; another is that of entering the negative quadrant. When the claims arrive according to a Poisson process we obtain a closed form expression for the ultimate ruin probability. In the general case we analyze the asymptotics of the ruin probability when the initial reserves of both companies tend to infinity under a Cramér light-tail assumption on the claim size distribution. In the proof we use new exact asymptotics of the finite time ruin probability when the time horizon and the level go to infinity in fixed proportion for a general Lévy process that admits exponential moments. The talk is based on the following papers:

1. Avram, F., Palmowski, Z. and Pistorius, M. (2008) Exit problem of a two-dimensional risk process from a cone: exact and asymptotic results. To appear in *Annals of Applied Probability.*

2. Avram, F., Palmowski, Z. and Pistorius, M. (2008) A two-dimensional ruin problem on the positive quadrant. To appear in *Insurance: Mathematics and Economics.*

3. Palmowski, Z. and Pistorius, M. (2008) Cramér asymptotics for finite time first passage probabilities for general Lévy processes. Submitted for publication.

313 Estimating error-correction model for panel cointegration

[CS 67,(page 50)]

Jiazhu PAN, *Department of Statistics and Modelling Science, University of Strathclyde, UK*

Important advances on dynamic models for panel data have been made in recent years. An efficient approach to panel cointegration is Error-Correction Model (ECM). This paper develops an inferential theory for panel error-correction model (PECM) in both homogenous case and heterogenous case. General high-order dynamic PECM with or without deterministic terms are considered. The focus is determination of cointegration rank (r) and estimation of cointegrating vectors, which is an attractive issue in the rapidly growing literature on panel cointegration analysis. We first establish the rate of convergence and the limiting distributions for estimates of cointegrating vectors, and then introduce an consistent estimator of r, based on the proposed penalized goodness-of-fit criterion. The theory is developed under a framework of large cross sections (N) and large time dimension (T).

314 On RandomTomography without Angular Information Arising in Structural Biology

[CS 75,(page 56)]

Victor PANARETOS, *Ecole Polytechnique Fédérale de Lausanne*

What can be said about an unknown density function in $\mathbf{R}^n$ given a finite collection of $(n-1)$-dimensional marginals at random and unknown orientations? This question arises in single particle electron microscopy, a powerful method that biophysicists employ to learn about the structure of biological macromolecules. The method images unconstrained particles, as opposed to particles fixed on a lattice (crystallography) and poses a variety of statistical problems. We formulate and study statistically one such problem, namely the estimation of a structural model for a biological particle given random projections of its Coulomb potential density, observed through the electron microscope. Although unidentifiable (ill-posed), this problem can be seen to be amenable to a statistical solution, once parametric assumptions are imposed. It can also be seen to present challenges both from a data analysis point of view (e.g. uncertainty estimation and presentation) as well as computationally.

315 Stability of the Gibbs sampler for Bayesian hierarchical models

[IS 10,(page 18)]

Omiros PAPASPILIOPOULOS, *Universitat Pompeu Fabra*

We characterise the convergence of the Gibbs sampler which samples from the joint posterior distribution of parameters and missing data in hierarchical linear models with arbitrary symmetric error distributions. We show that the convergence can be uniform, geometric or sub-geometric depending on the relative tail behaviour of the error distributions, and on the parametrisation chosen. Our theory is applied to characterise the convergence of the Gibbs sampler on latent Gaussian process models. We indicate how the theoretical framework we introduce will be useful in analyzing more complex models.

316 Determining the effect of a rapid response team on hospital-wide mortality and code rates in a childrens hospital

[CS 81,(page 58)]

Layla PARAST, *Harvard University Department of Biostatistics*
Paul J. SHAREK, *Department of Pediatrics, Stanford University School of Medicine*
Stephen J. ROTH,
Kit LEONG,

Autoregressive integrated moving average models with a 12-month seasonality effect were implemented to determine the effect on hospital-wide mortality rates and code rates outside of the ICU setting after a rapid response team implementation at an academic childrens hospital.

317 Goodness-of-fit tests for multiplicative models with dependent data

[CS 17,(page 19)]

Juan Carlos PARDO-FERNÁNDEZ, *Departamento de Estatística e IO. Universidade de Vigo (Spain)*
Holger DETTE, *Fakultät für Mathematik. Ruhr-Universität Bochum (Germany)*
Ingrid VAN KEILEGOM, *Universite catholique de Louvain*

Several classical time series models can be written as a regression model of the form $Y_t = m(X_t) + \sigma(X_t)\varepsilon_t$, where (X_t, Y_t), $t = 0, 1, 2, \ldots$, is a bivariate strictly stationary process. Some of those models, such as ARCH models, share the property of proportionality of the regression function, m, and the scale function, σ. In this paper, we work in a nonparametric setup and derive a test for the null hypothesis $H_0 : m(\cdot) = c\sigma(\cdot)$, where c is a fixed positive value, in general unknown, versus the general alternative hypothesis $H_1 : m(\cdot) \neq c\sigma(\cdot)$.

In econometric and financial models some relationship might be expected between the return (here represented by the regression function, m) and the risk (or the scale function, σ), so the proportionality of m and σ is a feature of interest. In other contexts, the problem of estimating and testing the regression function under the assumption of a constant coefficient of variation also corresponds to the situation described above.

The proposed testing procedure is based on a comparison of two empirical processes of the standardized nonparametric residuals calculated under the hypothesis of a multiplicative structure and the alternative of a general nonparametric regression model. Some asymptotic results establishing weak convergence of the difference of the processes and the asymptotic distributions of the tests statistics are stated. The asymptotic distributions of the proposed test statistics are complicated and difficult to use in practice. Thus we describe a consistent bootstrap procedure to approximate the critical values of the test and present the results of a small simulation study which illustrates the finite sample properties of the bootstrap version of the test.

318 Semiparametric efficient estimation in partially linear additive models with unknown error density

[IS 19,(page 35)]

Byeong U. PARK, *Seoul National University*
Kyusang YU, *University of Mannheim*
Enno MAMMEN, *University of Mannheim*

We discuss asymptotically efficient estimation in partially linear additive models, where the additive function and the error density are infinite dimensional nuisance parameters. This model has wider application than fully nonparametric additive models since it may accommodate categorical random variables as regressors. Also, the additivity of the nonparametric function allows one to avoid the curse of dimensionality that one suffers from with partially linear models. Two main issues are to ensure the additive structure in fitting the nonparametric part of the regression function, and to adapt the estimation procedure to the unknown error density. We apply nonparametric additive regression techniques and use an estimator of the efficient influence function to obtain semiparametric efficient estimators of the finite-dimensional parameters in the parametric part of the regression function. We also discuss the finite sample properties of the estimators from a numerical study.

319 Persistent threshold-GARCH processes : An introduction

[PS 2,(page 16)]

Jinah. A PARK, *Sookmyung women's university*
Sun Young HWANG, *Sookmyung women's university*
Jisun S. BAEK, *Sookmyung women's university*
Moonsun S. CHOI, *Sookmyung women's university*

Over the decade, there has been a growing interest in nonlinear modeling for conditionally heteroscedastic time series. In particular, as a natural extension of the standard ARCH/GARCH models of Engle(1982) and Bollerslev (1986), threshold ARCH/GARCH process discussed by Li and Li(1996) and Hwang and Basawa(2004) provided an asymmetric account for the volatility (conditional variance) in such a way that positive and negative shocks have different effects on volatility. This article introduces near non-stationary model, say, integrated threshold-GARCH time series where a shock to the current volatility will remain for a long time and thus we have a persistent effect of a current shock on the future volatilities. Some probabilistic structures of the integrated threshold-GARCH process are investigated. Also, real data analysis is illustrated using financial series in KoreaOver the decade, there has been a growing interest in nonlinear modeling for conditionally heteroscedastic time series. In particular, as a natural extension of the standard ARCH/GARCH models of Engl(1982) and Bollerslev(1986), threshold ARCH/GARCH process discussed by Li and Li(1996) and Hwang and Basawa(2004) provided an asymmetric account for the volatility(conditional variance) in such a way that positive and negative shocks have different effects on volatility. This article introduces near non-stationary model, say, integrated threshold-GARCH time series where a shock to the current volatility will remain for a long time and thus we have a persistent effect of a current shock on the future volatilities. Some probabilistic structures of the integrated threshold-GARCH process are investigated. Also, real data analysis is illustrated using financial series in Korea.

320 Predictions in ARMA-GARCH models via transformation and back-transformation approach

[PS 3,(page 29)]

Juyeon PARK, *Department of Statistics, Sookmyung Women's University, Seoul, Korea*
In-Kwon YEO, *Department of Statistics, Sookmyung Women's University, Seoul, Korea*

One of main aspects of time series analysis is to forecast future values of series based on values up to a given time. The prediction is usually performed under the normality assumption. When the assumption is seriously violated, a transformation of data may permit the valid use of the normal theory. We investigate the prediction problem for future values in the original scale when transformations are applied in ARMA-GARCH models. The transformation and back-transformation approach is applied to obtain the prediction value and interval. We introduce a modified smearing estimation to provide a bias-reduced estimator for prediction value. Empirical studies on some financial indices are executed to compare the coverage probability and bias of existing methods and the proposed method.

321 Fast high-dimensional bayesian classification and clustering

[CS 11,(page 12)]

Vahid PARTOVI NIA, *Swiss Federal Institute of Technology Lausanne*
Anthony C. DAVISON, *Swiss Federal Institute of Technology Lausanne*

In this talk we introduce a hierarchical Bayesian mixture model applicable to high-dimensional continuous response data, for which explicit calculations are possible, and show how the model can be used for high-dimensional classification. An avenue from classification to clustering is built via an exchangeable prior for partitions, and a fast algorithm for clustering is proposed based on agglomerative method, using the log-posterior as a natural distance. A generalisation of the model which allows Bayesian variable selection for clustering is then proposed. The effectiveness of the new method is studied for simulated data and applications are given to replicated metabolomic and cDNA microarray data.

322 Stepdown multiple test procedure controlling the error rates in finite samples

[PS 2,(page 16)]

Dariusz PARYS, *University of Lodz*
Czeslaw DOMANSKI, *University of Lodz*

In this paper we present the stepdown methods that control the familywise error rate. We consider the case of finite samples. The methods which are connecting with stepwise character cannot always achieve strong control the error rates. We discuss the familywise error rate (FWE) and procedures which control it in a strong manner. We present the application of closure principle of Marcus et al. (1976) to create the algorithm of new stepdown method. Also we discuss how the asymptotic approach is connected with our inferences.

323 On the performance of the r-convex hull of a sample

[CS 23,(page 22)]

Beatriz PATEIRO-LOPEZ, *Universidade de Santiago de Compostela*

Alberto RODRIGUEZ-CASAL, *Universidade de Santiago*

The reconstruction of a compact set S in R^d from a finite set of points taken in it is an interesting problem that has been addressed in many fields like computational geometry. A large body of literature assumes that the set of interest is convex. In this paper a less restrictive assumption on the set we want to estimate is considered. It is assumed that S is r-convex, which means that a ball of radius r can roll freely outside the set. Under this assumption the r-convex hull of the sample S_n is a natural estimator. The performance of S_n is evaluated through the expectation of the distance in measure between the estimator and the target S. Its convergence rate is obtained. The expected number of extreme points of the estimator, that quantifies the degree of complexity of the set, is also provided.

324 Densities of composite Weibullized generalized gamma variables

[CS 76,(page 56)]

J PAUW, *University of Pretoria, South Africa*

JJJ ROUX, *University of Pretoria, South Africa*

A BEKKER, *University of Pretoria, South Africa*

The Weibullized generalized gamma distribution, which is a very flexible distribution due to its richness in parameter structure, is derived. In this paper we study densities of composite Weibullized generalized gamma variables.

325 Nonparametric testing for image symmetries

[CS 18,(page 20)]

Miroslaw PAWLAK, *University of Manitoba*

Hajo HOLZMANN, *Institute for Stochastics, University of Karlsruhe*

Symmetry plays an important role in image understanding and recognition. In fact, symmetric patterns are common in nature and man-made objects and the detection of an image symmetry can be useful for designing effcient algorithms for object recognition, robotic manipulation, image animation, and image compression. This paper formulates the problem of assessing reflection and rotations symmetries of an image function observed under an additive noise. Rigorous nonparametric statistical tests are developed for testing image invariance under reflections or under rotations through rational angles, as well as joint invariance under both reflections and rotations. The symmetry relations are expressed as restrictions for Fourier coefficients with respect to a class of radial orthogonal functions. Therefore, our test statistics are based on checking whether the estimated radial coefficients approximately satisfy those restrictions. We derive the asymptotic distribution of the test statistics under both the hypothesis of symmetry and under fixed alternatives. The former result is used to construct asymptotic level α tests for lack of symmetry, whereas the latter result can be used to estimate the power of these tests, or to construct tests for validating the approximate symmetry of the image. Our results model the performance of the tests on grids which become increasingly fine. The theoretical developments are based on the theory of the asymptotic behavior of quadratic forms of random variables.

326 Statistical inference of interspike intervals from short time windows

[CS 47,(page 42)]

Zbynek PAWLAS, *Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic*

Petr LANSKY, *Institute of Physiology, Academy of Sciences of the Czech Republic*

Statistical characteristics of interspike intervals are studied based on the simultaneous observation of spike counts in many independent short time windows. This scenario corresponds to the situation in which a target neuron occurs. It receives information from many neurons and has to respond within a short time interval. We are mainly interested in the estimation of statistical moments of interspike intervals. Not only mean but also coefficient of variation is estimated and the precision of the estimation procedures is examined. We consider two stationary models of neuronal activity – a renewal process with gamma distribution of interspike intervals and a doubly stochastic point process. Both, moment and maximum likelihood estimators are investigated. In accordance with our expectations, numerical studies confirm that the estimation of mean interspike interval is more reliable than the estimation of coefficient of variation. The error of estimation increases with increasing mean interspike interval which is equivalent to decreasing the size of window (less events are observed in a window) and with decreasing number of neurons (lower number of windows).

327 Bayesianity of the maximum likelihood estimator, under absolute error loss

[CS 42,(page 38)]

Amir T. PAYANDEH, *Department of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran*
Dan KUCEROVSKY, *Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada E3B 5A3*
Eric MARCHAND, *Departement de mathematiques, Universite de Sherbrooke, Sherbrooke, QC, Canada, J1K 2R1*
William E. STRAWDERMAN, *Department of Statistics, Rutgers University, 501 Hill Center, Busch Campus, Piscataway, N.J., USA*

We consider the problem of estimating, under absolute-value (L1) loss, the location parameter of a unimodal and symmetric distribution, when the parameter is bounded to an interval. After a brief review of related literature and the introduction of some useful mathematical tools, we present a development showing that the maximum likelihood estimator is proper Bayes, hence admissible, under absolute-value loss. This contrasts with the case of square-error (L2) loss, where the MLE is neither Bayes, nor admissible. This extends previously established results by Iwasa and Moritani for the normal case. Applications are given and we also discuss the case of a lower bound constraint.

328 Stochastic spatial Lotka-Volterra Models, superprocesses and pde's

[IS 30,(page 10)]

Ed PERKINS, *The University of British Columbia*
Ted COX, *Syracuse University*
Richard DURRETT, *Department of Mathematics, Cornell University*
Mathieu MERLE, *The University of British Columbia*

We study both high density and low density limit theorems for the spatial Lotka-Volterra model introduced by Neuhauser and Pacala. The former produce superprocess limits and the latter reaction-diffusion equations. In earlier work with Ted Cox, the former were used to show results on survival (in two and more dimensions) and coexistence (in three and more dimensions). In ongoing work with Ted Cox and Rick Durrett, a high density limit theorem is used to show these results are locally sharp in three and more dimensions. Time permitting, some ongoing work on coexistence in two dimensions will be described–the latter is joint also with Mathieu Merle.

329 Random sampling of long-memory stationary processes

[CS 39,(page 34)]

A. PHILIPPE, *Universite Nantes, Laboratoire de mathematiques Jean Leray*
M.C. VIANO, *Universite Lille 1*

The effect of random sampling on the memory of a stationary second order discrete time process is the subject of this talk. The main results concern a family of long memory processes including the so-called FARIMA (p, d, q) processes.

We start from $(X_n)_{n>0}$, a stationary discrete time second order process with covariance sequence $S_X(h)$ and a random walk $(T_n)_{n>0}$ independent of $(X_n)_{n>0}$. The sampling intervals $T_j - T_{j-1}$ are i.i.d. with a common distribution supported by the set of strictly positive integers. We consider the sampled process Y defined by

$$Y_n = X_{T_n}$$

We first show that short memory is always preserved. Then we present examples proving that even if it does not affect the intensity of long memory, deterministic sampling can affect the seasonal effects of this memory. For example the number of singular frequencies of the spectral density of the sampled process can be reduced.

The main results of the paper, concerning processes with covariances going arithmetically to zero, we show that the intensity of memory is preserved if the sampling intervals $T_1 - T_0$ belongs in L^1. In the other cases, the memory of the sampled process is reduced according to the largest finite moment of

T_1.

330 Bayesian inference for an impatient M—M—1 queue with balking

[CS 29,(page 27)]

Chandrasekhar PICHIKA, *Loyola College (Autonomous) Chennai 600034, India*

Assuming that the stationary distribution in an M—M—1 balking situation is Negative Binomial, maximum likelihood estimator (MLE) and Bayes estimator of the parameter p based on the number of observations present at several sampled time points are obtained. Further, the minimum posterior risk associated with Bayes estimator and minimum Bayes risk of the estimator are obtained

331 Degenerate diffusions in gene duplication models

[IS 14,(page 5)]

Lea POPOVIC, *Concordia University*
Richard DURRETT, *Department of Mathematics, Cornell University*

We consider two processes that have been used to study gene duplication, Watterson's double recessive null model, and Lynch and Force's subfunctionalization model. Though the state spaces of these diffusions are 2 and 6 dimensional respectively, we show in each case that the diffusion stays close to a curve. Using ideas of Katzenberger we show that the one dimensional projections of the processes converge to a diffusion on a curve, and we obtain asymptotics for the time to loss of one gene copy. As a corollary we find that the probability of subfunctionalization decreases exponentially fast as the population size increases, which rigorously confirms the result Ward and Durrett found by simulation that the likelihood of subfunctionalization for gene duplicates decays exponentially fast as the population size increases.

332 Description Length and Dimensionality Reduction in Functional Data Analysis

[CS 64,(page 49)]

D. S. POSKITT, *Department of Econometrics and Business Statistics, Monash University*
Arivalzahan SENGARAPILLAI, *Department of Econometrics and Business Statistics, Monash University*

In the analysis of functional data considerable advantages can be obtained by expressing each function in terms of a low dimensional, finite basis. In this paper we explore the consequences of using selection criteria based upon description length principles to select an appropriate number of such basis functions. As part of our analysis we provide a flexible definition of the dimension of a random function that depends on the signal-to-noise ratio derived from a signal-plus-noise type decomposition. The decomposition is constructed directly from the Karhunen–Loeve expansion of the process and does not depend on assuming that the observations actually consist of a signal embedded in noise, although such situations are encompassed as a special case. Our definition is closely related to ideas of variance decomposition and we show that the dimension chosen by the classical variance decomposition technique will be consistent for the true dimension of a function at a pre-assigned signal-to-noise ratio. Description length criteria do not require the practitioner to assign a signal-to-noise ratio. Nevertheless, we also show that description length criteria behave in a coherent manner and that in low noise settings they will produce consistent estimates of the true finite dimension of the signal. Two examples, taken from mass-spectroscopy and climatology, are used to illustrate the practical impact of the different methods and some simulations that demonstrate the workings of our theoretical results are also presented.

333 Two-dimensional Poisson point process in modeling of tectonic earthquake in Java and Bali

[CS 3,(page 7)]

Hasih PRATIWI, *Sebelas Maret University, Surakarta, Indonesia*
SUBANAR, *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Indonesia*
J.A.M. van der WEIDE, *Delft University of Technology, Delft, Netherlands*

Indonesia with its complicated tectonic setting, crossed by the boundaries of three tectonic plates is highly endangered by earthquakes. The shaking ground can cause buildings and bridges to collapse and disrupt gas, electric, and phone services. Especially in Indonesia earthquakes also trigger landslides and huge, destructive ocean waves called tsunami. This paper discusses the modeling of earthquake in Java and Bali using two-dimensional Poisson point process approach. The data consist of tectonic earthquake data in Java and Bali between the years 1966 and 2000. By considering truncated data and Gutenberg-Richter's Law, we conclude that the tectonic earthquake pattern in Java and Bali can be expressed as two-dimensional Poisson point process in time and magnitude. Mean number of tectonic earthquakes within five years is 17 events while mean of magnitude is 5.272 Richter scale. For magnitude bigger than 5.8 Richter scale, the model appropriates with the data.

334 A Bayesian synthesis of evidence for estimating HIV incidence among men who have sex with men in London

[CS 81,(page 59)]

A. M. PRESANIS, *Medical Research Council Biostatistics Unit, Cambridge, UK*
D. DE ANGELIS, *Statistics, Modelling and Bioinformatics Unit, Health Protection Agency Centre for Infections, London and Medical Research Council Biostatistics Unit, Cambridge, UK*
A. GOUBAR, *Institut de Veille Sanitaire, France*
A. E. ADES, *Department of Community Based Medicine, University of Bristol, UK*

The implementation and evaluation of public health policies aimed at prevention and control of epidemics rely crucially on knowledge of fundamental aspects of the disease of interest, such as prevalence and incidence. These are typically not directly measurable and, increasingly, are being estimated through the synthesis of diverse sources of evidence, particularly using a Bayesian framework [1,2].

The prevalence of Human Immunodeficiency Virus (HIV) in England and Wales is estimated annually [3,4] by dividing the general population into mutually exclusive risk groups g and regions r, and for each pair (g, r) and at a single time point t, synthesising surveillance and other data, D_t, to estimate: the proportion of the population in the group, $\rho_{t,g,r}$; HIV prevalence, $\pi_{t,g,r}$; and the proportion of infections which are diagnosed, $\delta_{t,g,r}$. Markov chain Monte Carlo (MCMC) is used to obtain the posterior distribution of each quantity.

Here we restrict attention to the men who have sex with men (g =MSM) group in London. By performing the above inference simultaneously over successive time points $1 \ldots T$, using data sets $D_t, t \in 1 \ldots T$, we may obtain yearly "snapshots" (with associated uncertainty) of the number of MSM in each of three compartments: uninfected; HIV-positive but undiagnosed; and diagnosed HIV-positive. Changes over time in the compartment sizes may be described by a system of differential equations. Use of further data on mortality, migration, aging and risk behaviour change allows estimation of the rates of transition between the states and, in particular, of the rate of movement between the uninfected and infected states, *i.e.* HIV incidence. This problem is analogous to the estimation of transition rates in a Markov multi-state model where the observed data are aggregate counts of the state occupancies at fixed times. Here, however, we simultaneously estimate the transition rates, the compartment sizes and $(\delta_{t,\mathrm{MSM}}, \pi_{t,\mathrm{MSM}}, \rho_{t,\mathrm{MSM}})$, given data sets $D_t, t \in 1 \ldots T$. MCMC is again employed and solutions to the system of equations are computed numerically at each iteration, hence obtaining the joint posterior distribution of the quantities of interest.

Furthermore, we consider parameterising HIV incidence in terms of HIV prevalence, both diagnosed and undiagnosed, and of the probabilities of contact and transmission given contact between the susceptible and infected states. In contrast to the largely deterministic mathematical models in the literature on infectious disease dynamics, this synthesis of evidence combines a dynamic transmission model with a statistical model, which fully and correctly propagates the uncertainty in the data sets D_t through to $(\delta_{t,\mathrm{MSM}}, \pi_{t,\mathrm{MSM}}, \rho_{t,\mathrm{MSM}})$, and hence through to HIV incidence.

References

1. Ades, A. E. and A. J. Sutton (2006): Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches, *J. R. Statist. Soc. A, 169,* 5-35

2. Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. van der Linde (2002): Bayesian measures of model complexity and fit, *J. R. Statist. Soc. B, 64,* 1-34

3. Presanis, A. M., D. De Angelis, D. J. Spiegelhalter, S. Seaman, A. Goubar and A. E. Ades (2008): Conflicting evidence in a Bayesian synthesis of surveillance data to estimate HIV prevalence, *J. R. Statist. Soc. A,* in press

4. Goubar, A., A. E. Ades, D. De Angelis, C. A. McGarrigle, C. H. Mercer, P. Tookey, K. Fenton and O. N. Gill (2008): Estimates of HIV prevalence and proportion diagnosed based on Bayesian multi-parameter synthesis of surveillance data, *J. R. Statist. Soc. A,* in press, with discussion

335 Palm likelihood for parameter estimation in inhomogeneous spatial cluster point processes

[CS 2,(page 6)]

Michaela PROKESOVA, *Charles University, Prague, Czech Republic*

The paper is concerned with parameter estimation for a class of inhomogeneous spatial cluster point processes which has the property of second-order intensity reweighted stationarity (Baddeley et al. 2000). Most common example of processes with this property are processes derived from homogeneous processes by location dependent thinning.

Our method is a generalization of the method for homogeneous point processes introduced in Tanaka et al. 2007. There the Poisson approximation to the likelihood of the homogeneous process (X-X) of the differences between the points of the original point process X was used to define the so called Palm likelihood (constructed from the Palm intensities), which is then used for estimation of the model parameters. The assumption of the homogeneity/stationarity of X is essential here for the Palm intensity being translation invariant and equivalent to the intensity of the process of the differences. Neverthless the second-order intensity reweighted stationarity enables a special form of desintegration of the second-order intensity function of X and thus generalization of the Palm likelihood estimation to the inhomogeneous case.

We propose a 2-step estimation procedure for the inhomogeneous cluster point processes, where in the first step we estimate the intensity parameters from the Poison likelihood score estimation function and in the second step we use the generalized Palm likelihood for estimation of the clustering parameters. The proposed method is compared with the existing estimation procedures for inhomogeneous cluster point processes (based on composite likelihood (Guan 2006) and minimum contrast estimation (Waagepetersen and Guan 2008)) and the use is illustrated by application to forest data.

References

1. Baddeley, A. J., Moller, J., Waagepetersen, R. (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns, Statist. Neerlandica, 54:329–350.
2. Guan, Y. (2006) A composite likelihood approach in fitting spatial point process models, J. Am. Stat. Assoc., 101:1502–1512.
3. Tanaka, U., Ogata, Y., Stoyan, D. (2007) Parameter Estimation and Model Selection for Neymann-Scott Point Processes, Biometrical Journal, 49:1–15.
4. Waagepetersen, R. P. and Guan, Y. (2008) Two-Step Estimation for Inhomogeneous Spatial Point Processes, J. R. Stat. Soc. B, submitted.

336 Sliced space-filling designs

[CS 43,(page 38)]

Zhiguang QIAN, *Department of Statistics University of Wisconsin-Madison*
Jeff WU, *Georgia Tech*

Design construction for computer experiments with qualitative and quantitative factors is an important but unsolved issue. In this work a general approach is proposed for constructing a new type of design called sliced space-filling design to accommodate these two types of factors. It starts with constructing a Latin hypercube design based on a special orthogonal array for the quantitative factors and then partition the design into groups corresponding to different level combinations of the qualitative factors. The points in each of these groups are guaranteed to have good space-filling properties in low dimensions.

337 Invariant measures for KdV

[IS 30,(page 10)]

Jeremy QUASTEL, *University of Toronto*

Gaussian white noise turns out to be invariant for the Kortweg-deVries equation on the circle. We will provide a little background on the equation, discuss what it means to solve KdV with distributional initial data, describe the proof that white noise is invariant, and speculate on the physical relevance of the resulting field. This is joint work with Benedek Valko.

338 Informatics Platform for Global Proteomic Profiling and Biomarker Discovery

[CS 9,(page 11)]

Dragan RADULOVICH, *Florida Atlantic University*

We have developed an integrated suite of computer algorithms, statistical methods and software applications to support large-scale liquid-chromatography-tandem-mass spectrometry-based shotgun profiling of complex protein mixtures. The programs automatically detects and quantifies large numbers of peptide peaks in feature-rich ion mass chromatograms, compensate for spurious fluctuations in peptide retention times and recorded signal intensities, and reliably match related peaks across different datasets. Application of this toolkit markedly facilitates pattern recognition and biomarker discovery in comparative global proteomic studies. We will present a short overview of our recent paper in Nature Genetics where we for the first time reduced in practice a large-scale protein mapping. We will describe the ramification of this result related to early detection and treatment of a disease. In particular we will describe our preliminary results related to Biomarkers discovery for classification and early detections of Leukemia in Humans.

339 Wavelet density estimation from contaminated data with repeated measurements

[CS 53,(page 44)]

Marc RAIMONDO, *Scool of Mathematics and Statistics, the University of Sydney*
Laurent CAVALIER, *Centre de Mathématiques et Informatique, Université Aix-Marseille 1*

We present an adaptive method for density estimation when the observations are contaminated by additive errors. The error distribution is not specified by the model but is estimated using repeated measurements. In this setting, we propose a wavelet method for density estimation which adapts both to the degree of ill-posedness of the problem (smoothness of the error distribution) and to the regularity of the target density. Our method is implemented in the Fourier domain via a square root transformation of the empirical characteristic function and yields fast translation invariant non-linear wavelet approximations with data driven choices of fine tuning parameters. For smooth error distributions we show that our proposal is near optimal over a wide range of density functions. When the observations are made without errors our method provides a natural implementation of direct density estimation in the Meyer wavelet basis. We illustrate the adaptiveness properties of our estimator with a range finite sample examples drawn from population with smooth and less smooth density function.

340 Goodness of fit for Auto-Copulas: testing the adequacy of time series models

[CS 16,(page 16)]

Pal RAKONCZAI, *Eotvos Lorand University, Probability Theory and Statistics Department,Budapest,Hungary*
Laszlo MARKUS, *Eötvös Loránd University, Budapest, Hungary*
Andras ZEMPLENI, *Eotvos Lorand University, Probability Theory and Statistics Department,Budapest,Hungary*

Copula models proved to be powerful tools in describing the interdependence structure of multivariate data sets. The advantage of using copulas lays in the fact that - unlike Pearson-correlations - they represent nonlinear dependencies as well, and make it possible to study the interdependence of high (or low) values of the variables. So, it is very natural to extend the use of copulas to the interdependence structure of time series. To the analogy of the autocorrelation function the use of auto-copulas for the lagged series can reveal the specifics of the dependence structure in a much finer way. Therefore the fit of the corresponding auto-copulas can provide an important tool for evaluating time series models. For making comparisons among competing models the fit of copulas has to be measured. Our suggested goodness of fit test is based on the probability integral transformation of the joint distribution, which reduces the multivariate problem to one dimension. We apply the proposed methods for investigating the auto-dependence of river flow time series with particular focus on the synchronised appearance of high values and we also consider how these methods could be improved by choosing different weights for more efficient detecting.

341 Measure-valued Process Limits for Many Server Systems

[IS 31,(page 51)]

Kavita RAMANAN, *Carnegie Mellon University*

We consider a system in which customers with independent and identically distributed service times from a general distribution arrive to a queue, and are served in the order of arrival. A Markovian representation of this process leads naturally to a measure-valued process. Under an asymptotic scaling that is of interest in many applications, we establish functional strong law and central limit theorems for these processes. The latter, in particular, gives rise to an SPDE, and the corresponding evolution of the number of customers in system is described by a functional stochastic differential equation with memory. Generalizations to include abandonments into the system are also considered. This is based on joint works with Weining Kang and Haya Kaspi.

342 Statistical analysis of equity indices- a probabilistic approach

[CS 78,(page 58)]

K.S.Madhava RAO, *Associate Professor, Department of Statistics, Univeristy of Botswana, Private Bag 00705, Gaborone,Botswana*
K.K. MOSEKI, *Lecturer, Department of Statistics, Univeristy of Botswana, Private Bag 00705, Gaborone,Botswana*

Forecasting volatility in stock indices and currency returns has been a major area of research in financial economics. A common approach to forecast the volatility uses the standard deviation of the returns of the last T trading days. Many traditional econometric methods forecast the conditional distribution of asset returns by a point prediction of volatility. The central contribution of this paper is that it suggests an alternative approach for modeling and related analysis of asset returns. In the new approach suggested here volatility in stock returns is classified into various states according to the predetermined perceptions of the market player. It is imperative that a fairly accurate knowledge of such states of volatility in asset returns will be of great help to a prospective or an existing market player. In this paper, we build a probability model for forecasting variations in the equity prices/indices and devise certain criteria to establish the significance of the empirically estimated parameters. As an application of the proposed model, we analyze Botswana stock market data from 1999-2005. The probability models are built based on weekly and monthly equity indices. The approach suggested here will be of interest to academicians, stock market investors and analysts.

343 Influence properties of partial least squares in measurement error models

[CS 7,(page 9)]

A. RASEKH, *Prof. of Statistics, Department of Statistics, Shahid Chamran University, Ahvaz, Iran*
F. MAZAREI, *Postgraduate Student, Department of Statistics, Shahid Chamran University, Ahvaz, Iran*

Partial least squares (PLS) regression is one of the most widely used chemometrical tools to estimate concentrations from measured spectra. In recent years partial least square regression has been extended to the measurement error models. As it is mostly a chemometrical tool, it has hitherto only been granted little attention in the statistical literature. One of these properties is the influence function (IF), which is of widespread use in the literature on robust and mathematical statistics. Indeed, one can define an estimator to be robust whenever its influence function is bounded, but also for non-robust, so called classical estimators (such as PLS), the influence function has major applicability. Serneel et al (2004) computed the influence function for partial least squares regression and used as a diagnostic tool to assess the influence of individual calibration samples on prediction. In this paper, we extend these results to the measurement error models and we derive the influence function for partial least squares regression in this context.

Sven Serneels, Christophe Croux, Pierre J. Van Espen (2004), Influence properties of partial least squares regression, Chemometrics and Intelligent Laboratory Systems, 71, 13

344 Marshall-Olkin q-Weibull distribution and autoregressive processes

[CS 28,(page 26)]

Shanoja RAVEENDRAN NAIK, *Centre for Mathematical Sciences, Pala campus, India*
Jose KANICHUKATTU KORAKUTTY, *Department of Statistics, St. Thomas College, Pala, Kerala-686574, India*

The Weibull distribution plays an important role in modeling survival and life time data. Recently various authors have introduced several q-type distributions such as q-exponential, q-Weibull, q-logistic and various pathway models in the context of information theory, statistical mechanics, reliability modeling etc. The q-Weibull distribution is a stretched model for Weibull distribution obtained by introducing a new pathway parameter q, which facilitates a slow transition to the Weibull as $q \to 1$. Picoli et al (2003) used the q-Weibull distribution in the context of modeling data on basketball baskets in a championship, tropical cyclone victims, brand name drugs by retail sales, highway lengths etc., and found that it is a more suitable model. Here we introduce the q-Weibull distribution as a special case of the pathway model of Mathai (2006). The q-Weibull distribution is characterised by heavy tailedness as well as cutoff points. The q-models are based on the relation

$$\mathrm{e}_q^{(-x)} \equiv \begin{cases} [1+(q-1)x]^{-\frac{1}{q-1}} \ ; & x \geq 0, q > 1 \\ [1-(1-q)x]^{\frac{1}{1-q}} ; & q < 1. \end{cases}$$

As $q \to 1$ the function approaches the standard exponential function. Here we made a detailed study of the properties of the q-Weibull distribution and applied it to a data on cancer remission times for which this distribution is a good model. Results relating to reliability properties, estimation of parameters and applications in stress-strength analysis are obtained. It is also established that the new model can be regarded as a compound extreme value model. The structure of this distribution yields a wider class of distributions known as Marshall-Olkin q-Weibull distributions which is an effective model for time series data and other modeling purposes. Various properties of the distribution and hazard rate functions are considered. The problem of estimation of parameters is discussed. The corresponding time series models are also developed to illustrate its application in times series modeling. We also develop different types of autoregressive processes with minification structure and max-min structure which can be applied to a rich variety of contexts in real life. Sample path properties are examined and generalization to higher orders are also made. The model is successfully applied to a data on daily discharge of Neyyar river in Kerala, India. Applications in statistical physics and financial modelling are also discussed.

345 Information bounds and MCMC parameter estimation for the pile up model with application to fluorescence measurements

[CS 62,(page 48)]

Tabea REBAFKA, *Telecom ParisTech, TSI / CNRS LTCI, 46 rue Barrault, 75634 Paris; CEA, LIST, 91191 Gif sur Yvette Cedex, France*

Fran ROUEFF, *Telecom ParisTech, TSI / CNRS LTCI, 46 rue Barrault, 75634 Paris*

Antoine SOULOUMIAC, *CEA, LIST, 91191 Gif sur Yvette Cedex, France*

The pile up model represents a special type of an inverse problem resulting from a nonlinear transformation of the original model. An observation of the pile up model is defined as the minimum of a random number of independent random variables distributed according to the original distribution. The pile up model is motivated by an application in fluorescence spectroscopy, where the TCSPC technique provides data from a pile up model for the recovery of lifetime distributions. Obviously, the amount of distortion of the original model depends on the distribution of the number of random variables over which the minimum is taken. In order to optimize the experimental conditions in the context of fluorescence measurements by a favorable choice of the tuning parameter, which is the average number of random variables over which the minimum is taken, a study of the Cramer-Rao bound is conducted. We focus on the following question: how the amount of distortion affects the information contained in the data. This study shows that the tuning parameter currently used for fluorescence measurements is by far suboptimal. An augmentation of the tuning parameter increases the information drastically and thus a significant reduction of the acquisition time shall be possible. However, data obtained at the optimal choice of the tuning parameter require an estimator adapted to the pile up distortion. Therefore, a Gibbs sampler is presented for the estimation of the parameters of the pile up model in the context of fluorescence measurements. In this case the original distribution is usually modeled as a mixture of exponential distributions and the number of random variables over which the minimum is taken as a Poisson distribution. The covariance matrix of the Gibbs estimator turns out to be close to the Cramer-Rao bound. Thus, applying this method one can reduce the acquistion time considerably in comparison to the standard method of recovering lifetime distributions for a given quadratic risk.

346 Asymptotic behaviour of solutions of the kinetic Kac equation

[CS 72,(page 54)]

Eugenio REGAZZINI, *Dipartimento di Matematica, Universita degli Studi di Pavia, Italy*

Emanuele DOLERA, *Dipartimento di Matematica, Universita degli Studi di Pavia, Italy*

Ester GABETTA, *Dipartimento di Matematica, Universita degli Studi di Pavia, Italy*

Let $f(.,t)$ be the probability density function which represents the solution of the Kac equation at time t, with initial data f_0, and let g_s be the Gaussian density with zero mean and variance s^2, s^2 being the value of the second moment of f_0. We prove that the total variation distance between $f(.,t)$ and g_s goes to zero, as t goes to infinity, with an exponential rate equal to $-1/4$. In the present talk, this fact is proved on the sole assumption that f_0 has finite fourth moment and its Fourier transform h_0 satisfies $|h_0(y)| = o(|y|^{-p})$ as $|y|$ goes to infinity, for some $p > 0$. These hypotheses are definitely weaker than those considered so far in the most advanced literature to obtain, in any case, less precise rates. Moreover, we determine a lower bound which decreases exponentially to zero with the same said rate, provided that f_0 has non-zero kurtosis coefficient. We show that sharper (than $-1/4$) upper bounds are valid when the $(4+d)$–ablsolute moment is finite for some d in $(0, 2)$ provided that the kurtosis coefficient of f_0 is zero.

347 The contact process in a dynamic random environment

[CS 73,(page 54)]

Daniel REMENIK, *Cornell University*

We study a contact process running in a random environment in $\mathbf{Z}^d$ where sites flip, independently of each other, between blocking and non-blocking states, and the contact process is restricted to live in the space given by non-blocked sites. We give a partial description of the phase diagram of the process, showing in particular that, depending on the flip rates of the environment, survival of the contact process may or may not be possible for large values of the birth rate. We prove block conditions for the process that parallel the ones for the ordinary contact process and use these to conclude that the critical process dies out and that the complete convergence theorem holds in the supercritical case.

348 Locally Stationary Representation of Ornstein-Uhlenbeck Processes

[CS 58,(page 47)]

Saeid REZAKHAH, *Amirkabir University of Technology*

Akram Kohansal KOHANCAL, *Amirkabir University of Technology*

Using Lamperti transformation, self-similar processes can be obtained from the stationary processes. The Ornstein-Uhlenbeck process which is stationary, combined with Lamperti transformation generates the Brownian motion process which is a self-similar process. Lim and Muniandy studied the generalized Ornstein-Uhlenbeck process and the associated self-similar process. The generalized Ornstein-Uhlenbeck process has many representations. One of its representations is regarded as Lamperti transformation of fractional Brownian motion. The associated process with the fractional Brownian motion (FBM) by the Lamperti transformation is $Y(t) = \frac{e^{-2aHt}}{\sqrt{4aH}} B_H(t)$. the other representations is the stationary solution of the Langevin equation.

The Langevin equation is defined by $\frac{dY(t)}{dt} + aY(t) = W(t)$, $a > 0$ where $W(t)$ is the standard whit noise. In fact we extend this equation to a fractional Langevin equation and solve it. The stationary solution of it can be assumed as a Ornstein-Uhlenbeck process. The present research was performed to investigate the properties of these processes. Finally we show that the generalized Ornstein-Uhlenbeck can be regarded as the locally Stationary representation of FBM.

349 Detecting periodicity in photon arrival times

[IS 4,(page 55)]

John RICE, *University of California, Berkeley*

I will discuss the statistical problem of detecting periodicity in a sequence of photon arrival times. This occurs, for example, in attempting to detect gamma-ray pulsars. A particular focus is on how auxiliary information, typically source intensity, background intensity, and incidence angles and energies associated with each photon arrival should be used to maximize the detection power. I will present a class of likelihood-based tests, score tests, which give rise to event weighting in a principled and natural way, and derive expressions quantifying the power of the tests. A search for pulsars over a broad frequency range can require very significant computation, and I will discuss a method that maximizes power subject to a computational constraint.

350 Fluctuation theory for positive self-similar Markov processes

[IS 8,(page 50)]

Victor RIVERO, *Centro de Investigacion en Matematicas A.C. Mexico*

Loic CHAUMONT, *University of Angers, France.*

Juan Carlos PARDO, *University of Bath, United Kingdom*

In recent years there has been a growing interest in the theory of positive self-similar Markov processes (pssMp), i.e. positive valued strong Markov processes with right continuous left-limited paths and with the scaling property. This class of processes has been introduced by Lamperti in 1972, in a seminal paper where he proved several important results. A result of particular interest in that article is the nowadays called Lamperti transformation. It establishes that any pssMp killed at their first hitting time of 0 is the exponential of a Levy process time changed by the right-continuous inverse of an additive functional. The Lamperti transformation allows to embed the theory of Levy processes into that of pssMp. This embedding has proved to be a powerful tool to unravel the asymptotic behaviour of pssMp, establish various interesting identities and to link these processes with other areas of applied probability such as mathematical finance, random walks in random environments, continuous state branching processes, fragmentation theory, etc.. Never-the-less the usage of the theory of Levy processes for those purposes has never been simple owing that the time change relating pssMp with Levy processes carries several modifications in the behaviour of a Levy process. For example, it is known that a non-decreasing Levy process with finite mean growths linearly, owing to the law of large numbers, whilst for a pssMp associated via the Lamperti transformation to such a Levy process, growths with a polynomial rate whose order is given by the index of self-similarity, (Bertoin and Caballero (2002) and Rivero (2003)). Our main purpose in this work is to establish a fluctuation theory, similar to and build on the fluctuation theory of Levy processes. We will provide several new identities for pssMp concerning first passages times, the past supremum and infimum, overshoots and undershoots. Moreover, we will construct an upward (downward) ladder process and to determine the existence of entrance laws for it.

351 Unbiased estimation of expected diffusion functionals: the intersection layer approach

[CS 2,(page 6)]

Gareth ROBERTS, *University of Warwick, UK*
Alexandros BESKOS, *University of Warwick, UK*

This presentation will report on current work which extends recent developments on the exact simulation of diffusion trajectories and their use in Monte Carlo estimation of expected diffusion functionals. The present work considers estimation of path functionals which are not analytically tractable for Brownian motion (such as non-linear functionals of non-linear path averages). The methodology derived extends the so-called EA3 algorithm by an iterative procedure which can be combined with a retrospective simulation technique to produce unbiased estimators.

352 Empirical saddlepoint approximations of the studentized mean under stratified random sampling

[CS 25,(page 24)]

John ROBINSON, *School of Mathematics and Statistics, University of Sydney NSW 2006, Australia.*
Zhishui HU, *Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui 230026, China*
Chunsheng MA, *Department of Mathematics and Statistics, Wichita State University,*
Wichita, Kansas 67260-0033, USA

We obtain a saddlepoint approximation for the Studentized mean under stratified random sampling. This is of theoretical interest only, since its calculation requires knowledge of the entire population. We also obtain an empirical saddlepoint approximation based on the stratified sample alone. This empirical approximation can be used for tests of significance and confidence intervals for the population mean and can be regarded as a saddlepoint approximation to a bootstrap approximation (Booth, J.G., Butler, R.W. and Hall, P.G., 1994. Bootstrap methods for finite populations. J. Amer. Statist. Assoc. 89, 1282-1289). We compare the empirical approximation to the true saddlepoint approximation, both theoretically, to obtain results on the relative error of the bootstrap, and numerically, using Monte Carlo methods to compare the true and bootstrap distributions with the saddlepoint and empirical saddlepoint approximations.

353 Dualist statisticians in current literature

[CS 4,(page 8)]

Paolo ROCCHI, *IBM*

Statisticians face the theoretical opposition between the frequentist school and the Bayesian school, but a significant circle of authors openly holds that both the methods should be used. I call ‘dualist’ writers those who ground the dual view of probability upon phisophical considerations instead ‘eclectic’ writers rarely enter into the philosophical nodes. The present paper illustrates a survey upon 70 books written by dualist and eclectic authors. I did not follow predetermined rules for sampling, since this work is

just a first attempt to clarify the shape of current literature. The survey examines three groups of books. Group A includes 22 textbooks and reference books in probability and statistics used in schools and professional practice. Group B has 37 books similar to the previous group but focusing on a special area i.e. medicine, business, economy, engineering. Group C contains 11 books which examine advanced topics in statistics e.g. circular statistics, asymptotic statistics, bootstrap, or go deep into paradoxes and special cases. Sixty-two books are written in English, the remanant in French, Spanish and Italian. The books have the following structures:

I 13 books basically follow the classical methods in statistics or otherwise the Bayesian methods, and add a few pages to mention the other school. The authors belonging to this class do not translate the dualist view into explicit directions. All the books falling into this class belong to group B.

II 45 works illustrate the different interpretations of probability and statistical inference in a manner more extensive than the style just seen, but the authors do not produce thorough works. Writers focus on some special topics of the Bayesian statistics (e.g. prior/posterior probabilities, decision theory, estimators) and of the classical statistics (e.g. hypothesis testing, student-test) but do not develop exhaustive illustration.

III 12 books present the concepts and the results underlying the Bayesian and the Fisherian approaches in an exhaustive manner.

Dualist authors of III go deep into both the inferential methods and compare them. Eclectics authors of I and II are inclined to neglect the diverging significance of the subjective and frequentist probabilities. The approach ‘see and do’ appears evident especially in class II. This approach seems rather disputable because of the odd meanings of the Bayesian and frequentist probabilities.

354 An application of non-homogeneous Poisson models with multiple change-points to air pollution problems

[CS 48,(page 42)]

Eliane R. RODRIGUES, *Instituto de Matematicas-UNAM, Mexico*
Jorge A. ACHCAR, *Faculdade de Medicina de Ribeirao Preto-SP, Brazil*
Guadalupe TZINTZUN, *Instituto Nacional de Ecologia-SEMARNAT, Mexico*

In this talk we consider some non-homogeneous Poisson models to estimate the probability that an air quality standard is exceeded a given number of times in a time interval of interest. We assume that the number of exceedances occur according to a non-homogeneous Poisson process. This Poisson process has rate function $\lambda(t)$, $t \geq 0$, which depends on some parameters that must be estimated. We take into account two cases of rate functions: the Weibull and the Goel-Okumoto. We consider models with and without change-points. When the presence of change-points is assumed, we may have the presence of either one, two or three change-points, depending of the data set. The parameters of the rate functions are estimated using a Gibbs sampling algorithm. The selection of the best model fitting the data is made using the Deviance Information Criterion. Results are applied to ozone data provided by the Mexico City monitoring network.

355 Plug-in choice for non fixed-kernel-shape density estimators

[CS 49,(page 42)]

Alberto RODRÍGUEZ-CASAL, *Universidade de Santiago*
Jose Enrique CHACÓN-DURÁN, *Universidad de Extremadura*

The kernel density estimator is nowadays a well-known tool. It is included in most statistical software packages and commonly used. It consists of estimating an unknown probability density function f by $f_{nh}(x) = n^{-1}\sum_{i=1}^{n} K_h(x-X_i)$ where $X_1, \ldots, X_n$ is a sample from the distribution with density f, the kernel K is an integrable function with $\int K = 1$, $h \in R^+$ is called the bandwidth and $K_h(x) = K(x/h)/h$. It is readily known that the choice of h is crucial in the performance of this estimator and so, many data-dependent methods for choosing this parameter have been proposed; see the survey papers by Cao, Cuevas and González Manteiga (1994).

An immediate generalization of the above estimator is given by $f_{nK}(x) = n^{-1}\sum_{i=1}^{n} K(x - X_i)$ where $K \equiv K_n$ is a sequence of kernel functions. The role of the bandwidth is played now by the whole function K which should be carefully chosen. The goal of this paper is to propose a data based-method for selecting the kernel K.

356 A test for continuous local martingales with application to exchange rate modelling

[CS 68,(page 51)]

David A. ROLLS, *Dept. of Mathematics and Statistics, University of Melbourne*
Owen D. JONES, *Dept. of Mathematics and Statistics, University of Melbourne*

Continuous local martingales, or equivalently, continuously time-changed Brownian motions, are a popular class of models in finance. We present a set of tests for whether observed data are from a continuously time-changed Brownian motion, based on the concept of the crossing tree. Results for simulated data from a range of processes suggest the test is more powerful than an alternative approach which uses the quadratic variation, particularly for shorter datasets. A feature of the method is that it easily identifies the scale at which a continuous local martingale model cannot be rejected. We apply our tests to several timeseries of log-transformed currency exchange rate tick data and show that for moderately large timescales the hypothesis of continuous time-changed Brownian motion cannot be rejected.

357 The exact asymptotic of the collision time tail distribution for independent Brownian particles with different drifts.

[CS 73,(page 54)]

Tomasz ROLSKI, *Wroclaw University*
Zbigniew PUCHALA, *Institute of Theoretical and Applied Informatics*

Let $W = \{x \in \mathbf{R}^n : x_1 < \ldots < x_n\}$ and $X_t^1, \ldots, X_t^n$ are independent, Brownian processes with drifts $a_1, \ldots, a_n$, each starting from $\mathbf{x} = (X_0^1, \ldots, X_0^n) \in W$. We study the collision time $\tau = \inf\{t > 0 : \mathbf{X}_t \notin W\}$. Since particles have different drifts one cannot use directly Karlin-McGregor formula. We show the exact asymptotics of $P_{\mathbf{x}}(\tau > t) = Ch(\mathbf{x})t^{-\alpha}e^{-\gamma t}(1 + o(1))$ as $t \to \infty$ and identify $C, h(\mathbf{x}), \alpha, \gamma$ in terms of the drifts. Different scenarios are described via the notion of stable partition of the drift vector $(a_1, \ldots, a_n)$. The paper will appear in **Probability Theory and Related Fields** .

358 Nonparametric tests for conditional independence via copulas

[CS 18,(page 20)]

Jeroen VK ROMBOUTS, *HEC Montreal,CIRANO and CORE*

This paper proposes nonparametric tests for conditional independence and particularly for Granger non-causality. The proposed test is based on the comparison of copula densities using Hellinger distance. The Bernstein density copula is used to estimated the unknown copulas. First, for $\beta-$mixing data, we investigate the asymptotic proprieties of the Bernstein density copula, i.e., we give its asymptotic bias and variance and we establish its uniform strong convergence integrated squared error properties, uniform strong consistency and asymptotic normality. Second, we show the asymptotic normality of the conditional independence test statistic. A detailed simulation study investigates the performance of the estimators. Applications using financial data are provided.

359 Generalized covariation function for stochastic process with finite first moments: some numerical properties

[CS 27,(page 25)]

Dedi ROSADI, *Department of Mathematics, Gadjah Mada University, Indonesia*

It has been known that many popular models in finance have been developed under assumption that the returns distribution is multivariate normal. However, from numerous empirical studies (see e.g., Rydberg 1997; Rachev and Mittnik, 2000), the normality assumption for many empirical asset returns data can not be justified. It has been shown that many asset returns are typically leptokurtic (heavy-tailed and peaked around the center).

The class of stable distributions (see Samorodnitsky and Taqqu, 1994) , of which the normal distribution is a special case, represents a natural generalization of the Gaussian distribution, and provides a popular alternative for modeling leptokurtic data. In many empirical studies (see e.g., Rachev and Mittnik, 2000), it has been shown that the non-Gaussian stable distributions with parameter index of stability $1 < \alpha < 2$ are more appropriate for modeling asset returns, while preserving the desirable properties of the normal. When the index α is less than 2, the second moments of the stable distribution are infinite, therefore, the dependence between two random variables can not be described using covariance.

However, when it is assumed that the mean of the returns exist ($\alpha > 1$), the dependence can be analyzed using covariation, a generalization of covariance function (several generalized dependence measures are discussed in Rosadi, 2004). Miller (1978) gave a definition of covariation function, however it is applied only for stable random variables with finite first moment.

Based on Gallagher (1998), in this paper we consider a linear dependence measure called as generalized covariation function, which is not only applicable for two stably distributed random variables, but also for two random variables with finite first moments, and morever, contained covariation and covariance function as a special case. We consider a moment type estimator for the function and investigate the numerical properties of this estimator using the simulated data and the real data from Jakarta Stock Exchange (JSX).

References

1. Belkacem, L., Vehel, J.L. and Walter, C., 2000, CAPM, Risk and Portofolio Selection in α-stable markets, Fractals, 8, 99-116
2. Gallagher, C.(1998) Fitting ARMA models to heavy tailed data. Ph.D. thesis, University of California, Santa Barbara.
3. Miller, G., 1978, Properties of Certain Symmetric Stable Distributions, Journal of Multivariate Analysis, 8, 346-360
4. Nikias,C.L. and Shao, M. (1995) Signal Processing with Alpha-Stable Distributions and Applications. New York: John Wiley & Sons.
5. Rachev, S.T. and Mittnik, S., 2000, Stable Paretian Models in Finance, Wiley, Chichester
6. Rydberg, 1997, Realistic Statistical Modeling of Financial data, Proceedings of 51th International Statistical Institute Conference, Istambul
7. Rosadi, D., 2004, The codifference function for a-stable processes: Estimation and Inference, Thesis Ph.D., Institute for Mathematical Methods, Research Unit : Econometrics and Operation Research, Vienna University of Technology, Vienna, Austria
8. Samorodnitsky, G. and Taqqu, M. S., 1994. Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance. Chapman and Hall. New York.

360 mathStatica: a symbolic approach to computational mathematical statistics

[CS 15,(page 16)]

Colin ROSE, *Theoretical Research Institute*

Living in a numerical versus symbolic computational world is not merely an issue of accuracy. Nor is it merely about approximate (numerical) versus exact (symbolic) solutions. More importantly, a symbolic approach to computational statistics significantly changes what one can do, and how one does it. Unlike almost any other package, mathStatica has been designed on top of a computer algebra system to provide a general toolset for doing exact (symbolic) mathematical statistics. It provides automated statistical operators for taking expectations, finding probabilities, deriving transformations of random variables, finding moments, order statistics, cumulative distribution functions, characteristic functions etc - all for completely arbitrary user-defined distributions.

Unlike the traditional approach to computational statistical software, we illustrate how a symbolic approach can significantly change the notion of what is difficult, what one can reasonably solve, how one solves it, and perhaps even the very notion of what is publishable. We argue that the shift from traditional numerical software to symbolic software has broader parallels, in particular to an evolving epistemology of statistical knowledge ... essentially a shift from a 19th C database conception of knowledge to a broader algorithmic one. These ideas will be illustrated in real-time using the forthcoming mathStatica v2 release, using new algorithms for automating transformations of random variables, finding products of piecewise random variables, solving many-to-one transformations, solviong problems such as finding the pdf of Min[X, Y, Z, ...], calculating order statistics with non-identical parent distributions, multivariate moments of moments, and other edible comestibles.

361 Sequential monitoring and randomization tests

[CS 63,(page 48)]

William F. ROSENBERGER, *Department of Statistics, George Mason University*
Yanqiong ZHANG, *Sanofi-Aventis*
R. T. SMYTHE, *Department of Statistics, Oregon State University*

Randomization provides a basis for inference, but it is rarely taken advantage of. We discuss randomization tests based on the family of linear rank tests in the context of sequential monitoring of clinical trials.

Such tests are applicable for categorical, continuous, and survival time outcomes. We prove the asymptotic joint normality of sequentially monitored test statistics, which allows the computation of sequential monitoring critical values under the Lan-DeMets procedure. Since randomization tests are not based on likelihoods, the concept of information is murky. We give an alternate definition of randomization and show how to compute it for different randomization procedures. The randomization procedures we discuss are the permuted block design, stratified block design, and stratified urn design. We illustrate these results by reanalyzing a clinical trial in retinopathy.

362 Decompositions and structural analysis of stationary infinitely divisible processes

[IS 8,(page 50)]

Jan ROSIŃSKI, *University of Tennessee, USA*

Many classes of infinitely divisible processes $\mathbf{X} = \{X_n\}_{n\in\mathbf{Z}}$ can be characterized by their path Lévy measures having a product convolution form of a certain fixed 'root' Lévy measure ρ on $\mathbf{R}$ and 'mixing' measures τ on $\mathbf{R}^{\mathbf{Z}}$. For example, selfdecomposable processes indexed by $\mathbf{Z}$ have their path Lévy measures of the form

$$\nu(A) = \int_0^\infty \tau(s^{-1}A)\,\rho(ds), \qquad A \subset \mathbf{R}^{\mathbf{Z}},$$

where $\rho(ds) = s^{-1}I_{(0,1]}(s)ds$ and the mixing measure τ on $\mathbf{R}^{\mathbf{Z}}$ satisfies $\int(x_n^2 \wedge |x_n|)\,\tau(dx) < \infty$, $n \in \mathbf{Z}$. Another example is the Thorin class of gamma mixtures, where the root Lévy measure is $\rho(ds) = s^{-1}e^{-s}I_{(0,\infty)}(s)ds$.

The mixing measure τ is assumed to be known (in some form); it plays the role of a spectral measure of the process. If $\mathbf{X}$ is a strictly stationary infinitely divisible process without Gaussian part, then τ is also invariant under the shift T. In this talk we relate properties of two dynamical systems: a probabilistic system $(\mathbf{R}^{\mathbf{Z}}, P_{\mathbf{X}}; T)$ and a 'deterministic' but possibly infinite $(\mathbf{R}^{\mathbf{Z}}, \tau; T)$ (here $P_{\mathbf{X}} = \mathcal{L}(\mathbf{X})$). We investigate $(\mathbf{R}^{\mathbf{Z}}, P_{\mathbf{X}}; T)$ as a factor of an infinitely divisible suspension $(\mathcal{M}(\mathbf{R}^{\mathbf{Z}}), P_\tau; T^*)$ over $(\mathbf{R}^{\mathbf{Z}}, P_{\mathbf{X}}; T)$, which has a rich structure and is directly related to $(\mathbf{R}^{\mathbf{Z}}, \tau; T)$. In addition, $L^2(\mathcal{M}(\mathbf{R}^{\mathbf{Z}}), P_\tau)$ admits chaotic decomposition

$$L^2(\mathcal{M}(\mathbf{R}^{\mathbf{Z}}), P_\tau) = \bigoplus_{n=0}^{\infty} \bigoplus_{i_1,\ldots,i_n\in\mathbf{N}} \mathcal{H}^{(i_1,\ldots,i_n)},$$

where $\mathcal{H}^{(i_1,\ldots,i_n)}$ are spaces of multiple stochastic integrals with respect to strongly orthogonal Teugels martingales. We will discuss ergodic decompositions of processes into parts with different long range memory structures. This approach applies not only to the underlying infinitely divisible processes but also to their nonlinear functionals such as multiple integrals.

363 On the asymptotic analysis of hierarchical and graphical log-linear models for binary data

[IS 7,(page 5)]

Alberto ROVERATO, *University of Bologna, Italy*

Discrete graphical models are a proper subclass of the hierarchical log-linear models so that all the theoretical results concerning statistical inference for hierarchical log-linear models also apply to the subclass of graphical models. However, graphical models satisfy useful properties that are not generally shared by an arbitrary hierarchical log-linear model. We consider binary data and provide a formulation of the asymptotic theory of hierarchical log-linear models that allows to fully exploit the specific features of graphical models. For hierarchical log-linear models our approach allows to derive, in an alternative way, certain results of asymptotic theory, thereby providing additional details with respect to the existing literature. For the subclass of graphical models, including the non-decomposable case, we give explicit rules for the local computation of the asymptotic variance of maximum likelihood estimates of the log-linear parameters. This is of theoretical interest but also of practical usefulness because it leads to an efficiency improvement in the computation of standard errors and allows the local computation statistical quantities that involve the determinant of the Fisher information matrix, such as the Jeffrey's non informative prior and the Laplace approximation to the Bayes factor.

364 Genetic information and Cryptanalysis of the human Genome using statistical tools

[CS 75,(page 56)]

Balai Chandra ROY, *The Institute of Radio Physics and Electronics Science College, Calcutta University*

Genetic information is the discipline concerned with the characterization in probabilistic terms of the information sources and the reliability of the genetic data when conveyed across the cells of the living organs. Within the human genome, individual genes carry information pertinent through specific characteristics. The statistical problems of uncertainty and complexity of the evolutionary processes are studied in the present paper. The optimality and the instability of the genetic codes are also analyzed by using the cryptographic tools.

The biological perspective of the DNA molecules consisting of the long strings of the nucleotide bases are taken into account in describing the sequence of the genetic codes. This genetic code provides the way of communicating the information from genes to proteins that help a cell to do its work. The cryptographic statistical analysis based on projection operators on the DNA bases is seen to have major impacts on the diagnosis of diseases.

365 On the representation of Fleming-Viot models in Bayesian nonparametrics

[CS 19,(page 20)]

Matteo RUGGIERO, *University of Pavia*
Stephen G. WALKER, *University of Kent*

Fleming-Viot processes are probability-measure-valued diffusions which arise as large population limits of a wide class of population genetics models. Motivated by the fact that the stationary distributions of some Fleming-Viot diffusions involve the Dirichlet process, which is widely used in Bayesian statistics, we provide several explicit constructions of Fleming-Viot processes based on some typically Bayesian models, as Polya prediction schemes, Gibbs sampling procedures and hierarchical mixtures. In particular, by means of known and newly defined generalised Polya urn schemes, several types of pure jump particle processes are introduced, describing the evolution in time of an exchangeable population. Then the process of empirical measures of the individuals converges in the Skorohod space to a Fleming-Viot diffusion, and the stationary distribution is de Finetti measure of the infinite sequence of individuals. The construction also suggests an MCMC sampling technique for simulating finite-population approximations to the measure-valued diffusions. Some work in progress along the same line on interacting systems of Fleming-Viot processes is also sketched.

366 Kingman's unlabeled n-coalescent

[CS 21,(page 21)]

Raazesh SAINUDIIN, *Biomathematics Research Centre, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*
Peter DONNELLY, *Department of Statistics, University of Oxford, Oxford, UK*
Robert C. GRIFFITHS, *Department of Statistics, University of Oxford, Oxford, UK*
Gilean MCVEAN, *Department of Statistics, University of Oxford, Oxford, UK*
Kevin THORNTON, *Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of California, Irvine, USA*

We derive the transition structure of a Markovian lumping of Kingman's n-coalescent (J.F.C. Kingman, *On the genealogy of large populations*, J. Ap. Pr., 19:27-43, 1982). Lumping a Markov chain is meant in the sense of Kemeny and Snell (def. 6.3.1, *Finite Markov Chains*, D. Van Nostrand Co., 1960). The lumped Markov process, referred as the unlabeled n-coalescent, is a continuous-time Markov chain on the set of all integer partitions of the sample size n. We derive the forward, backward and the stationary probabilities of this chain. We show that the likelihood of any given site-frequency-spectrum, a commonly used statistics in genome scans, from a locus free of intra-locus recombination, can be directly obtained by integrating conditional realizations of the unlabeled n-coalescent. We develop an importance sampler for such integrations that relies on an augmented unlabeled n-coalescent forward in time. We apply the methods to population-genetic data from regions of the human genome with low recombination to conduct demographic inference at the empirical resolution of the site-frequency-spectra.

367 Maximum likelihood estimation of a multidimensional log-concave density

[CS 53,(page 44)]

Richard SAMWORTH, *University of Cambridge, UK*
Madeleine CULE, *University of Cambridge, UK*
Robert GRAMACY, *University of Cambridge, UK*
Michael STEWART, *University of Sydney, Australia*

We show that if $X_1, ..., X_n$ are a random sample from a log-concave density f in d-dimensional Euclidean space, then with probability one there exists a unique maximum likelihood estimator $\hat{f}_n$ of f. The use of this estimator is attractive because, unlike kernel density estimation, the estimator is fully auto-

matic, with no smoothing parameters to choose. The existence proof is non-constructive, but by reformulating the problem as one of non-differentiable convex optimization, we are able to develop an iterative algorithm that converges to the estimator. We will also show how the method can be combined with the EM algorithm to fit finite mixtures of log-concave densities, and hence perform clustering. The talk will be illustrated with pictures from the R package LogConcDEAD - Log-Concave Density Estimation in Arbitrary Dimensions.

368 Preservation of reliability classes under mixtures of renewal processes

[CS 32,(page 31)]

C. SANGUESA, *Departamento de Metodos Estadisticos. Universidad de Zaragoza, Spain*
F. G. BADIA, *Departamento de Metodos Estadisticos. Universidad de Zaragoza, Spain*

In this work we provide sufficient conditions for the arrival times of a renewal process so that the number of its events occurring before a randomly distributed time, T, independent of the process preserves the aging properties of T.

369 Generalized weighted Simes test

[IS 12,(page 18)]

Sanat SARKAR, *Temple University*

In this talk, a generalized version of the weighted Simes test of Benjamini and Hochberg (1997, *Scandinavian Journal of Statistics*) for testing an intersection null hypothesis will be presented based on the generalized Simes test recently proposed by Sarkar (2007, *Annals of Statistics*). The control of the Type I error rate, both under independence and some form of positive dependence, will be given. A potential use of it in developing a generalized gatekeeping strategy will be discussed.

370 Statistical analysis for a three station tandem queue with blocking and infinite queue infront of station 1

[CS 13,(page 13)]

Paul R. SAVARIAPPAN, *Luther College, Decorah, Iowa-52101-1045*

A maximum likelihood estimator (MLE), a consistent asymptotically normal (CAN) estimator and asymptotic confidence limits for the expected number of customers in the system in a three station tandem queue with blocking and infinite queue capacity infront of station 1 and zero queue capacity infront of stations 2 and 3 are obtained.

371 Identifying influential model choices in Bayesian hierarchical models

[CS 19,(page 20)]

Ida SCHEEL, *Department of Mathematics, University of Oslo and $(sfi)^2$- Statistics for Innovation*
Peter J. GREEN, *Department of Mathematics, University of Bristol*
Jonathan C. ROUGIER, *Department of Mathematics, University of Bristol*

Real-world phenomena are frequently modelled by Bayesian hierarchical models. The building-blocks in such models are the distribution of each variable conditional on parent and/or neighbour variables in the graph. The specifications of centre and spread of these conditional distributions may be well-motivated, while the tail specifications are often left to convenience. However, the posterior distribution of a parameter may depend strongly on such arbitrary tail specifications. This is not easily detected in complex models. In this paper we propose a graphical diagnostic which identifies such influential statistical modelling choices at the node level in any chain graph model. Our diagnostic, *the local critique plot*, examines local conflict between the information coming from the parents and neighbours (local prior) and from the children and co-parents (lifted likelihood). It identifies properties of the local prior and the lifted likelihood that are influential on the posterior density. We illustrate the use of the local critique plot with applications involving models of different levels of complexity. The local critique plot can be derived for all parameters in a chain graph model, and is easy to implement using the output of posterior sampling.

372 CvM and KS two-sample test based on regression rank scores

[PS 3,(page 29)]

Martin SCHINDLER, *Charles University in Prague, Czech Republic*

We derive the two-sample Cramér-von Mises and Kolmogorov-Smirnov type test of location when a nuisance linear regression is present. The test is based on regression rank scores and provides a natural extension of the classical CvM or KS test of location.

Their asymptotic distributions under the hypothesis and the local alternatives are similar to those of the classical tests.

373 Waiting for two mutations: with applications to DNA regulatory sequence evolution and the limits of Darwinian evolution

[IS 14,(page 5)]

Deena SCHMIDT, *Institute for Mathematics and its Applications, University of Minnesota*
Richard DURRETT, *Department of Mathematics, Cornell University*

Results of Nowak and collaborators concerning the onset of cancer due to the inactivation of tumor suppressor genes give the distribution of time until some individual in a population has experienced two prespecified mutations, and the time until this mutant phenotype becomes fixed in the population. We apply and extend these results to obtain insights into DNA regulatory sequence evolution in *Drosophila* (fruit flies) and humans. In particular, we examine the waiting time for a pair of mutations, the first of which inactivates an existing transcription factor binding site and the second which creates a new one. Consistent with recent experimental observations for *Drosophila*, we find that a few million years is sufficient, but for humans with a much smaller effective population size, this type of change would take more than 100 million years. In addition, we use these results to expose flaws in some of Michael Behe's arguments concerning mathematical limits to Darwinian evolution.

374 On the Distribution of the Adaptive LASSO Estimator

[CS 26,(page 25)]

Ulrike SCHNEIDER, *University of Vienna*
Benedikt M. POETSCHER, *University of Vienna*

The LASSO estimator is a penalized least-squares (LS) estimator which acts simultaneously as variable selection and coefficient estimation method. A variant of the LASSO, the so-called adaptive LASSO estimator, was introduced by Zou (2006) and shown to possess an 'oracle'-property, i.e. its asymptotic distribution under fixed parameters coincides with the one of the OLS estimator of the smallest correct model and is, in particular, normal.

We study the distribution of the adaptive LASSO estimator for a linear regression model with orthogonal regressors and Gaussian errors in finite samples as well as in the large-sample limit. Both types of distributions are derived for the case where the tuning parameter of the adaptive LASSO estimator is chosen so that the estimator performs conservative model selection (i.e. a procedure that asymptotically only selects correct, but possibly overparametrized models), as well as for the case where the tuning results in consistent model selection. We show that these distributions are typically highly non-normal regardless of the choice of tuning and mention similar results by Ptscher and Leeb (2007) for other well-known penalized LS estimators. Moreover, the uniform convergence rate is obtained and shown to be slower than $n^{-1/2}$ in case the estimator is tuned to perform consistent model selection. In this context, we also discuss the questionable statistical relevance of the 'oracle'-property of the estimator.

We present simulation results for the case of non-orthogonal regressors to complement and confirm our theoretical findings for the orthogonal case.

Finally, we provide an impossibility result regarding the estimation of the distribution function of the adaptive LASSO estimator.

375 Dynamical and near-critical percolation

[BS-IMS Lecture 2,(page 15)]

Oded SCHRAMM, *Microsoft Research*
Christophe GARBAN, *Université Paris-Sud*
Gábor PETE, *Microsoft Research*

In one of several mathematical models of percolation, the edges (or sites) of a lattice are selected with some probability p, independently, and the connectivity properties of the resulting graph are studied. There is a critical value p_c, such that when $p > p_c$ there is with probability one an infinite connected component in the percolation subgraph and when $p < p_c$, the probability for an infinite component is zero. In dynamical peroclation, the bits determining the percolation configuration switch on and off according to Poisson clocks, independently. We will describe some recent results concerning dynamical and near-critical percolation on two-dimensional lattices, as well as applications to the theory minimal spanning trees in two dimensions.

376 On parametric estimation for linear ARCH processes

[CS 8,(page 11)]

Martin SCHUETZNER, *Department of Mathematics and Statistics, University of Konstanz, Germany*
Jan BERAN, *Department of Mathematics and Statistics, University of Konstanz, Germany*

Several studies have shown that financial time series, such as asset returns or exchange rates, often exhibit long memory in volatility, in the sense of slowly decaying, possibly non-summable, autocorrelations in the squared values. A model with this property is the linear ARCH (LARCH) process, introduced by Robinson (1991), with the square root of the conditional variance being, upto a possible change of sign, a linear combination of past observations. Basic probabilistic properties of LARCH processes have been investigated in recent papers by Giraitis, Robinson and Surgailis (2000), Berkes and Horvath (2003), Beran (2006), and Beran and Feng (2007), among others. However, one of the main reasons why to date LARCH processes have not been used in practice is the lack of theoretical results on parameter estimation.

In this paper, three different parameter estimators are considered. Asymptotic results are derived under short- and long-memory conditions respectively. The first method consists of a modified conditional maximum likelihood estimator. Consistency and asymptotic $n^{1/2}$-rate of convergence is derived under the assumption that the conditional variance can be computed exactly. For a computable version of the estimator, $n^{1/2}$-convergence still holds in the case of short memory, whereas this is no longer true for long-memory LARCH models. The second estimator is essentially a method of moments based on empirical autocovariances of the squared observations X_t^2. Limit theorems for partial sums of the products $X_t^2 X_{t+k}^2$ and the estimator are derived, with the rate of convergence depending on the strength of long memory. Finally, we investigate a local Whittle estimator based on the process X_t^2. A simulation study illustrates the asymptotic results.

References

1. Beran, J. (2006) On location estimation for LARCH processes, Journal of Multivariate Analysis, 97, 1766-1782.
2. Beran, J. and Feng, Y. (2007) Weighted averages and local polynomial estimation for fractional linear ARCH processes. J. Statistical Theory and Practice, 1, 149-166.
3. Berkes, I. and Horvath, L. (2003) Asymptotic results for long memory LARCH sequences, Ann. Appl. Probab., 13, 641-668.
4. Giraitis, L., Robinson, P.M. and Surgailis, D. (2000) A model for long memory conditional heteroskedasticity, Ann. Appl. Probab., 10, 1002-1024.
5. Robinson, P.M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, J. Econometrics, 47, 67-84.

377 Loop-erased random walk on finite graphs and the Rayleigh process

[IS 17,(page 31)]

Jason SCHWEINSBERG, *University of California, San Diego*

Let $(G_n)_{n=1}^{\infty}$ be a sequence of finite graphs, and let Y_t be the length of a loop-erased random walk on G_n after t steps. We show that for a large family of sequences of finite graphs, which includes the case in which G_n is the d-dimensional torus of size-length n for $d \geq 4$, the process $(Y_t)_{t=0}^{\infty}$, suitably normalized, converges to the Rayleigh process introduced by Evans, Pitman, and Winter. Our proof relies heavily on ideas of Peres and Revelle, who used loop-erased random walks to show that the uniform spanning tree on large finite graphs converges to the Brownian continuum random tree of Aldous.

378 Parameter estimation for integer-valued autoregressive processes with periodic structure

[CS 16,(page 16)]

Manuel SCOTTO, *Departamento de Matemática, Universidade de Aveiro, Portugal*
Magda MONTEIRO, *Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, Portugal*
Isabel PEREIRA, *Departamento de Matemática, Universidade de Aveiro, Portugal*

This talk is primarily devoted to the problem of estimating the parameters of the periodic integer-valued autoregressive process of order one with period T, defined by the recursive equation

$$X_t = \phi_t \circ X_{t-1} + Z_t,\ t \geq 1,$$

with $\phi_t = \alpha_j \in (0,1)$ for $t = j+kT, (j = 1, \ldots, T,\ k \in N_0)$, where the *thinning* operator $\circ$ is defined as

$$\phi_t \circ X_{t-1} \overset{d}{=} \sum_{i=1}^{X_{t-1}} U_{i,t}(\phi_t),$$

being $(U_{i,t}(\phi_t))$, for $i = 1, 2, \ldots$, a periodic sequence of independent Bernoulli random variables with success probability $P(U_{i,t}(\phi_t) = 1) = \phi_t$. Furthermore it is assumed that $(Z_t)_{t\in N}$ constitutes a periodic sequence of independent Poisson-distributed random variables with mean υ_t, with $\upsilon_t = \lambda_j$ for $t = j + kT, (j = 1, \ldots, T, k \in N_0)$, which are assumed to be independent of X_{t-1} and $\phi_t \circ X_{t-1}$.

Several methods for estimating the parameters of the model are discussed in detail. Their asymptotic properties and corresponding finite sample behavior are also investigated.

379 An eigenfunction problem and the multivariate Kaplan-Meier estimator

[CS 28,(page 26)]

Arusharka SEN, *Department of Mathematics and Statistics, Concordia University, Canada*

Winfried STUTE, *Mathematics Institute, Justus-Liebig University, Germany*

We start with the observation that a multivariate survivor function is an eigenfunction of an integral operator given by the corresponding cumulative hazard function. This yields, under multivariate random censoring, the multivariate Kaplan-Meier estimator as a matrix eigenvector. It reduces to the familiar product-limit formula in the univariate case. We obtain the influence function of the estimator and demonstrate its efficiency. We also obtain an estimator of its asymptotic variance and present some simulation results.

380 Restricted regression estimation of parameters in a measurement error model

[CS 7,(page 9)]

SHALABH, *Department of Mathematics and Statistics Indian Institute of Technology Kanpur Kanpur - 208016, India*

Gaurav GARG, *Indian Institute of Technology Kanpur, Kanpur - 208016, Uttar Pradesh, India*

Neeraj MISRA, *Department of Mathematics and Statistics Indian Institute of Technology Kanpur Kanpur - 208016, India*

A multivariate ultrastructural measurement error model is considered and it is assumed that some prior information is available in the form of exact linear restrictions on regression coefficients. Using the prior information along with the additional knowledge of covariance matrix of measurement errors associated with explanatory vector and reliability matrix, we have proposed three methodologies to construct the consistent estimators which also satisfy the given linear restrictions. Asymptotic distribution of these estimators is derived when measurement errors and random error component are not necessarily normally distributed. Dominance conditions for the superiority of one estimator over the other under on the criterion of Loewner ordering are obtained for each case of the additional information. Some conditions are also proposed under which the use of a particular type of information will give more efficient estimator.

381 A New Method for assigning ,the eigenvalues sign in equation $(Ax = \lambda x)$ and $(Ax = \lambda Bx)$

[PS 2,(page 17)]

Maryam SHAMS SOLARY, *Guilan University(PhD student)*

Hashem SABERI NAJAFI, *Guilan University(PhD)*

The inertia of an $n \times n$ complex matrix A,is defined to be an integer triple ,$In(A) = (\pi(A), \nu(A), \delta(A))$,where $\pi(A)$ is the number of eigenvalues of A with positive real parts,$\nu(A)$ is the number of eigenvalues with negative real parts and $\delta(A)$ is the number of eigenvalues with zero real parts.We are interested in computing the Inertia for large unsymmetric generalized eigenproblem (A,B) for equation $A\varphi = \lambda B\varphi$ Where A and B are $n \times n$ large matrices. For standard eigenvalues problem let $B = Identity\ \ matrix$. An obvious approach for determine Inertia of pair(A,B) ,is to transform this to a standard eigenproblem by inverting either A or B. Many important characteristics of physical and engineering systems, such as stability ,can often be determined only by knowing the nature and location of the eigenvalues.In this paper we show that the eigenvalues sign can be computed by assigning the interval that including all the eigenvalues and this method is compared by results in Matlab.

382 Exponential survival entropies and their properties

[CS 28,(page 26)]

Dilip Kumar SHARMA, *Jaypee Institute of Engineering and Technology A.B. Road, Raghogarh*

D.S. HOODA, *Jaypee Institute of Engineering and Technology, Guna, India*

In the present communication the multivariate survival function of a random variable X is used to define four new classes of exponential survival entropies. It is shown that logarithm of these classes give well known cumulative residual entropies studied by various authors. Explicit expressions of the new measures for specific distribution functions are derived and some important properties of the proposed classes are also studied.

383 A tight frame approach for missing data recovery in images

[IS 3,(page 55)]

Zuowei SHEN, *Department of Mathematics, National University of Singapore*

In many practical problems in image processing, the observed data sets are often in complete in the sense that features of interest in the image are missing partially or corrupted by noise. The recovery of missing data from incomplete data is an essential part of any image processing procedures whether the final image is utilized for visual interpretation or for automatic analysis. In this talk, we will discuss our new iterative algorithm for image recovery for missing data which is based on tight framelet systems constructed by the unitary extension principle. We consider in particular few main applications in image processing, inpainting, impulse noise removal and super-resolution image reconstruction

384 Space-time modeling of biomass burning and regional aerosols in Southeast Asia

[IS 6,(page 30)]

Tao SHI, *Department of Statistics, The Ohio State University*
Kate CALDER, *The Ohio State University*
Darla MUNROE, *The Ohio State University*
Ningchuan XIAO, *The Ohio State University*

Scientists and policy makers have become increasingly concerned about the implications of the consistent brown haze covering Southeast Asia in terms of human health and climate change. The emergence of this haze is due to increased atmospheric concentrations of carbonaceous aerosols, which are generated by anthropogenic activities including both shifting/swidden agriculture and fossil fuel combustion. Our research focuses on determining the relative contribution of these two types of emissions to the total aerosol burden over the region.

We propose a space-time model for regional carbonaceous aerosol composition and concentration, given atmospheric circulation processes and observed fire occurrence. Our model synthesizes a variety of types of data including remote sensing imagery, output from atmospheric transport models, and estimates of biomass emissions for various vegetation types.

This is joint work with Prof. Kate Calder (Statistics, OSU), Prof. Darla Munroe (Geography, OSU), and Prof. Ningchuan Xiao (Geography, OSU). This project is supported by NASA Land-Cover/Land-Use Change program.

385 Multifractal moment-scalings in birth and death processes

[PS 2,(page 17)]

Narn-Rueih SHIEH, *Department of Mathematics, National Taiwan University, Taipei 10617, Taiwan*

We investigate the properties of multifractal products of the exponential of a birth and death process with certain given marginal discrete distribution and given covariance structure. The conditions on the mean, variance and covariance functions of the resulting cumulative processes are interpreted in terms of the moment generating functions. We provide four illustrative examples of Poisson, Pascal, binomial and hypergeometric distributions. We establish the corresponding log-Poisson, log-Pascal, log-binomial and log-hypergeometric scenarios for the limiting processes, including their Rényi functions and dependence structures. This is a joint work with V. Anh (Brisbane) and N.N. Leonenko (Cardiff).

386 Sleepy walkers: progress, conjectures and challenges

[CS 73,(page 54)]

Vladas SIDORAVICIUS, *IMPA, Rio de Janeiro and CWI, Amsterdam*

We study an infinite system of random walkers on the integer lattice Z^d. The particles can exist in two states, active or inactive (sleeping); only active particles perform simple symmetric random walk. The number of particles is conserved during the evolution, and active particles do not interact among themselves. At some positive rate particles which are alone at given vertex can fall asleep, and remain so until joined by another active particle at the same vertex. The state with all particles sleeping is absorbing state of the system. Whether activity continues at long times depends on the relation between the particle density and the rate at which particles go to sleep. In spite of existence of very extensive and detailed physics literature on this model, almost no rigorous results are available. This is mostly due to the fact that above described interaction makes the dynamics not monotone, and, approaching the critical line interaction becomes highly non-local. I will discuss difficulties and some partial progress for the general case, and then, for the one-dimensional case, will show the existence of non-trivial phase transition between an active phase (for sufficiently large particle densities and/or small sleep rate) and an absorbing one. The phase transition appears to be continuous in both the symmetric and asymmetric versions of the process, but the critical behavior is very different. I will also focus on the behaviour of the system at the critical line, where system does not fixate and where an unusual form of the ageing phenomena appears. Finally I will discuss an interesting transient phenomenon of damped oscillations in the density of active particles. Talk is based on joint works with R. Dickman, C. Hoffman, L. Rolla

387 The generalized t-distribution on the circle

[CS 76,(page 57)]

Hai-Yen SIEW, *The Institute of Statistical Mathematics*
Shogo KATO, *The Institute of Statistical Mathematics*
Kunio SHIMIZU, *Keio University*

An extended version of t-distribution on the unit circle is generated by conditioning a normal mixture distribution, which is broadened to include not only unimodality and symmetry, but also bimodality and asymmetry, depending on the values of parameters. After reparametrization, the distribution contains four circular distributions as special cases: symmetric Jones–Pewsey, generalized von Mises, generalized cardioid and generalized wrapped Cauchy distributions. As an illustrative example, the proposed model is fitted to the number of occurrences of the thunder in a day.

388 Bayesian inference for mixed graph models

[IS 7,(page 5)]

Ricardo SILVA, *Statistical Laboratory, University of Cambridge*

Directed acyclic graphs (DAGs) have been widely used as a representation of conditional independence. Moreover, hidden or latent variables are often an important component of graphical models. However, DAG models suffer from an important limitation: the family of DAGs is not closed under marginalization. This means that in general we cannot use a DAG to represent the independencies over a subset of variables in a larger DAG. Directed mixed graphs (DMGs) are a representation that includes DAGs as a special case, and overcomes this limitation. In this talk, we will present algorithms for performing Bayesian inference in Gaussian and probit DMG models. An important requirement for inference is the characterization of the distribution over parameters of the models. We introduce a new distribution for covariance matrices of Gaussian DMGs. We also introduce more advanced algorithms for scaling-up the computation of marginal likelihoods for Gaussian models of marginal independence.

389 Study of a stochastic model for mobile networks

[CS 32,(page 31)]

Florian SIMATOS, *INRIA Paris-Rocquencourt*
Danielle TIBI, *Université Paris 7*

A stochastic model for mobile networks is studied. Users arrive at random times into the network, and then perform independent Markovian routes between nodes, where they receive service according to the Processor-Sharing policy. Once their initial service requirement is satisfied, customers leave the system.

The stability region is identified via a fluid limit approach, and strongly relies on a "homogenization" property. This property translates to the fact that, at the fluid level, the customers are instantaneously distributed across the network according to the stationary distribution of the routing dynamic. Note that because customers move independently and according to the same dynamic, this result intuitively follows from the Law of Large Numbers. Besides this instantaneous behavior, stronger results concerning the long-time behavior are proved. Namely, in the unstable case, this property is shown to hold almost surely as time goes to infinity: because of a reinforcing effect, customers stay distributed according to this distribution forever. In the stable case, customers stay distributed as such as long as the network is not empty. One of the technical achievements of this paper is the construction of a family of martingales associated to the multi-dimensional process of interest, which makes it possible to get estimates of hitting times, crucial in establishing the aforementioned properties.

390 Taxicab non symmetrical correspondence analysis

[CS 65,(page 49)]

Biagio SIMONETTI, *Universit del Sannio*

Non Symmetrical correspondence analysis (NSCA) is a variant of the classical Correspondence Analysis (CA) for analyze two-way contingency table with a structure of dependence between two variables. In order to overcome the influence due to the presence of outlier, in this paper, it is presented Taxicab Non Symmetrical Correspondence Analysis (TNSCA), based on the taxicab singular value decomposition. It will show that TNSCA it is more robust than the ordinary NSCA, because it gives uniform weights to all the points. The visual map constructed by TNSCA has a larger sweep and clearer perspective than the map obtained by correspondence analysis. In the full paper we'll present the analysis of two data sets. The first data set that does not contain any outliers, that cross-classifies the daily consumption of wine with the attained level of education for liver patients. The data are based on the findings of a 2003 survey of 826 patients suffering from liver sickness which was conducted by the Department of Clinic Medicine and Infectious Disease, Second University of Naples. The second data set contains some influential points or outliers and is found in Bradley, Katti and Coons (1962) concerning a sample of 210 individuals that were asked to reflect their impression of five foods on a five point scale. It will be seen that for these two data sets NSTCA produces more interpretable results than the ordinary NSCA.

391 Some aspects of diallel cross designs with correlated observations

[CS 43,(page 38)]

Karabi SINHA, *Department of Biostatistics, University of California Los Angeles (UCLA)*

The purpose of this talk is to present some theoretical results for analysis of diallel cross designs in blocked situations with a possible correlation structure within each block.

392 Application of algebraic statistics for statistical disclosure limitation

[IS 2,(page 23)]

Aleksandra B. SLAVKOVIC, *Penn State University*
Stephen E. FIENBERG, *Carnegie Mellon University*

Statistical disclosure limitation applies statistical tools to the problems of limiting sensitive information releases about individuals and groups that are part of statistical databases while allowing for proper statistical inference. The limited releases can be in a form of arbitrary collections of marginal and conditional distributions, and odds ratios for contingency tables. Given this information, we discuss how tools from algebraic geometry can be used to give both complete and incomplete characterization of discrete distributions for contingency tables. These problems also lead to linear and non-linear integer optimization formulations. We discuss some practical implication, and challenges, of using algebraic statistics for data privacy and confidentiality problems.

393 The application of Self Organizing Maps and evolutionary algorithms to distinguish bacterial proteomes

[PS 3,(page 29)]

Maciej SOBCZYNSKI, *Department of Genomics, Faculty of Biotechnology, Wroclaw University, Poland*
Pawel MACKIEWICZ, *Department of Genomics, Faculty of Biotechnology, Wroclaw University, Poland*
Piotr LIPINSKI, *Institute of Computer Science, Wroclaw University, Poland*
Stanislaw CEBRAT, *Department of Genomics, Faculty of Biotechnology, Wroclaw University, Poland*

Rapid increase of many completely sequenced genomes delivers huge data that require special large-scale analyses carried out by sufficient statistical and data mining methods. We used Self-Organizing Maps (SOM) and evolution algorithms to analyze differences in amino acid composition of bacterial proteins. Every protein is represented as a point in twenty-dimension space on the SOM, therefore every proteome is a set of these usually thousands proteins. Then we are able to present species in multidimensional space and calculate fenetic distances between them. These distances reflect specific amino acid compositions of analyzed proteomes. Next, an evolutionary algorithm was used to find weights for particular amino acids that best distinguish the fenetic distance between a pair of proteomes. We showed that there is relationship between the weights of some amino acids. It may be related to some correlations present in amino acid composition of proteins. The presented method enables to identify amino acids that distinguish even very similar proteomes.

394 Average number of level crossings of a random polynomial with Cauchy distributed coefficients

[CS 71,(page 53)]

Somanath BAGH, *Sambalput University, Orissa India*

The present paper provides an estimate of the average number of level crossings of a random algebraic polynomial of degree n with the line $Y = CX$ where the coefficients are independent and identically Cauchy distributed random variables with the common characteristic function $\exp(-IzI)$ and C is a constant independent of x. The polynomials with Cauchy distributed coefficients are interesting as they indicate the behaviour different from a hidden class of distributions of domain of attraction of the Normal law.

395 Efficient estimation for ergodic SDE models sampled at high frequency

[CS 20,(page 21)]

Michael SØRENSEN, *Department of Mathematics Sciences, University of Copenhagen*

Simple and easily checked conditions are given that ensure rate optimality and efficiency of estimators for ergodic SDE models in a high frequency asymptotic scenario, where the time between observations goes to zero while the observation horizon goes to infinity. For diffusion models rate optimality is important because parameters in the diffusion coefficient can be estimated at a higher rate than parameters in the drift. The criteria presented in the talk provide, in combination with considerations of computing time, much needed clarity in the profusion of estimators that have been proposed for parametric diffusion models. The focus is on approximate martingale estimating functions for discrete time observations. This covers most of the previously proposed estimators, and the few that are not covered are likely to be less efficient, because non-martingale estimating functions, in general, do not approximate the score function as well as martingales.

Optimal martingale estimating functions in the sense of Godambe and Heyde have turned out to provide simple estimators for many SDE models. These estimating functions are approximations to the score functions, which are rarely explicitly known, and have often turned out to provide estimators with a surprisingly high efficiency. This can now be explained: the estimators are, under weak conditions, rate optimal and efficient in the high frequency asymptotics considered in the talk.

396 Orthogonal polynomials and polynomial kernels in dependent Dirichlet populations, in their countable representations, and in related distributions.

[CS 19,(page 20)]

Dario SPANÒ, *University of Warwick*
Robert C. GRIFFITHS, *Department of Statistics, University of Oxford, Oxford, UK*

Multivariate Jacobi polynomials, orthogonal with respect to the Dirichlet distribution on d points, are constructed by exploiting the property of right-neutrality of the Dirichlet atoms. By using the same property, multivariate Hahn polynomials, orthogonal with respect to the d-dimensional Dirichlet-Multinomial distribution, are obtained as posterior mixtures of Jacobi polynomials. Similar constructions hold in the limit as $d \to \infty$: Polynomials in the GEM distribution, in the Poisson-Dirichlet factorial moment measure, and in the Ewens sampling formula are derived.

Orthogonal polynomials can be used to characterize reversible stochastic processes with a given stationary distribution, or copulae with given marginals. This is a classical topic which dates back to the works of Bochner and Lancaster, in the middle of last century. In a multivariate setting, however, the issue is that systems of orthogonal polynomials are not unique. To overcome such a problem, one can restrict the attention to reformulations of the Bochner/Lancaster problem in terms of *polynomial kernels* of the form

$$Q_n(x, y) = \sum_{|m|=|n|} P_m(x)P_m(y),$$

where $\{P_m : m \in N^d\}$ are orthogonal Polynomial with respect to the target stationary measure. Polynomial kernels are unique. We find an explicit description of Jacobi kernels, and we show again a representation of Hahn kernels as posterior mixtures of Jacobi kernels.

By using known facts from the distribution theory, multivariate Meixner and Laguerre polynomials (on multiple Negative Binomial and Gamma distributions, respectively) are similarly constructed.

397 A picture is worth a thousand numbers: communicating uncertainties following statistical analyses

[Bernoulli Lecture,(page 28)]

David SPIEGELHALTER, *University of Cambridge and MRC Biostatistics Unit*

Subtle statistical analyses are increasingly used to analyse large and complex datasets with the intention of influencing public perceptions and government policy. And yet the level of risk literacy in society appears to remain low, with communication officers and the media often showing a general lack of understanding of statistical arguments that express uncertainty about current parameters, future events, or underlying scientific understanding.

The Gapminder initiative of Hans Rosling has led the way in showing how large datasets can be animated using modern programming techniques, with changes in time being represented by smooth animated movement. Inspired by this, we will describe attempts by ourselves and others to explicitly represent uncertainty in the conclusions of statistical analyses. Simple interval bounds over-emphasise arbitrary limits, and so we focus on graphics that display levels of confidence or probability as a continuum. Animation can be used to represent changes over time, with multiple paths indicating possible futures. These ideas will be illustrated with applications drawn from applications including finance, gambling, sport, lifestyle risks to survival, drug safety and teenage pregnancy.

This work forms part of the Winton Programme for the Public Understanding of Risk at the University of Cambridge.

398 Subjective Bayesian analysis for binomial proportion in zero-numerator problems

[CS 42,(page 38)]

Mamandur Rangaswamy SRINIVASAN, *Department of Statistics University of Madras, Chennai-600 005, TN, India*

M. SUBBIAH, *Department of Statistics University of Madras, Chennai-600005, Tn, India*

The problem of estimation of binomial parameters, witness variety of competing intervals to choose from, using Frequentist and Bayesian methods, but could still prove to be interesting. The advent of computers has made the computation of posteriors, in a Bayesian process, through powerful methods like Monte-Carlo, more realistic and programs such as BUGS easier. Beta distribution B(a, b) is a natural choice of conjugate prior for the probability of success (p) of the binomial distribution. However, care needs to be taken with the choice of the parameters of beta distribution and the usual choices of noninformative priors include a = b = 0, 0.5, or 1. In addition, small and moderate samples need suitable modification in the boundaries of the parameter space.

Tuyl et al (2008) discussed a set of informative prior in the problem of estimating binomial proportion with zero events and concluded that the use of a prior with a, b less than 1 should be avoided, both for noninformative and informative priors. A Bayesian method for constructing confidence interval for the binomial proportion is developed by considering the prior parameters as a = 1 and the choice of b is based on a historical value of the problem. The appropriateness of the fully Bayesian approach by allowing more flexibility with the additional parameter for estimating the binomial proportion with zero successes has been illustrated. In the absence of suitable historical information, an alternative prior based on, , - Rule of Three (Jovanovic and Levy, 1997 and Winkler et al, 2002) gives an approximation for an upper 95 percent confidence bound for a proportion in a zero numerator problem. The performance of the intervals has been studied under varied parametric values by making a comparative study with consensus posterior corresponding to the Bayes-Laplace prior.

References

1. Tuyl, F., Gerlach, R., and Mengersen, K., 2008, A comparison of Bayes-Laplace, Jeffreys, and other priors: The case of zero events. The American Statistician, 62, 40
2. Winkler, R.L., Smith, J.E., and Fryback, D.G., 2002, The role of informative priors in zero-numerator problems: being conservative versus being candid. The American Statistician, 56, 1
3. Jovanovic, B. D., and Levy, P. S. 1997, A Look at Rule of Three, The American Statistician, 51, 137

399 A study of generalized skew-normal distribution

[CS 76,(page 56)]

Nan-Cheng SU, *Nan-Cheng Su*

Wen-Jang HUANG, *Wen-Jang Huang*

Following the paper by Genton and Loperfido (2005. Generalized skew-elliptical distributions and their quadratic forms. Ann. Inst. Statist. Math. 57, 389-401) we say that Z has a generalized skew-normal distribution, if its probability density function (p.d.f.) is given by $f(z) = 2\phi_p(z;\mu,\Omega)\pi(z-\mu)$, $z \in R^p$, $\Omega > 0$, $\phi_p(z;\mu,\Omega)$ is the p-dimensional normal p.d.f. with mean vector μ and covariance matrix Ω, and π is a skew function, that is $0 \leq \pi(z) \leq 1$ and $\pi(-z) = 1 - \pi(z)$, $\forall z \in R^p$. First some basic properties, such as moment generating functions and moments of Z and its quadratic form, distribution under linear transformation and marginal distribution are derived. Then the joint moment generating functions of a linear compound and a quadratic form, two linear compounds, and two quadratic forms, and conditions for their independence are given. Finally explicit forms for the above moment generating functions, distributions and moments are derived when $\pi(z) = G(\alpha' z)$, where $\alpha \in R^p$ and G is one of the normal, Laplace, logistic or uniform distribution function.

400 New procedures for model selection in feedforward neural networks for time series forecasting

[CS 30,(page 27)]

SUBANAR, *Mathematics Department, Gadjah Mada University, Indonesia*
SUHARTONO, *Statistics Department, Institut Teknologi Sepuluh Nopember, Indonesia*

The aim of this paper is to propose two new procedures for model selection in Neural Networks (NN) for time series forecasting. Firstly, we focus on the derivation of the asymptotic properties and asymptotic normality of NN parameters estimator. Then, we develop the model building strategies based on statistical concepts particularly statistics test based on the Wald test and the inference of R2 incremental. In this paper, we employ these new procedures in two main approaches for model building in NN, i.e. fully bottom-up or forward scheme by using the inference of R2 incremental, and the combination between forward (by using the inference of R2 incremental) and top-down or backward (by implementing Wald test). Bottom-up approach starts with an empty model, whereas top-down approach begins with a large NN model. We use simulation data as a case study. The results show that a combination between statistical inference of R2 incremental and Wald test is an effective procedure for model selection in NN for time series forecasting.

401 The Brownian passage time model (bpt) for earthquake recurrence probabilities

[CS 3,(page 7)]

Nurtiti SUNUSI, *Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
Sutawanir DARWIS, *Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
Wahyu TRIYOSO, *Department of Geophysics, Faculty of Earth Sciences, Institut Teknologi Bandung, Indonesia*
I Wayan MANGKU, *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Indonesia*

Estimation of the time interval until the next large earthquake in a seismic source region has significant contribution to seismic hazard analysis. Conditional probabilities for recurrence times of large earthquakes are a reasonable and valid form for estimating the likelihood of future large earthquakes. in this study, we estimate the interval time for the occurrence of the next large seismic event assuming that the conditional probability of earthquake occurrence is a maximum, provided that a large earthquake has not occured in the elapsed time since the last large earthquake. We assume a probability model that is based on a simple physical model of the earthquake recycle, that is the Brownian Passage Time (BPT) model for the earthquake recurrence time intervals. In the statistics literarture, BPT model known as the Inverse Gaussian distribution. BPT model is specified by the mean time to failure and the aperiodicity of the mean (cofficient of variation). The proposed model will compared with existing known model. As an illustration, we estimate the interval time for the occurrence of the next large earthquake in the Nusatenggara region,Indonesia.

402 On estimation of response probabilities when missing data are not missing at random

[CS 45,(page 39)]

Michail SVERCHKOV, *BAE Systems IT and Bureau of Labor Statistics*

Most methods that deal with the estimation of response probabilities assume either explicitly or implicitly that the missing data are missing at random (MAR). However, in many practical situations this assumption is not valid, since the probability to respond often depends directly on the outcome value. The case where the missing data are not MAR (NMAR) can be treated by postulating a parametric model for the distribution of the outcomes before non-response and a model for the response mechanism. The two models define a parametric model for the joint distribution of the outcomes and response indicators, and therefore the parameters of these models can be estimated by maximization of the likelihood corresponding to this distribution. Modeling the distribution of the outcomes before non-response, however, can be problematic since no data is available from this distribution. In this paper we propose an alternative approach that allows to estimate the parameters of the response model without modelling the distribution of the outcomes before non-response. The approach utilizes relationships between the population, the sample and the sample complement distributions derived in Pfeffermann and Sverchkov (1999, 2003) and Sverchkov and Pfeffermann (2004).

403 Comparison of queueing networks: time properties

[CS 13,(page 13)]

Ryszard SZEKLI, *Mathematical Institute, University of Wroclaw, Poland*
Hans DADUNA, *Mathematics Department, University of Hamburg, Germany*
Pawel LOREK, *Mathematical Institute, University of Wroclaw, Poland*

For a class of unreliable queueing networks with product form stationary distribution π, we show that the existence of spectral gap (and consequently geometric ergodicity) is equivalent to the property that marginal distributions of π have light tails. Further we prove formulas which enable to translate correlation properties of the routing process in networks into correlation properties of networks. For example we show for $\mathbf{X} = (X_t)$ an ergodic Jackson network processes with a routing matrix R and $\widetilde{\mathbf{X}}$ another Jackson network processes having the same service intensities but with the routing matrix $\widetilde{R} = [\widetilde{r}_{ij}]$ such that the *traffic solutions* of the traffic equation for R and for $\widetilde{R}$ coincide, that for arbitrary real functions f, g, and the corresponding generators

$$\begin{aligned}&(f, Q^{\mathbf{X}}g)_\pi - (f, Q^{\widetilde{\mathbf{X}}}g)_\pi \\ =&\lambda/\xi_0 E_\pi(tr(W^{f,g}(X_t) \cdot diag(\xi) \cdot (R - \widetilde{R}))),\end{aligned}$$

where $diag(\xi) = (\delta_{i,j} \cdot \xi_i : 0 \le i, j \le J))$ is the diagonal matrix with entries of the probability solution (ξ_i) of the traffic equation, and $W^{f,g}(\mathbf{n}) = [g(\mathbf{n} + \mathbf{e}_i)f(\mathbf{n} + \mathbf{e}_j)]_{i,j}$. Formulas of such a form we utilize to compare speed of convergence, asymptotic variance, and to formulate dependence orderings of supermodular type, for networks. To modify routings in networks we use Peskun ordering, majorization and supermodular ordering if the set of network nodes is partially ordered.

404 Disconnection of discrete cylinders and random interlacements

[Kolmogorov Lecture,(page 41)]

Alain-Sol SZNITMAN, *ETH Zurich*

The disconnection by random walk of a discrete cylinder with a large finite connected base, i.e. the so-called problem of the "termite in a wooden beam", has been a recent object of interest. In particular it has to do with the way paths of random walks can create interfaces. In this talk we describe results concerning the disconnection of a cylinder based on a large discrete torus. We explain how this problem is related to the model of random interlacements and specifically to some of its percolative properties.

405 Some results on minimality and invariance of Markov bases

[IS 2,(page 23)]

Akimichi TAKEMURA, *Graduate School of Information Science and Technology, University of Tokyo*

We present some results on minimality and invariance of Markov bases. In particular we discuss indispensable moves and indispensable monomials of Markov bases. We mainly focus on Markov bases for hierarchical models of contingency tables.

406 Multi-dimensional Bochner's subordination

[CS 44,(page 39)]

Shigeo TAKENAKA, *Department of Applied Mathematics, Okayama University of Science*

Let $Y(t),\ t \ge 0$, be a positive stable motion of index $\beta,\ 0 < \beta < 1$, and $X(t),\ t \ge 0$, be a symmetric stable motion of index $\alpha,\ 0 < \alpha \le 2$. It is well known that the process $Z(t) = X(Y(t))$ becomes a symmetric stable motion with index $\alpha \cdot \beta$ (Bochner's subordination).

Let us consider an multi-dimensional version of the above. Fix a convex cone $V \subset R^n$. A curve $\mathbf{L}(t) \in R^n$, $t \geq 0$ is called time like curve if $\mathbf{L}(0) = O$, $\mathbf{L}(t) \in \mathbf{L}(s) + V$, $\forall t \geq s$, that is, $\mathbf{L}(t)$ is monotone increasing with respect to the partial order induced by V.

Consider V-valued positive stable motion $\mathbf{Y}(t)$ of index β, that is,

1. $\mathbf{Y}(\cdot;\omega)$ is a time like curve $a.a.\omega$.
2. $\mathbf{Y}(t) - \mathbf{Y}(s)$ is independent to $\mathbf{Y}(s)$ and is subject to the same law of $\mathbf{Y}(t-s)$.

These processes are characterized by the measures on $V \cap S^{n-1}$(H.Tanida 2003).

V-parameter stable process $X(\mathbf{t})$ of index α is called uniformly additive if

1. the restriction $X|_{\mathbf{L}}(t) = X(\mathbf{L}(t))$ for any time like curve $\mathbf{L}(\cdot)$ is an additive process.
2. the difference $X(\mathbf{t}) - \mathbf{X}(\mathbf{s})$ is subject to the same law of $X(\mathbf{t} - \mathbf{s})$, for any $\mathbf{t} \in \mathbf{s} + V$.

Such processes are characterized by the measures on $V^* \cap S^{n-1}$, where V^* means the dual cone of V (S.Takenaka 2003).

Then an anologous result holds:

$Z(t) = X(\mathbf{Y}(t))$ is a symmetric stable motion with index $\alpha \cdot \beta$.

407 Parameter estimation and bias correction for diffusion processes

[CS 24,(page 24)]

Cheng Yong TANG, *Department of Statistics Iowa State University*
Song Xi CHEN, *Department of Statistics Iowa State University*

This paper considers parameter estimation for continuous-time diffusion processes which are commonly used to model dynamics of financial securities including interest rates. To understand why the drift parameters are more difficult to estimate than the diffusion parameter as observed in many empirical studies, we develop expansions for the bias and variance of parameter estimators for two mostly employed interest rate processes. A parametric bootstrap procedure is proposed to correct bias in parameter estimation of general diffusion processes with a theoretical justification. Simulation studies confirm the theoretical findings and show that the bootstrap proposal can effectively reduce both the bias and the mean square error of parameter estimates for both univariate and multivariate processes. The advantages of using more accurate parameter estimators when calculating various option prices in finance are demonstrated by an empirical study on a Fed fund rate data.

408 Activity signature plots and the generalized Blumenthal-Getoor index

[CS 44,(page 39)]

George TAUCHEN, *Duke University*
Viktor TODOROV,

We consider inference about the activity index of a general Ito semimartingale; the index is an extension of the Blumenthal-Getoor index for pure-jump Levy processes. We define a new concept termed the quantile activity signature function, which is constructed from discrete observations of a process evolving continuously in time. Under quite general regularity conditions, we derive the asymptotic properties of the function as the sampling frequency increases and show that it is a useful device for making inferences about the activity level of an Ito semimartingale. A simulation study confirms the theoretical results. One empirical application is from finance. It indicates that the classical model comprised of a continuous component plus jumps is more plausible than a pure-jump model for the spot US/DM exchange rate over 1986–1999. A second application pertains to internet traffic data at NASA servers. We find that a pure-jump model with no continuous component and paths of infinite variation is appropriate for modeling this data set. These two quite different empirical outcomes illustrate the discriminatory power of the methodology.

409 Excursion sets of stable random fields

[IS 5,(page 18)]

Jonathan TAYLOR, *Stanford University*
Robert J. ADLER, *Technion*
Gennady SAMORODNITSKY, *Cornell University*

Studying the geometry generated by Gaussian and Gaussian related random fields via their excursion sets is now a well developed and well understood subject. The purely non-Gaussian scenario has however, not been studied at all. We look at three classes of stable random fields, and obtain asymptotic formulae for the mean values of various geometric characteristics of their excursion sets over high levels.

While the formulae are asymptotic, they contain enough information to show that not only do stable random fields exhibit geometric behaviour very different from that of Gaussian fields, but they also differ significantly among themselves.

410 On the extension of random mappings

[CS 71,(page 53)]

Dang Hung THANG, *Department of Mathematics,Mechanics and Informatics University of Natural Sciences Vietnam National University*

Let X and Y be separable metric spaces. By a random mappings from X into Y we mean a rule Φ that assigns to each element $x \in X$ an unique Y-valued random variable Φx . A random mapping can be regarded as an action which transforms each deterministic input $x \in X$ into a random output random Φx.

Taking into account many circumstances in which the inputs are also subject to the influence of a random enviroment there arises the need to define the action of Φ on some random inputs, i.e. to extend the domain of Φ to some class $D(\Phi)$ of X-valued random variables such that the class $D(\Phi)$ must be as wide as possible and at the same time the extension of Φ should enjoy many good properties similar to those of Φ.

In this talk, some procedures for extending random mappings will be proposed. Some conditions under which a random mapping can be extended to apply to all X -valued random variables will be presented.

411 A study on semi-logistic distribution

[CS 41,(page 37)]

Thomas Mathew THAZHAKUZHIYIL, *M D College, Pazhanji, Thrissur, Kerala, India-680542*

Semi logistic and generalized semi logistic distributions are studied. These distributions can be useful for modeling real data that exhibit periodic movements and in situations in which the logistic and generalized logistic become unrealistic. It is shown that among the distributions on R, semi logistic distribution is the only distribution that possesses maximum stability with respect to geometric summation. First order autoregressive semi logistic process is introduced and its properties are studied. A generalization of this model is considered. Other characterizations of semi logistic and generalized semi-logistic distributions are obtained. Applications of some of these results in time series modelling are discussed. The Marshall-Olkin scheme of introducing an additional parameter in to a family of distribution is generalized by introducing a second parameter. The logistic distribution is generalized using the Marshall-Olkin scheme and its generalization. Marshall-Olkin logistic and Marshall-Olkin semi-logistic distributions are studied. First order autoregressive time series models with these distributions as marginals are developed and studied. Some applications are discussed. Bivariate semi-logistic and Marshall-Olkin bivariate semi-logistic distributions are considered. Some properties and characterizations of these distributions are studied. First order autoregressive processes with bivariate semi-logistic and Marshall-Olkin bivariate semi-logistic distributions as marginals are introduced and studied.

412 Marshall-Olkin logistic processes

[CS 67,(page 49)]

Alice Thomas THERMADOM, *Department of Statistics, Vimala College, Thrissur, Kerala-680009, India*

Kanichukattu Korakutty JOSE, *Department of Statistics, St. Thomas College, Pala, Kerala-686574, India*

Miroslav M RISTIČ, *Faculty of Sciences and Mathematics, University of Nis, Serbia*

Ancy JOSEPH, *Department of Statistics, B.K. College, Amalagiri, Kerala-686036, India*

Marshall-Olkin univariate logistic and semi-logistic distributions are introduced and studied. Autoregressive time series models of order 1 as well as k (AR(1) and AR(k)) are developed with minification structure, having these stationary marginal distributions. Various characterizations are also obtained. Bivariate logistic and semi-logistic processes are considered and their characterizations are obtained. The un-known parameters of the processes are estimated and some numerical results of the estimations are given.

413 Genome sharing among related individuals: an approximate answer to the right question.

[Tukey Lecture,(page 3)]

Elizabeth A. THOMPSON, *Department of Statistics, University of Washington*

Similarities among individuals for traits determined in whole or in part by their DNA arise from their coancestry. Related individuals share segments of their genome, in the sense that these segments derive from a single DNA segment in a common ancestor. Such segments are said to be identical-by-descent or ibd and have high probability of being of the same biochemical type. While a known pedigree relationship gives a probability distribution on the marginal probability of ibd and on lengths of ibd segments, modern genomic data permits much more precise inference of shared genome. However, human individuals are diploid: even a pair of individuals have a total of four haploid genomes. Thus, at a minimum, models for sharing among four genomes are required, and these models must be such that conditional probabilities of ibd segments given dense genomic data can be computed. A new model for genome sharing along four genomes will be described, and the resulting inferences of ibd segments illustrated.

414 Some approaches to parallel computing in R

[IS 1,(page 5)]

Luke TIERNEY, *Statistics and Actuarial Science, University of Iowa*

This talk outlines two approaches to adding parallel computing support to the R statistical computing environment. The first approach is based on implicitly parallelizing basic R operations, such as vectorized arithmetic operations; this is suitable for taking advantage of multi-core processors with shared memory. The second approach is based on developing a small set of explicit parallel computation directives and is most useful in a distributed memory framework.

415 Most powerful test for fuzzy hypotheses testing using r-Level Sets

[CS 34,(page 32)]

Hamzeh TORABI, *Department of Statistics, Yazd University, Yazd, Iran*
Eisa MAHMOUDI, *Department of Statistics, Yazd University, Yazd, Iran*

Decision making in the classical statistical inference is based on crispness of data, random variables, exact hypotheses, decision rules and so on. As there are many different situations in which the above assumptions are rather irrealistic, there have been some attempts to analyze these situations with fuzzy set theory proposed by Zadeh.

One of the primary purpose of statistical inference is to test hypotheses. In the traditional approach to hypotheses testing all the concepts are precise and well defined. However, if we introduce vagueness into hypotheses, we face quite new and interesting problems.

In this paper, we redefine some concepts about fuzzy hypotheses testing, and then we give a new version of Neyman-Pearson lemma for fuzzy hypotheses testing using r-levels. Finally, we give some examples.

416 Erdos-Renyi random graphs + forest fires = self-organized criticality

[IS 30,(page 10)]

Balint TOTH, *Institute of Mathematcs, Budapest University of Technology*
Balazs RATH, *Institute of Mathematcs, Budapest University of Technology*

We modify the usual Erdos-Renyi random graph evolution by letting connected clusters 'burn down' (i.e. fall apart to disconnected single sites) due to a Poisson flow of lightnings. In a range of the intensity of rate of lightnings the system sticks to a permanent critical state. The talk will be based on joint work with Balazs Rath.

417 Extreme values statistics for Markov chains via the (pseudo-) regenerative method

[CS 61,(page 47)]

Jessica TRESSOU, *INRA-Met@risk & HKUST-ISMT*
Patrice BERTAIL, *University Paris X & CREST-LS*
Stephan CLÈMENÇON, *Telecom Paristech*

This presentation introduces specific statistical methods for extremal events in the markovian setup, based on the regenerative method and the Nummelin technique.

Exploiting ideas developed in Rootzén (1988) [*Adv. Appl. Probab.* 20:371–390], the principle underlying our methodology consists of generating first (a random number l of) approximate pseudo-renewal times $\tau_1, \tau_2, \ldots, \tau_l$ for a sample path $X_1, \ldots, X_n$ drawn from a Harris chain X with state space E, from the parameters of a *minorization condition* fulfilled by its transition kernel, and computing then submaxima over the *approximate cycles* thus obtained: $\max_{1+\tau_1 \leq i \leq \tau_2} f(X_i), \ldots, \max_{1+\tau_{l-1} \leq i \leq \tau_l} f(X_i)$ for any measurable function f.

Estimators of tail features of the sample maximum $\max_{1 \leq i \leq n} f(X_i)$ are then constructed by applying standard statistical methods, tailored for the i.i.d. setting, to the submaxima as if they were independent and identically distributed. In particular, the asymptotic properties of extensions of popular inference procedures based on (conditional) maximum likelihood theory, such as the Hill estimator for the tail index in the case of the Fréchet maximum domain attraction, are thoroughly investigated. Using the same approach, we also consider the problem of estimating the extremal index of the sequence $\{f(X_n)\}_n$ under suitable assumptions.

Practical issues related to the application of this methodology are discussed and some simulation results are displayed.

A preprint of the paper is available at http://hal.archives-ouvertes.fr/hal-0165652.

418 Nonparametric meta-analysis for identifying signature genes in the integration of multiple genomic studies

[IS 24,(page 10)]

George C. TSENG, *Department of Biostatistics University of Pittsburgh USA*
Jia LI, *Department of Biostatistics University of Pittsburgh USA*

With the availability of tons of expression profiles, the need for meta-analyses to integrate multiple microarray studies is obvious. For detection of differentially expressed genes, most of the current efforts are focused on comparing and evaluating gene lists obtained from each individual dataset. The statistical framework is often not rigorously formulated and a real sense of information integration is rarely performed. In this paper, we tackle two often-asked biological questions:What are the signature genes significant in one or more data sets? and What are the signature genes significant in all data sets?. We illustrated two statistical hypothesis settings and proposed a best weighted statistic and a maximum p-value statistic for the two questions, respectively. Permutation analysis is then applied to control the false discovery rate. The proposed test statistic is shown to be admissible. And we further show the advantage of our proposed test procedures over existing methods by power comparison, simulation study and real data analyses of a multiple-tissue energy metabolism mouse model data and prostate cancer data sets.

419 Efficient sparse recovery with no assumption on the dictionary

[IS 22,(page 23)]

Alexandre B. TSYBAKOV, *CREST and University of Paris 6*

There exist two main approaches to sparse statistical estimation. The first one is the BIC: it has nice theoretical properties without any assumption on the dictionary but is computationally infeasible starting from relatively modest dimensions. The second one is based on the Lasso or Dantzig selector that are easily realizable for very large dimensions but their theoretical performance is conditioned by severe restrictions on the dictionary. The aim of this talk is to propose a new method of sparse recovery that realizes a compromise between the theoretical properties and computational efficiency. The theoretical performance of the method is comparable with that of the BIC in terms of sparsity oracle inequalities for the prediction risk. No assumption on the dictionary is needed, except for the standard normalization. At the same time, the method is computationally feasible for relatively large dimensions. It is designed using the exponential weighting with suitably chosen priors, and its analysis is related to the PAC-Bayesian methodology in statistical learning. We obtain some new PAC-Bayesian bounds with leading constant 1 and we develop a general technique to derive sparsity oracle inequalities from the PAC-Bayesian bounds. This is a joint work with Arnak Dalalyan.

420 Stochastic flows, planar aggregation and the Brownian web

[CS 27,(page 25)]

Amanda TURNER, *Lancaster University, UK*
James NORRIS, *University of Cambridge, UK*

Diffusion limited aggregation (DLA) is a random growth model which was originally introduced in 1981 by Witten and Sander. This model is prevalent in nature and has many applications in the physical sciences as well as industrial processes. Unfortunately it is notoriously difficult to understand, and only one rigorous result has been proved in the last 25 years. We consider a simplified version of DLA known as the Eden model which can be used to describe the growth of cancer cells, and show that under certain scaling conditions this model gives rise to a limit object known as the Brownian web.

421 Quantum entangled states generation and nonlinear localized modes in coupled Kerr nonlinear chains

[CS 27,(page 25)]

B.A. UMAROV, *Department of Computational and Theoretical Sciences, Faculty of Science, International Islamic University Malaysia, Jalan Istana,,Bandar Indera Mahkota, 25200, Kuantan Pahang,, Malaysia*
M.R.B. WAHIDDIN, *Cyber Security Lab, MIMOS Berhad, Technology Park Malaysia, 57000, Kuala Lumpur, Malaysia*

The investigation of the nonclassical properties of light propagating in nonlinear optical systems is currently the subject much theoretical and experimental efforts in quantum optics. The third order (Kerr) nonlinearity was among the first proposed for generation of squeezed light, for quantum nondemolition measurements, and recently for entangled states generation [1-3]. In this paper we will consider the system consisting of the three coupled nonlinear materials with Kerr nonlinearity. The coupling is realized via evanescent overlaps of the modes. The interaction Hamiltonian in the rotating-wave approximation can be written as

$$\hat{H} = \hbar\omega \sum_{i=1}^{3} (\hat{a}_i^\dagger \hat{a}_i + g\hat{a}_i^\dagger \hat{a}_i^\dagger \hat{a}_i \hat{a}_i) + \hbar k(\hat{a}_1^\dagger \hat{a}_2 + \hat{a}_2^\dagger \hat{a}_1 + \hat{a}_3^\dagger \hat{a}_2 + \hat{a}_2^\dagger \hat{a}_3),$$

where $\hbar$ is the Planck constant, ω is the frequency common for all three waveguides and g is the coupling constant proportional to the third order susceptibility $\chi^{(3)}$ responsible for the self-action process (we do not consider the nonlinear coupling or the cross-action between the two modes), k is the linear exchange coupling coefficient between waveguides and $\hat{a}_i$ are the photon annihilation operators in the ith waveguides, respectively. Here $i = 1, 2, 3$ determines the channel in the NLDC. This Hamiltonian describes the behavior of continuous wave (CW) fields in NLC in coupled-mode approximation. By application of standard techniques for the given Hamiltonian we obtain the master equation for the density matrix of the system, which in positive P representation [4] can be converted to the Fokker-Planck equation for the quasiprobability distribution function $P(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3)$.Using the Ito rules one can obtain from the Fokker-Planck equation the Langevin stochastic equations for the α_i and β_i variables with complex Gaussian noise terms. These equations were numerically integrated and averaged over many trajectories to get the information about system. The inseparability of density matrix criteria obtained in [5] was applied to check and to show the possibility of genuine tripartite entangled states generation and also bipartite entangled states in the system under consideration

References

1. W.Leonski and A.Miranowicz, J. Opt. B: Quant. Semiclass. Opt. 6, S37 (2004),
2. R.S.Said, M.R.B.Wahiddin and B.A.Umarov, J. Phys. B: At. Mol.Opt. Phys. 39 (2006) 1269-1274.
3. M.K.Olsen, arXiv:quant-ph/0603037, (2006).
4. P.D. Drummond and C.W.Gardiner, J.Phys. A:13, 2353 (1980)
5. P. van Loock and A.Furusawa, Phys.Rev.A 67, 052315, (2003)

422 Multiple Stratonovich integral and Hu-Meyer formula for Lévy processes

[CS 44,(page 39)]

Frederic UTZET, *Universitat Autonoma de Barcelona*
Merce FARRE, *Universitat Autonoma de Barcelona*
Maria JOLIS, *Universitat Autonoma de Barcelona*

Combining the main ideas of Hu and Meyer [1] and Rota and Wallstrom [3], we will present an Ito multiple integral and a Stratonovich multiple integral with respect to a Lévy processes with finite moments up to a convenient order. The Stratonovich multiple integral is an integral with respect to a product measure, and the Ito multiple integral is the integral with respect to a measure that give zero mass to the diagonal sets, like $\{(s_1, \ldots, s_n) \in R_+^n,\ s_1 = s_2\}$. The principal tool is the powerful combinatorial machinery introduced by Rota and Wallstrom for random measures, where the diagonal sets of R_+^n are identified with the partitions of the set $\{1, \ldots, n\}$. In this context, the key point is to understand how the product of stochastic measures works on the diagonal sets, and that leads to the diagonal measures defined by Rota and Wallstrom. For a Lévy process those measures are related to the powers of the jumps of the process, and hence with a family of martingales introduced by Nualart and Schoutens [2], called Teugels martingales, which enjoy very nice properties. With all these ingredients we prove a general Hu–Meyer formula. As particular cases, we deduce the classical Hu–Meyer formulas for the Brownian motion and for the Poisson process.

References

[1] Hu, Y. Z. and Meyer, P. A., Sur les intégrales multiples de Stratonovitch, Séminaire de Probabilités XXII (1988) pp. 72–81 Lecture Notes in mathematics, 1321, Springer, New York, 1988

[2] Nualart, D. and Schoutens, W., Chaotic and predictable representation for Lévy processes. *Stochastic Process. Appl.* **90** (2000) 109–122.

[3] Rota, G-C. and Wallstrom, T. C., Stochastic integrals: A combinatorial approach, Ann. prob., **25** (1997) 1257–1283.

423 Stability and entropy of statistical distributions

[CS 38,(page 34)]

VIJAYAKUMAR, *Multimedia University*

This paper discusses the effect of entropy in statistical distributions. In particular, we will discuss the physical implications of the characteristic functions (moment-generating functions) of Gaussian distributions using concepts of entropy. These distributions are a special case of Levy stable distributions with finite variance. We have used simulated data for illustration.

424 Extending the two-sample empirical likelihood method

[CS 56,(page 44)]

Janis VALEINIS, *University of Latvia, Zellu 8, LV-1002, Riga, Latvia*
Axel MUNK, *Georg-August University of Göttingen, Göttingen, Germany*
Edmunds CERS, *University of Latvia, Riga, Latvia*

Since Owen (1988, 1990) has introduced the empirical likelihood method for statistical inference, there have been several attempts to generalize it for the two-sample case. Qin and Zhao (2000) established the empirical likelihood method for mean and distribution function differences in the two-sample case. In the PhD thesis of Valeinis (2007) it has been shown that this result can be applied also for P-P, Q-Q plots, ROC curves and structural relationship models. Moreover, this setup basically generalizes the results of Claeskens *et al.* (2003), Jing and Zhou (2003), where ROC curves and quantile differences have been analyzed in the two-sample case.

Consider the two-sample problem, where i.i.d. random variables $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are independent and have some unknown distribution functions F_1 and F_2, respectively. Assume we are interested in making inference for some function $t \to \Delta(t)$ defined on an interval T (further on we write Δ). Let θ_0 be some univariate parameter associated with one of the distributions F_1 or F_2. We assume that all information about θ_0, Δ, F_1 and F_2 is available in the known form of unbiased estimating functions, i.e.,

$$E_{F_1} w_1(X, \theta_0, \Delta, t) = 0, \quad E_{F_2} w_2(Y, \theta_0, \Delta, t) = 0.$$

If $\Delta = \theta_1 - \theta_0$, where θ_0 and θ_1 are univariate parameters associated with F_1 and F_2 respectively we have exactly the setup of Qin and Zhao (2000). The reason of our formulation of the problem is to have some flexibility in the function Δ.

For example, let $\theta_0 = F_2^{-1}(t)$ and $\Delta = F_1(F_2^{-1}(t))$, which is the P-P plot of functions F_1 and F_2. In this case

$$w_1(X, \theta_0, \Delta, t) = I_{\{X \le \theta_0\}} - \Delta,$$
$$w_2(Y, \theta_0, \Delta, t) = I_{\{Y \le \theta_0\}} - t.$$

It is well known that in case of quantiles the smoothed empirical likelihood method has additional advantage (see Chen and Hall (1993)). Thus, additionally we establish the smoothed empirical likelihood method for P-P and Q-Q plots and simulate their pointwise confidence bands. We end our analysis by constructing simultaneous confidence bands for P-P and Q-Q plots for real data combining the empirical likelihood and the bootstrap method (see Owen and Hall (1993)).

425 Empirical comparisons of computer models for stellar evolution

[IS 4,(page 55)]

David A. VAN DYK, *University of California, Irvine, California, USA*
Steven DEGENNARO, *University of Texas at Austin, Texas, USA*
Ted VON HIPPEL, *University of Texas at Austin, Texas, USA*
William JEFFERY, *University of Vermont, Burlington, Vermont, USA*
Nathan STEIN, *University of Texas at Austin, Texas, USA*
Elizabeth Jeffery JEFFERY, *University of Texas at Austin, Texas, USA*

Color Magnitude Diagrams (CMDs) are plots that compare the magnitudes (luminosities) of stars in different wavelengths of light (colors). High non-linear correlations among the mass, color, and surface temperature of newly formed stars induce a long narrow curved point cloud in a CMD that is known as the main sequence. After millions or billions of years, depending on the initial mass of the star, the physical processes that cause a star to shine change. This in turn causes dramatic shifts in the color, spectrum, and density of stars. These aging stars form new CMD groups that correspond to red giants and later, white dwarfs. The physical processes that govern stellar formation and evolution are studied with complex computer models that are used to predict the plotted magnitudes on a set of CMDs as a function of parameters of scientific interest such as distance, stellar age, mass, and metallicity (a measure of the abundance of elements heaver than Helium). Here, we describe how we use these computer models as a component in a complex likelihood function and how we use Bayesian methods to fit the resulting statistical model in order to evaluate and compare the stellar evolution models.

We focus on developing methods for the analysis of CMDs of the stars in a so-called open cluster. Stars in these clusters were formed from the same molecular cloud at roughly the same time. This simplifies statistical analysis because we expect the stars to have the same metallicity and age; only their masses differer. Unfortunately, the data are contaminated with stars that are in the same line of sight as the cluster but are not part of the cluster. Because these stars are of different ages and metallicities than the cluster stars, their coordinates on the CMDs are not well predicted from the computer models. A second complication arises from multi-star systems in the cluster. These stars are of the same age and metallicity as the cluster, but we typically cannot resolve the individual stars in the system and thus observe only the sums of their magnitudes in different colors. This causes these systems to appear systematically offset from the main sequence. Because the offset is informative as to the individual stellar masses we can formulate a model so as to identify the masses. In this talk we show how Bayesian analysis of appropriate highly-structured models can overcome these challenges.

426 Empirical likelihood for non-smooth criterion functions

[CS 56,(page 44)]

Ingrid VAN KEILEGOM, *Universite catholique de Louvain*

Elisa MOLANES LOPEZ, *Universidad Carlos III de Madrid*

Noel VERAVERBEKE, *Universiteit Hasselt*

Suppose that $X_1, \ldots, X_n$ is a sequence of independent random vectors, identically distributed as a d-dimensional random vector X. Let $\mu \in R^p$ be a parameter of interest and $\nu \in R^q$ be some nuisance parameter. The unknown, true parameters (μ_0, ν_0) are uniquely determined by the system of equations $E\{g(X, \mu_0, \nu_0)\} = 0$, where $g = (g_1, \ldots, g_{p+q})$ is a vector of $p + q$ functions. In this paper we develop an empirical likelihood method to do inference for the parameter μ_0. The results in this paper are valid under very mild conditions on the vector of criterion functions g. In particular, we do not require that $g_1, \ldots, g_{p+q}$ are smooth in μ or ν. This offers the advantage that the criterion function may involve indicators, which are encountered when considering e.g. differences of quantiles, copulas, ROC curves, to mention just a few examples. We prove the asymptotic limit of the empirical log-likelihood ratio, and carry out a small simulation study to test the performance of the proposed empirical likelihood method for small samples.

427 The Bouligand influence function: checking robustness of support vector machines

[CS 70,(page 52)]

Arnout VAN MESSEM, *Vrije Universiteit Brussel (Belgium)*

Andreas CHRISTMANN, *University of Bayreuth (Germany)*

A short introduction to support vector machines will be given. These kernel methods are inspired by convex risk minimization in infinite dimensional Hilbert spaces.

First we will propose the Bouligand influence function (BIF) as a new concept for robust statistics. The BIF is a modification of F.R. Hampel's influence function (IF) and is based on a special cone derivative (the Bouligand derivative) instead of the usual Gâteaux-derivative. There exists a nice relationship between the BIF and the IF: if the BIF exists, then the IF does also exist and both are equal. The usefulness of these Bouligand-derivatives to robust statistics is explained.

In the second part of the talk we apply the BIF to support vector machines based on a non-smooth loss function for which the classical influence function was unknown. We show for the regression case that many support vector machines based on a Lipschitz continuous loss function and a bounded kernel have a bounded BIF and hence also a bounded IF. In this respect such SVMs are therefore robust. Special cases are SVMs based on the ϵ-insensitive loss, Huber's loss, and kernel based quantile regression based on the pinball loss, all used in combination with a bounded kernel such as the classical Gaussian RBF kernel.

References

1. Christmann, A. and Steinwart I. (2007). Consistency and robustness of kernel based regression in convex risk minimization. *Bernoulli*, **13**, 799-819.
2. Christmann, A. and Van Messem A. (2007). Bouligand derivatives and robustness of support vector machines for regression. *Tentatively accepted*.
3. Robinson, S.M. (1991). An implicit-function theorem for a class of nonsmooth functions. *Mathematics of Operations Research*, **16**, 292-309.
4. Schölkopf, B. and Smola, A. (2002). *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, Cambridge, Massachusetts.
5. Vapnik, V. (1998). *Statistical Learning Theory.* Wiley & Sons, New York.

428 On a randomized poly-nuclear growth model with a columnar defect

[CS 73,(page 54)]

Maria Eulalia VARES, *CBPF - Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, BRAZIL*

The starting point of the present work is the following question: how localized microscopic defect can affect macroscopic behavior of a growth system? This is an important question in non-equilibrium growth: is the asymptotic shape changed (faceted) in the macroscopic neighborhood of such a defect at any value of its strength, or, when the defect is too weak, then the fluctuations of the bulk evolution become predominant and destroy the effects of the obstruction in such a way that its presence becomes macroscopically undetectable. Such a vanishing presence of the macroscopic effect as a function of the strength of obstruction represents what is called dynamic phase transition. The existence of such a transition, its scaling properties, the shape of the density profile near the obstruction, and also whether information percolates through the obstruction, are among of the most important issues. In many cases these questions can be translated into the language of pinning of directed polymer in presence of bulk disorder, or to questions if in a driven flow presence of a static obstruction, such as a slow bond in one dimensional totally asymmetric simple exclusion process (TASEP), always results in change of a current, as it is predicted by mean-field theory. Similar questions are often raised in the setting of first-passage percolation (FPP).

The process that we introduce and study in this work, the Randomized Poly-Nuclear Growth (RPNG), is a variant of poly-nuclear growth in 1+1 dimensions, where the level boundaries of the growing droplet perform continuous-time, simple symmetric random walks. This dynamics corresponds to the modified Glauber dynamics for the two-dimensional Ising model on a solid surface at temperature 0, with additional nucleation.

We study how the asymptotic behavior of the speed of growth is affected by the presence of a columnar defect placed along one fixed line, and prove that the system undergoes dynamical phase transition; there is a non-trivial transition in the strength of the perturbation, above which the law of large numbers for the height function is modified. We also identify the shape of the faceted (macroscopic) region. The talk is based on the joint work with V. Beffara and V. Sidoravicius.

429 A goodness-of-fit process for ARMA(p,q) time series models based on a modified residual autocorrelation sequence

[CS 39,(page 34)]

Santiago VELILLA, *Universidad Carlos III de Madrid*

The asymptotic distribution of the usual goodness-of-fit process for ARMA(p, q) models, introduced by Durbin (1975), is obtained rigourously, using techniques of weak convergence in the space C$[0, 1]$. As it turns out, this process, that depends on the standard residual autocorrelations, converges weakly to a Gaussian limit process whose covariance function depends on unknown parameters. This fact motivates the consideration of a modified process based on a suitable transformation of the residual autocorrelations. This new process is shown to converge weakly to the Brownian bridge. Thus, functionals based on it are adequate for goodness-of-fit purposes. The behavior of these functionals is analyzed and compared, by simulation, to that of the standard Ljung-Box and cumulative periodogram statistics. The new method seems to improve over standard procedures, at the cost involved in computing the transformed autocorrelation sequence.

430 A decomposition of Pearson-Fisher and Dzhaparidze-Nikulin chi-squared goodness-of-fit statistics and some ideas for a construction of the more powerful test

[CS 17,(page 19)]

Vassilly VOINOV, *KIMEP*

An explicit decomposition of the Pearson-Fisher test, based on minimum chi-squared or equivalent estimates of parameters for grouped data, and a decomposition of the Dzhaparidze-Nikulin test, based on any $\sqrt{n}$-consistent estimates for raw data, on independent components distributed in the limit as chi-square with one degree of freedom is presented. Despite of the difference of those tests, the decomposition is formally the same because of the similarity of their limit distributions. The decomposition is valid for any partitioning of a sample space and helps to realize the idea of Cochran - to use the most powerful with respect to a specified alternative component or group of components for compound hypotheses testing. Some numerical examples illustrating the idea are given. Another way to construct more powerful tests is to use vector-valued statistics, components of which can be not only different scalar tests based on the same sample, but also scalar tests based on components or groups of components of the same statistic. Numerical examples, which illustrate the idea are presented.

431 Asymptotic probability for the deviations of dependent Bootstrap means from the sample mean

[CS 37,(page 33)]

Andrei VOLODIN, *University of Regina, Canada*
Kamon BUDSABA, *Thammasat University, Thailand*
Jiraroj TOSASUKUL, *Thammasat University, Thailand*

The asymptotic probability for the deviations of dependent bootstrap means from the sample mean is obtained, without imposing any assumptions on joint distribution of the original sequence of random variables from which the dependent bootstrap sample is selected. A nonrestrictive assumption of stochastic domination by a random variable is imposed on the marginal distributions of this sequence. Coverage probabilities and lengths of confidence intervals based on the dependent bootstrap procedure are compared to those based on the traditional bootstrap and standard normal theory.

432 Comparing Bayesian and frequentist approaches to analyzing genome-wide association studies

[PS 2,(page 17)]

Damjan VUKCEVIC, *Department of Statistics, University of Oxford*
Jonathan MARCHINI, *Department of Statistics, University of Oxford*
Chris HOLMES, *Department of Statistics, University of Oxford*
Peter DONNELLY, *Swiss Federal Institute of Technology Lausanne*

Due to recent technological advances, genome-wide association studies (GWAS) have become a tool of choice for investigating the genetic basis of common complex human diseases. These studies involve collecting DNA from a large number of individuals, and reading their genetic types at a very large number of places along the genome (‘typing’ their ‘alleles’ at a very large number of ‘loci’ along the genome). A typical large study would sample a few thousand individuals, each typed at hundreds of thousands of loci, giving rise to very large datasets. Here we focus only on the case-control design, the most common one for GWAS.

Analysis of GWAS typically first proceeds by testing each typed locus individually for an association with the disease of interest. (Short-range correlations across the genome will mean that untyped loci that are associated with the disease can still be detected using nearby typed loci, albeit with reduced power.)

To date, most GWAS have adopted the frequentist paradigm, reporting p-values as measures of evidence of association. In particular, the χ^2 one degree of freedom (Cochran-Armitage) trend test is quite commonly used. This is the score test for a logistic regression model where each additional copy of the risk allele within a locus additively increases the log-odds of developing the disease.

We develop a Bayesian model based on this additive logistic regression model, and use the Bayes factor (BF) as measure of evidence of association. Using data from a large GWAS, we compare the BF and the p-value, highlighting the effect of varying power across loci on both. We determine a family of priors which give the same rankings of loci, when ranked by either the BF or the p-value. This directly highlights some of the assumptions inherent in the trend test. In particular, we see that larger genetic effects are assumed at loci with rarer genetic variants in the population, with the prior variance on the effect size going to infinity as the frequency of the risk allele goes to zero.

433 A limit theorem for the trajectory of a particle in the Markov approximation of the Lorentz model

[CS 74,(page 54)]

Vladislav VYSOTSKY, *St.Petersburg State University*

We study the following model of motion of particle through a random medium in R^d under acceleration of a constant external field. Initially, the particle stays at the origin and has a deterministic speed v_0. It starts to move with the constant acceleration a; from time to time, the particle experiences collisions with obstacles that form the random medium. Let η_n be the length of the particle's trajectory between the nth and the $(n+1)$st collisions, and let η_0 be the length of the trajectory before the first collision. We assume that $\{\eta_n\}_{n\geq 0}$ are i.i.d. exponentially distributed r.v.'s with the mean λ that is a parameter equal to the mean free path of the particle. Collisions with obstacles happen as follows. Let $\{\sigma_n\}_{n\geq 1}$ be i.i.d. random vectors that are uniformly distributed on the unit sphere $S^{d-1} \subset R^d$, and let $\{\eta_n\}_{n\geq 0}, \{\sigma_n\}_{n\geq 1}$ be independent. We assume that at the nth collision, the particle's speed V_n changes to $V_n - \frac{1+\alpha}{2}\big(V_n + |V_n|\sigma_n\big)$, where $\alpha \in [0,1]$ is the restitution coefficient.

This model of motion is introduced as the Markov approximation of the classical Lorentz model. The latter describes a particle moving through a medium consisting of immobile spherical obstacles, randomly distributed in R^d; the particle is driven by a constant field a, and at collisions with obstacles, it inelastically reflects with the restitution coefficient α. We show how our Markov approximation is deduced from the conditions of the Lorentz model. We explain in which sense the models are close to each other and that the Markov approximation is the limiting case of the Lorentz model.

Our main goal is to study the asymptotics of $X(T)$ as $T \to \infty$, where $X(T)$ denotes the particle's position at time T in the Markov approximation model. Assuming that $0 < \alpha < 1$, we prove that, first, the particle drifts with constant speed $v \in R^d$ in the direction of the field a, i.e., $\frac{X(T)}{T} \xrightarrow{P} v$. Second, $T^{-1/2}(X(sT) - vsT) \xrightarrow{d} \big(c_1W_1(\cdot), \ldots, c_1W_{d-1}(\cdot), c_2W_d(\cdot)\big)$ in $C[0,1]$, where W_i are independent Wiener processes, $c_{1,2}$ are some constants, and the basis of R^d is such that $v = (0, \ldots, 0, |v|)$. This result significantly differs from the one of Ravishankar and Triolo (1999) who studied a similar model with elastic collisions ($\alpha = 1$). Our proof is based on theory of Markov chains.

434 Forecasting stochastic loss ratio using Kalman filter:bivariate case

[CS 39,(page 34)]

Sri WAHYUNINGSIH, *Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
Sutawanir DARWIS, *Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
Agus Yodi GUNAWAN, *Industrial and Financial Mathematics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia*
Asep Kurnia PERMADI, *Reservoir Engineering Research Division, Faculty of Petroleum Engineering, Institut Teknologi Bandung, Indonesia*

The Arps hyperbolic decline equation has been used for many years to predict the well performance from the prediction of production data. Many different methods have been developed to determine the hyperbolic decline curve parameter. The hyperbolic decline can be recognized by the fact that it has stochastic loss ratios. The previous results of our research showed that the univariate Kalman filter approach is a good approach to predict next observation having stochastic loss ratios. In this talk, we apply Kalman Filter approach to predict the next observation in bivariate case. The present results will be compared with univariate case in which two wells are analysed separately. As an illustration, we shall apply the technique to predict next production of geothermal production data from two wells which is relatively close. The results show that the production prediction with stochastic loss ratio using bivariate Kalman filter is more realistic compared to univariate case.

435 Shrinkage tuning parameter selection with a diverging number of parameters

[CS 46,(page 41)]

Hansheng WANG, *Peking University*
Bo LI, *Tsinghua University*
Chenlei LENG, *National University of Singapore*

Contemporary statistical research frequently deals with problems involving a diverging number of parameters, for which various shrinkage methods (e.g., LASSO, SCAD, etc) are found particularly useful for the purpose of variable selection (Fan and Peng, 2004; Huang, Ma, and Zhang, 2007). Nevertheless, the desirable performances of those shrinkage methods heavily hinge on an appropriate selection of the tuning parameters. With a fixed predictor dimension,Wang, Li, and Tsai (2007) and Wang and Leng (2007) demonstrated that the tuning parameters selected by a BIC-type criterion is able to identify the true model consistently. In this work, similar results are further extended to the situation with a diverging number of parameters for both unpenalized and penalized estimators (Fan and Peng, 2004, Huang, Ma, and Zhang, 2007). As a result, our theoretical results further enlarges not only the applicable scope of the traditional BIC criterion but also that of those shrinkage estimation methods (Tibshirani, 1996; Huang, Ma, and Zhang, 2007; Fan and Li, 2001; Fan and Peng, 2004).

436 Genetic substructure and population stratification in a multiethnic breast cancer study

[CS 9,(page 12)]

Hansong WANG, *Cancer Research Center of Hawaii*
Dan STRAM, *University of Southern California*

Population substructure may lead to confounding in case-control association studies. We analyzed the genetic substructure of a multiethnic breast cancer study consisting of five ethnic groups: Caucasians, African Americans, Latin Americans, Japanese and Native Hawaiians, based on about 1400 tagSNPs selected in 60 candidate genes. Our results indicate that self-reported ethnicity information represents individual genetic composition for the majority of the study population. A wide range of European admixture is observed for African Americans, Latinos and Hawaiians in this study. The genetic structure did not distort the p-value distribution for the breast cancer association study within each ethnic group or for all groups combined. For studies of other disease traits, the extent of confounding is likely to be modest and the confounding can generally be controlled by commonly-used methods. We discuss several aspects of detecting/controlling for stratification in a candidate-gene sized association study.

437 The superiority of Bayes and empirical Bayes estimators of a seemingly unrelated regression model

[PS 3,(page 29)]

Lichun WANG, *Department of Mathematics Beijing Jiaotong University Beijing 100044 P.R.China*

The paper considers the estimation of the vector β_1 in a system of two seemingly unrelated regressions

$$Y_1 = X_1\beta_1 + \epsilon_1, \quad Y_2 = X_2\beta_2 + \epsilon_2,$$

in which ϵ_1 and ϵ_2 are correlated random vectors. The aim of this paper is to construct a reasonable estimator of β_1, which makes use of all information of regressions. To adopt the Bayes and empirical Bayes approach, we assume that the β_1 is a random vector with the normal prior distribution $N(\beta_{01}, \sigma^2_{\beta_1}\Sigma_{\beta_1})$. Under any quadratic loss function, the Bayes estimator $\hat{\beta}_1^{(BE)}$ of β_1, given Y_1, clearly depends on $Y = (Y_1', Y_2')'$ only through Y_1. To improve $\hat{\beta}_1^{(BE)}$ we use the covariance adjustment technique and obtain a sequence $\hat{\beta}_1^{(BE)}(k)$ of decision rules, having the property $Cov(\hat{\beta}_1^{(BE)}(k+1)) \leq Cov(\hat{\beta}_1^{(BE)}(k))$ for all $k \geq 1$. This sequence tends to the estimator $\hat{\beta}_1^{(BE)}(\infty)$, which not only satisfies $Cov(\hat{\beta}_1^{(BE)}(\infty)) \leq Cov(\hat{\beta}_1^{(BE)}(k))$, $k \geq 1$, but also contains all information of β_1 in the regressions. Next the paper proves that if the loss of estimation is measured by the mean square error matrix (MSEM), then the decision $\hat{\beta}_1^{(BE)}(\infty)$ is better than the best linear unbiased estimator $\hat{\beta}_1^{(BLUE)}$, i.e. MSEM $(\hat{\beta}_1^{(BE)}(\infty)) <$ MSEM $(\hat{\beta}_1^{(BLUE)})$. This result is derived in the case where the covariance matrix of errors is known. In the second part of the paper, we assume that this matrix is unknown and, in an analogous manner, prove the MSEM superiority of the empirical Bayes estimator of β_1.

438 Male dominance rarely yields long term evolutionary benefits

[CS 21,(page 21)]

Joseph C WATKINS, *Department of Mathematics University of Arizona*

Many studies have argued that reproductive skew biased towards dominant or high-ranking men is very common in human communities. While variation in male fitness is known to occur, an important unanswered question is whether such differences are heritable and persist long enough to have evolutionary consequences at the population level. Highly variable polymorphisms on the non-recombining portion of the Y chromosome can be used to trace lines of descent from a common male ancestor. Thus it is possible to test for the persistence of differential fertility among patrilines. We examine haplotype distributions defined by 12 short tandem repeats in a sample of 1269 men from 41 Indonesian communities. After establishing that a cut-off phenomena holds for an infinite alleles Moran model associated with this genetic history, we test for departures from neutral mutation-drift equilibrium based on the Ewens sampling formula. Our tests reject the neutral model in only five communities. Analysis and simulations show that we have sufficient power to detect such departures under varying demographic conditions including founder effects, bottlenecks and migration, and at varying levels of social dominance. We conclude that patrilines are seldom dominant for more than a few generations, and thus traits or behaviors that are strictly paternally inherited are unlikely to be under strong cultural selection.

439 Asymptotic distributions of U-statistics based on trimmed and winsorized samples

[CS 80,(page 58)]

Neville WEBER, *University of Sydney*
Yuri BOROVSKIKH, *Transport University, St Petersburg.*

Let $X_1, X_2, \ldots$ be a sequence of real valued, independent random variables with a common distribution function $F(x)$ and let $X_{n1} \leq \cdots \leq X_{nn}$ denote the order statistics, based on the sample $X_1, \cdots, X_n$. For any integer $n \geq 1$ and $0 < \gamma < 1$ let $n_\gamma = [\gamma n]$, where $[\cdot]$ denotes the integer part. For $0 < \alpha < \beta < 1$ the trimmed sample is $X_{n,n_\alpha+1}, \cdots, X_{nn_\beta}$. The Winsorized sample is $W_{n1}, \ldots, W_{nn}$, where $W_{n1} = \cdots = W_{nn_\alpha} = X_{nn_\alpha}$, $W_{nk} = X_{nk}$ for $n_\alpha + 1 \leq k \leq n_\beta$, and $W_{n,n_\beta+1} = \cdots = W_{nn} = X_{n,n_\beta+1}$.

Let h be a real valued, symmetric function. The U-statistic with kernel h is defined as $U = \binom{n}{m}^{-1} \sum_{\sigma_{nm}} h(X_{i_1}, \cdots, X_{i_m})$, where $\sigma_{nm} = \{(i_1, \cdots, i_m) : 1 \leq i_1 < \cdots < i_m \leq n\}$. The U-statistic based on the Winsorized sample is denoted by U_W, and, given $0 < \alpha < \beta < 1$, the U-statistic with kernel h based on the trimmed sample is U_T.

When $m = 1$ and $h(x) = x$ then the above statistics reduce to the trimmed and Winsorized sample mean. The asymptotic behaviour of these statistics has a long history. For the trimmed mean Stigler (Ann. Statist.(1973), 1, 472-477) demonstrated that the limiting behaviour is a function of a three dimensional normal random vector with a covariance matrix depending on the quantile function $F^{-1}(x)$. The limiting distribution of the trimmed mean is normal if and only if α and β are continuity points of $F^{-1}(x)$. The corresponding limit result for non-degenerate U-statistics is established.

Asymptotic results to date for trimmed and Winsorized U-statistics have focussed on conditions that ensure U_T and U_W have a limiting normal distribution. However the class of asymptotic distributions associated with U-statistics is much broader and includes, for example, distributions represented in terms of multiple Ito-Wiener stochastic integrals. The corresponding non-Gaussian limit behaviour for U_T and U_W will be investigated.

440 Quantifying prediction uncertainty in computer experiments with fast Bayesian inference

[IS 6,(page 30)]

William J. WELCH, *University of British Columbia*
Bela NAGY, *University of British Columbia*
Jason L. LOEPPKY, *University of British Columbia*

Complex computer codes often require a computationally less expensive surrogate to predict the response at new, untried inputs. Treating the computer-code function as a realization of a random function or Gaussian process is now a standard approach for building such a surrogate from limited code runs. Scientific objectives such as visualization of an input-output relationship and sensitivity analysis can be conducted relatively quickly via the surrogate. The random function approach respects the deterministic nature of many computer codes, yet it also provides statistical confidence or credibility intervals around the predictions to quantify the prediction error. Whether these intervals have the right coverage is a long-standing problem, however. In this talk we introduce a new, simple, and computationally efficient Bayesian method for constructing prediction intervals that have good frequentist properties in terms of matching nominal coverage probabilities. This is demonstrated by simulation and illustrated with climate-model codes.

441 A Bayesian spatial multimarker genetic random-effect model for fine-scale mapping

[PS 1,(page 4)]

Shu-Hui WEN, *Department of Public Health, College of Medicine, Tzu-Chi University*
Miao-Yu TSAI, *Institute of Statistics and Information Science, College of Science, National Changhua University of Education*
Chuhsing Kate HSIAO, *Department of Public Health and Institute of Epidemiology, College of Public Health, National Taiwan University*

Multiple markers in linkage disequilibrium (LD) are usually used to localize the disease gene location. These markers may contribute to the disease etiology simultaneously. In contrast to the single-locus tests, we propose a genetic random effects model that accounts for the dependence between loci via their spatial structures. In this model, the locus-specific random effects measure not only the genetic disease risk, but also the correlations between markers. We consider two different settings for the spatial relations, the relative distance function (RDF) and the exponential decay function (EDF). The inference of the genetic parameters is fully Bayesian with MCMC sampling. We demonstrate the validity and the utility of the proposed approach with two real datasets and simulation studies. The analyses show that the proposed model with either one of two spatial correlations performs better as compared with the single locus analysis. In addition, under the RDF model, a more precise estimate for the disease locus can be obtained even when the candidate markers are fairly dense. In each of the simulations, the inference under the true model provides unbiased estimates of the genetic parameters, and the model with the spatial correlation structure does lead to greater confidence interval coverage probabilities.

442 Stein's identity for Bayesian inference

[CS 42,(page 37)]

Ruby Chiu-Hsing WENG, *National Chengchi University*

This talk describes applications of a version of Stein's Identity in Bayesian inference. First, we show that the iterative use of Stein's Identity leads to an infinite expansion of the marginal posterior densities. Next, we derive second order approximations and assess its accuracy by the logit model with moderate sample size. Then, we study the performance of analytic approximations when the sample size is small or the likelihood is not unimodal. In such cases, second order approximations do not work well, but analytic approximations with higher moments are still promising. Some examples are provided for illustration.

443 Are frontiers always symmetric?

[BS-IMS Lecture 1,(page 15)]

Wendelin WERNER, *universite Paris-Sud 11 and Ecole Normale Superieure*
Pierre NOLIN, *Ecole Normale Superieure*

We will consider the following question: Are random interfaces necessarily (statistically) symmetric on large scale, or is it possible to detect asymmetric features?

We will give a few results, some heuristics and some conjectures.

444 Sparse latent factor modelling

[IS 20,(page 50)]

Mike WEST, *Duke University*

A number of recent studies in the application of sparse latent factor models have highlighted the utility of such modelling approaches in high-dimensional problems in areas such as genomics and finance.

I will discuss aspects of the approach, including development of model form and specification, sparsity prior structuring and summaries of posterior distributions in contexts of both experimental and observational studies in cancer genomics.

Issues of computation and model search are paramount. I will highlight the utility and importance of adopting computational strategies that are customised to specific predictive and discovery goals and that raise interesting new research questions in computational statistics in such high-dimensional model contexts.

445 Estimation of the reproductive number and serial interval with an application to the 1918 influenza pandemic

[IS 21,(page 31)]

Laura Forsberg WHITE, *Boston University School of Public Health*
Marcello PAGANO, *Harvard School of Public Health*

Estimation of basic epidemiological parameters useful in quantifying infectious disease epidemics is an important challenge. Understanding of these parameters leads to a more informed public health response in a current epidemic and the ability to prepare for future epidemics. In this talk methods for estimating the reproductive number, defined as the number of secondary cases produced by an infected individual, and the generation interval, or probability distribution describing the time between symptomatic cases, will be presented. Among these is a likelihood-based method that requires only information on the number of new cases of disease each day. These methods are applied to data from documented influenza outbreaks during the 1918 Influenza Pandemic. From this analysis, it is clear that the dynamics of influenza spread vary according to population structure.

446 Inference for data graphics

[IS 1,(page 5)]

Hadley WICKHAM, *Iowa State University*

Humans love to tell stories and see patterns in pictures. How can we rein in these natural tendencies to ensure that we doní make false inferences (and bad decisions) after looking look at pictures of our data? In this talk, I will discuss some methods for graphical inference, based around the idea of randomisation (and permutation tests) for visualisation. Randomisation based methods are particularly powerful because they can be used for a wide range of graphics, and we can use our brains to generate the test statistics of visual difference. This builds on work by Andreas Buja, Dianne Cook, and Heike Hofmann.

447 Identifying differently expressed correlated genes in DNA microarray experiments

[CS 21,(page 21)]

Manel WIJESINHA, *Pennsylvania State University USA*
Dhammika AMARATUNGA, *Johnson and Johnson USA*

DNA microarray experiments commonly involve comparison of two groups such as the gene expression differences between a treated and a control group of patients. This comparison can be done individually gene by gene using simple t-tests. Due to very little replication, however, such individual t-tests may not yield high power. Also, these t-tests are carried out individually, under the unrealistic assumption that the genes are not correlated. This paper develops a procedure for the comparison of gene expression profiles using the Seemingly Unrelated Regression Equations (SURE) model which takes into account possible correlations among genes. Also, this collective approach of combining data across all genes will yield more efficient estimates for the unequal variances occurring within separate genes. Results of a power comparison between this method with other existing approaches will also be presented.

448 The distortion of Copulas: an application to drought

[CS 14,(page 14)]

Geraldine WONG, *School of Mathematical Sciences, The University of Adelaide*
Andrew V. METCALFE, *School of Mathematical Sciences, The University of Adelaide*
Martin F. LAMBERT, *School of Civil and Environmental Engineering, The University of Adelaide*

Drought is a global phenomenon and is a common characteristic of climate. It is widely considered the worldś costliest natural disaster in annual average terms [1], and the effects are especially devastating to the agricultural and social economy. Drought forecasting would enable decision makers to mitigate these effects by management of water systems and agriculture. The essential characteristics of drought are peak intensity, average intensity and duration. These variables are highly correlated among themselves and are also found to be dependent on climatic indices such as the Southern Oscillation Index (SOI). The correlation structure of these drought characteristics can be described by copulas. Copulas are multivariate uniform distributions, which allows for separate marginal and joint behaviour of variables to be modelled. Complex hydrological systems, such as droughts, are often heavy tail-dependent. It has been shown that trivariate asymmetric Gumbel copulas can model the multivariate dependence structure of drought characteristics in rainfall districts in Australia [2]. However, these Gumbel copulas have a restriction that the outer correlations are equal. The multivariate t-copula is another family of copulas with the ability to model symmetric upper and lower tail dependence. However, in the hydrological context, the dependence structure may not be well modelled by a t-copula, which is symmetric about the mean. The construction of an alternative t-copula is introduced through a transformation called a distortion. Appropriate distortion functions will be applied to the t-copula and the tail dependence resulting from the distorted t-copula will be examined through simulations from the copula, and

compared with the asymmetric Gumbel copula and the t-copula. Statistical tests demonstrate there is statistically significant difference between SOI states. Separate asymmetric copulas modulated by the SOI will be fitted to historical data and any improvements in characterizing potential droughts discussed. The use of copulas in forecasting and simulation will be discussed.

References

1. J. Keyantash and J. A. Dracup, *The Quantification of Drought: An Evaluation of Drought Indices*, Bulletin of American Meteorological Society. 83, 1167-1180 (2002).
2. G. Wong, M.F. Lambert and A.V. Metcalfe, *Trivariate copulas for characterization of droughts*, ANZIAM J. 49, C306 (2007).

449 Score test statistics for testing the equality of the survival functions of current status data

[CS 28,(page 26)]

Kam-Fai WONG, *Institute of Statistics No.700,Kaohsiung University Rd.,Nan Tzu Dist., 811.Kaohsiung,Taiwan*

Current status data arise when the failure time of interest is unable to observe. The observation consists only of a monitoring time and knowledge of whether the failure time occured before or after the monitoring time. For testing the equality of the survival functions of the failure time among different groups, a test statistic provided by Sun and Kalbfleisch (1993) can be applied. Coincidentally, the score test statistics under proportional hazards model, additive hazards model and accelerated life time model, respectively, are the weighted version of the test statistic given by them. In fact, under the regular condition given by them, the weighted version of the statistic still maintain the asymptotic properties that they showed. A series of simulation studies are presented to evaluate the finite sample performance of these four test statistics under different models.

450 Non-negative least squares random field theory

[IS 5,(page 18)]

Keith J. WORSLEY, *Department of Statistics, University of Chicago*
Jonathan TAYLOR, *Stanford University*

We fit a linear model by non-negative least squares at each point in a Gaussian random field. We wish to detect the sparse locations where the coefficients of the linear model are strictly greater than zero. In general, we wish to detect a sparse cone alternative. To do this we calculate the usual Beta-bar statistic for testing for a cone alternative at each point in the Gaussian random field. Such a Beta-bar random field has been proposed in the neuroscience literature for the analysis of fMRI data allowing for unknown delay in the hemodynamic response. However the null distribution of the maximum of this 3D random field of test statistics, and hence the threshold used to detect brain activation, was unsolved. Our solution approximates the P-value by the expected Euler characteristic (EC) of the excursion set of the Beta-bar random field. Our main result is the required EC density, derived using Taylor's Gaussian kinematic formula. We apply this to a set of fMRI data on pain perception.

451 Explicit and implicit algebraic statistical models

[IS 2,(page 23)]

Henry P. WYNN, *London School of Economics*

Algebraic Statistical models are those can be expressed using algebraic structures. The two principal areas in which there has been considerable recent activity are (i) polynomial regression models, which can be derived via quotient operation when the experimental design is expressed as a zero dimensional algebraic variety and (ii) polynomial models for categorical data which arise naturally from independence, conditional independence and other models. An *explicit* model is one for which the response mean, probability distribution, or similar feature, is express explicitly in terms of unknown parameters. An *implicit* model is one in which the parameters are eliminated to induce an implicit relationship on the response quantity. The equation $p_{00}p_{11} - p_{10}p_{01} = 0$ from a 2×2 contingency table is a simple example. One of the most important uses of algebraic statistics is to track the relationship between the implicit and explicit forms.

A general class of examples is discussed which puts a special condition on the lattice consisting of all cumulative probability functions $F(x_1, \ldots, x_n)$ and all marginals, $F(x_1), F(x_1, x_2)$ etc. This generalises the idea of a junction tree, in the theory of graphical models. It is seen that the natural setting for discussing this object is a simplicial complex, rather than a graph. The condition makes the log-density (or log-likelihood in the parametric case) have special additive properties on the lattice. The condition is the same as the tube condition studied by the authors and a number of co-authors, in particular D Q Naiman. However, whereas in that work the condition was on the range of the random variables, here the condition is on the index set. But recognising the connection allows several new examples, with an algebraic and geometric quality, to be given.

In so far as the special conditions are implicit, one can use the duality between implicit and explicit representations to give an explicit functional form for distributions possessing the condition. The exponential version gives a general type of exponential family. The structures have yet another representation in terms of commutativity of certain conditional expectations. The Gaussian and binary cases are given special attention.

452 The context-dependent DNA substitution process

[CS 21,(page 21)]

Von Bing YAP, *National University of Singapore*

A DNA molecule can be represented as a sequence of bases, of which there are four types: {T, C, A, G} = S. The substitution process refers to the replacement of some bases in the DNA by some other bases. A complicated mixture of mutations, selection and fixation, the substitution process is often modelled as a continuous-time Markov chain. The simplest model assumes that the bases evolve independently according to a chain on S specified by a substitution rate matrix, containing up to 12 free parameters. This class of models is the workhorse for many applications, such as estimating relative rates of evolution and estimating a phylogenetic tree. However, it is known that bases do not evolve independently, the most famous example being the significantly higher rate of a C to T substitution in the motif CG. To account for such context dependence, a number of investigators proposed that the rate of a base being substituted by another also depends on the two neighbouring bases. Despite its simplicity, the generalisation makes exact computation of the equilibrium distribution and transition matrices difficult. Approximations such as MCMC and pseudo-likelihood have been employed in statistical inference. There are two ways of viewing the process. First, a DNA sequence of length n evolves according to a Markov chain on S^n according to a sparse rate matrix. Second, it is a spatial process (Kelly, *Reversibility and Stochastic Networks*, 1979), where the DNA sequence is represented as a graph. Even though the state space is huge for most practical values of n, the alternative views may yield better or computationally more efficient approximations, given that the number of rate parameters is constant: $12 \times 16 = 192$.

453 Covariance estimation and hypothesis testing for generalized Wishart models

[CS 34,(page 32)]

Ahmad YASAMIN, *Indiana University*

Abstract. In this paper we consider problems of covariance estimating and hypotheses testing for statistical models parameterized by a symmetric cone and invariant under a group action. In particular, we extend the definition of some non-central distributions, derive the joint density of the eigenvalues of a generalized Wishart distribution, and propose a test statistic for testing homoscedasticity across a sample population taken from a generalized Wishart model. This test statistic is analog to the Bartlett's test, which tests the equality of variances across a normally distributed population, and furthermore, generalizes the Bartlett's test to all types of generalized Wishart distributions, namely, real, complex, quaternion, Lorentz and exceptional types. In the final chapter of the dissertation. Our main approach to these problems is to disintegrate the probability distribution of the parameterized model to the product of the transformed measure, under a maximal invariant statistic, and a quotient measure. We prove that densities of these two measures, with respect to the restrictions of Lebesgue measures , are functions of the eigenvalues of the generalized Wishart distribution. Our methodology in this approach is grounded on the analysis of symmetric cones and the structure of Euclidean Jordan algebras, which intrinsically differs from the classical method of using differential forms in deriving the marginal measures.

454 A non parametric control chart to detect shifts in mean and variance

[CS 52,(page 43)]

Nihal YATAWARA, *Curtin University of Technology, Australia*

Pairoj KHAWSITHIWONG, *Silpakorn University, Thailand*

Industrial data often does not follow a normal distribution. Hence, control charts based on Gaussian assumptions could give misleading signals. Khawsithiwong and Yatawara(2007),(2006) developed a general class of control charts for correlated data arising from elliptically contoured distributions. Instead of using the usual 3-sigma limits, the out of control signals in these charts were based on the well known Kullback-Leibler minimum discrimination information(MDI) function. In this paper we propose a non-parametric control chart as an alternative. We treat the data to have been independently generated in a control state and in a monitoring state of a process. The densities of data pertaining to these two states are estimated by using non parametric density estimators and a control chart based on minimum Hellinger distance is developed for simultaneous monitoring of the mean and the variance. The performance of the chart is highlighted through simulation studies.

455 Variance bias and selection

[IS 12,(page 18)]

Daniel YEKUTIELI, *Tel Aviv University*

Estimation error is usually considered as the result of the tradeoff between variance and bias. In recent work with Yoav Benjamini, we have suggested viewing False Discovery Rate adjusted inference as inference for selected parameters, and we have shown that selection causes more false discoveries, and more generally, higher estimation error. In my talk, I will discuss the relation between multiple testing and estimation error. I will show how FDR adjusted confidence intervals can be used to gauge the increase in estimation error due to selection, and illustrate how hierarchical FDR testing procedures decrease selection and produce smaller estimation error. I will also discuss a general Bayesian framework for providing inference for selected parameters

456 On solutions of a class of infinite horizon FBSDEs

[CS 77,(page 57)]

Juliang YIN, *Department of Statistics, Jinan University, P.R.China*

Keywords: FBSDEs; Adapted solution; Infinite horizon; Contraction mapping theorem

AMS(2001)Subject Classification: 60H10; 60H20

In this talk, I focous on the solvability of a class of infinite horizon forward-backward stochastic differential equations (FBSDEs, for short), where the coefficients in such FBSDEs do not satisfy the traditional Lipschitz condition and the weak monotonicity condition. Under some mild assumptions on the coefficients, the existence and uniqueness result of adapted solutions is stablished. The method adopted here is based on constructing a contraction mapping related to the solution of forward SDE in the FBSDEs.

457 Multivariate GARCH models with correlation clustering

[CS 39,(page 34)]

Iris W. H. YIP, *HKUST*
Mike K. P. SO, *HKUST*

This paper proposes a clustered correlation multivariate GARCH model (CC-MGARCH) which allows the conditional correlations forming clusters and each cluster follows the same dynamic structure. One main feature of our model is to form a natural grouping of the correlations among the series while generalizing the time-varying correlation structure proposed by Tse and Tsui (2002). To estimate our proposed model, we adopt Markov Chain Monte Carlo methods. Forecasts of volatility and value at risk can be generated from the predictive distributions. The proposed methodology is illustrated using both simulated and actual international market with high dimensional data.

458 SVM sensitivity analysis and its robust variants

[CS 70,(page 53)]

Chia Hsiang YU, *Department of Statistics, National Central University*

In this talk we will discuss SVM robustness by its sensitivity analysis. A class of new SVM variants is proposed based on a minimum psi-principle. The proposed method will place less weight to outlying points or mislabeled instances, and thus, behaves more resistant to data contamination and model deviation. Adaptive tuning procedures will be discussed and numerical examples be presented as well. To choose a good parameter setting for a better performance, we adopt a nested 3D uniform design (UD) search scheme. The nested 3D-UD search is to select the candidate set of parameters and employs a k-fold cross-validation to find the best parameters combination. The nested UD is used to replace less efficient grid search. Our numerical results indicate that the minimum psi-based SVM performs more robust than various conventional SVM algorithms against data contamination, model deviation and class membership mislabeling.

Key words and phrases: Support vector machine, uniform design, robust statistics.
AMS subject classifications: 62G35.

459 Bandwidth selection for kernel density estimators for randomly right-censored data

[CS 53,(page 43)]

Hui-Chun YU, *Department of Statistics, National Cheng-Kung University, Tainan, Taiwan 70101*
Tiee-Jian WU, *Department of Statistics, National Cheng-Kung University, Tainan, Taiwan 70101*

Based on randomly right-censored sample of size n, the problem of selecting the global bandwidth in kernel estimation of lifetime density f is investigated. We proposed a stabilized bandwidth selector, which is an extension to censored data of the complete-sample selector of Chiu (Biometrika 79:771-782, 1992). The key idea of our selector is to modify the weighted sample characteristic function beyond some cut-off frequency in estimating the integrated squared bias. It is shown that under some smoothness conditions on f and the kernel, our selector is asymptotically normal with the optimal root n relative convergence rate and attains the (conjectured) information bound. In simulation studies the excellent performances of the proposed selector at practical sample sizes are clearly demonstrated. In particular, the proposed selector performs conclusively better than the one selected by cross-validation.

460 Kernel quantile based estimation of expected shortfall

[CS 68,(page 51)]

Keming YU, *CARISMA, Brunel University,UK*
Shanchao YANG, *Guangxi Normal University, China*
Guglielmo Maria CAPORALE, *Centre for Empirical Finance, Brunel University, UK*

Expected shortfall (ES), a newly proposed risk measure, is currently attracting a lot of attention in market practice and financial risk measurement due to its coherence property. However, there is very little research on the estimation of ES as opposed to value at risk (VaR), particularly in the area of nonparametric estimation which provides a distribution-free estimation. In this paper we consider a few kernel-based ES estimators, including jackknife based bias-correction estimators with theoretically improving bias from $O(h^2)$, as the smoothing parameter $h \to 0$, to $O(h^4)$. Bias-reduction is particularly effective to reduce the tail estimation bias as well as the consequential bias that arises in kernel smoothing and finite sample fitting while risk estimators usually correspond to the extreme quantiles of underlying asset prices distributions.

In particular, by taking advantage of ES as an integral of quantile function we propose new kernel-based plug-in ES estimators. The comparative simulation study suggest that a new plug-in bias-reduction kernel estimator of ES which has an analytic expression should be used in practical applications.

461 On semiparametric inference in partially linear additive regression models

[CS 36,(page 33)]

Kyusang YU, *Universty of Mannheim*
Enno MAMMEN, *Universty of Mannheim*
Byeong U PARK, *Seoul National University*

Many studies have been done for the partial linear regression models. Most studies, however, are rather focused on the cases with the single dimensional (or at least low dimensional) function parameter in the regression model since the dimensionality costs higher order smoothness in theory and poor small sample performance in practice. In this paper we consider partial linear models with additive nonparametric regression parameters given by

$$Y = m_0 + X^T b + m_1(Z_1) + ... + m_d(Z_d)$$

for the response Y and the $(p+d)-$dimensional covarates $(X, Z_1, ..., Z_d)$. This model circumvents the curse of dimensionality as in the nonparametric additive regressions. We show that under this partially linear additive models we have smaller (in a certain sense) information bound than under the usual partial linear models. We also present asymptotically efficient estimators for the finite dimensional parameter b based on nonparametric additive estimators under certain techniacl conditions.

462 Determining the contributors to SPC charts signal in a multivariate process

[CS 52,(page 43)]

Yuehjen E. SHAO, *Graduate Institute of Applied Statistics, Fu Jen Catholic University*
Bo-Sheng HSU, *Graduate Institute of Applied Statistics, Fu Jen Catholic University*

Quality is one of the most important features to describe the products. Because the statistical process control (SPC) chart has practical monitoring capability, it becomes an important quality control technique. The SPC charts are able to effectively and correctly detect the process disturbances when they were introduced in the process. Nevertheless, the SPC charts still have some limitations, especially in monitoring a multivariate process. A multivariate process would have two or more variables (or quality characteristics) to be monitored. When an out-of-control signal is triggered by the SPC multivariate control chart, the process personnel usually only know that the process is wrong. However, it is very difficult to determine which of the monitored quality characteristics is responsible for this out-of-control signal. In this study, we propose the machine learning mechanisms to solve this problem. We integrate the neural network (NN) and support vector machine (SVM) with the multivariate SPC charts to determine the contributors to an SPC signal. The fruitful results are demonstrated through the use of a series of simulations.

463 Parametric estimation by generalized moment method

[CS 5,(page 8)]

ZDENĚK FABIÁN, *Institute of Computer Science, Academy of Sciences of the Czech republic, Pod vodárenskou věží 2, 18200 Prague, Czech republic*

Let $\theta \in \Theta \subseteq R$ and $X_1, ..., X_n$ be i.i.d. according to F_θ with density f_θ. The standard problem of the estimation of θ has standard solutions. However, the maximum likelihood estimation equations are often cumbersome due to the normalizing factor, which can be a non-elementary function of θ. The more simple moment equations

$$\frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = ES^k(\theta) \qquad k = 1, ..., m,$$

have serious drawbacks if $S(x;\theta) \equiv x$: EX^k may not exist or the estimates may be very inefficient.

We find a scalar inference function S which 'fits' family $\{F_\theta, \theta \in \Theta\}$ so that moments $ES^k(\theta)$ exist and moment equations are sufficiently simple. Such a function is known for a long time for distributions G supported by R; it is the score function $Q_\theta(x) = -g'_\theta(x)/g_\theta(x)$.

We suggested to view any F_θ supported by $\mathcal{X} \neq R$ as a transformed 'prototype' $G : F_\theta(x) = G(\eta(\theta))$ and to take as the inference function the transformed score function of its prototype, $S_\theta(x) = Q(\eta(\theta))$. By a consistent use of one concrete, support-dependent mapping $\eta : R \to \mathcal{X}$ (the Johnson mapping), it is possible to compare 'center points' and variabilities of distributions within a large class of distributions [1]. To obtain as simple as possible equations, it is to find the 'most convenient' η, which is often, but not necessarily the Johnson one. We show that the generalized moment estimates appear to be serious competitors of the maximum likelihood ones: they are robust in cases of heavy-tailed distributions.

The work was supported by MSMT under grant ME 949.

References: [1] Fabián, Z.: New measures of central tendency and variability of continuous distributions, *Communication in Statistics, Theory Methods* **37** (2008), pp.159-174.

464 How geometry influences statistics of zeros and critical points

[IS 5,(page 18)]

Steve ZELDITCH, *Johns Hopkins University*

The talk concerns the zeros and critical points of Gaussian random functions on Riemannian and Kahler manifolds. The metric is used to define an inner product and Gaussian measure. We describe how the statistics of zeros and critical points reflects the choice of metric. We mainly discuss asymptotics as either the 'degree' of the functions tends to infinity or as the dimension of the manifold tends to infinity. We briefly discuss applications to the statistics of vacua in string theory.

465 Continuous time principal-agent problems

[IS 27,(page 10)]

Jianfeng ZHANG, *University of Southern California*
Jaksa CVITANIC, *California Institute of Technology*
Xuhu WAN, *Hong Kong University of Science and Technology*

A pricipal hires an agent to run a bussiness and pays him compensation based on a contract. The two parties have their own utility, and the principal-agent problem is to find the optimal contract which maximizes the principal's utility, by expecting that the agent would choose optimal actions to maximize the agent's utility. Mathematically, this is a sequential stochastic control problem. The main difficulty lies in the asymetric information. The agent's action and/or type are unknown to the principal. We take the stochastic maximum principle approach. The optimal contract and the agent's optimal action are characterized as the solutions to a forward backward stochastic differential equations. We solve some examples explicitly.

466 Admissibility in the general multivariate linear model with respect to restricted parameter set

[PS 2,(page 16)]

Shangli ZHANG, *School of Science, Beijing Jiaotong University, Beijing 100044, P.R.China*
Gang LIU, *School of Information, Renmin University of China, Beijing 100872, P.R.China*
Wenhao GUI, *Department of Statistics, Florida State University, Tallahassee, FL 32306, USA*

In this paper, using the methods of linear algebra and matrix theory, we obtain the characterization of admissibility in the general multivariate linear model with respect to restricted parameter set. In the classes of homogeneous and general linear estimators, the necessary and sufficient conditions that the estimators of regression coefficient function are admissible are established.

467 Empirical likelihood inference for the Cox model with time-dependent coefficients via local partial likelihood

[CS 56,(page 45)]

Yichuan ZHAO, *Georgia State University*
Yangqing SUN, *University of North Carolina at Charlotte*
Rajeshwari SUNDARAM, *National Institute of Child Health and Human Development*

The Cox model with time-dependent coefficients has been studied by a number of authors recently. In this paper, we develop empirical likelihood (EL) pointwise confidence regions for the time-dependent regression coefficients via local partial likelihood smoothing. The EL simultaneous confidence bands for a linear combination of the coefficients are also derived based on the strong approximation methods. The empirical likelihood ratio is formulated through the local partial log-likelihood for the regression coefficient functions. The proposed EL pointwise/simultaneous confidence regions/bands have asymptotically desired confidence levels. Our numerical studies indicate that the EL pointwise/simultaneous confidence regions/bands have satisfactory finite sample performances. Compared with the confidence regions derived directly based on the asymptotic normal distribution of the local constant estimator, the EL confidence regions are overall tighter and can better capture the curvature of the underlying regression coefficient functions. Two data sets, the gastric cancer data and the Mayo Clinic primary biliary cirrhosis data, are analyzed using the proposed method.

468 Boundary Proximity of SLE

[IS 29,(page 24)]

Wang ZHOU, *National University of Singapore*
Oded SCHRAMM, *Microsoft Research*

This paper examines how close the chordal SLE$_\kappa$ curve gets to the real line asymptotically far away from its starting point. In particular, when $\kappa \in (0,4)$, it is shown that if $\beta > \beta_\kappa := 1/(8/-2)$, then the intersection of the SLE$_\kappa$ curve with the graph of the function $y = x/(logx)^\beta$, $x > e$, is a.s. bounded, while it is a.s. unbounded if $\beta = \beta_\kappa$:. The critical SLE$_4$ curve a.s. intersects the graph of $y = x^{-(loglogx)^\alpha}$, $x > e^e$, in an unbounded set if $\alpha \leq 1$, but not if $\alpha > 1$. Under a very mild regularity assumption on the function y(x), we give a necessary and sufficient integrability condition for the intersection of the SLE$_\kappa$ path with the graph of y to be unbounded. We also prove that the Hausdorff dimension of the intersection set of the SLE$_\kappa$ curve and real axis is $2 - 8/\kappa$ when $4 < \kappa < 8$.

469 Local linear quantile estimation for non-stationary time series

[CS 39,(page 35)]

Zhou ZHOU, *The University of Chicago*
Wei Biao WU, *The University of Chicago*

We consider estimation of quantile curves for a general class of non-stationary processes. Consistency and central limit results are obtained for local linear quantile estimates under a mild short-range dependence condition. Our results are applied to environmental data sets. In particular, our results can be used to address the problem of whether climate variability has changed, an important problem raised by IPCC (Intergovernmental Panel on Climate Change) in 2001.

470 Asymptotic properties of a marked branching process and interpretation to an empirical law in seismology

[CS 3,(page 7)]

Jiancang ZHUANG, *Institute of Statistical Mathematics*

David VERE-JONES, *Victoria University of Wellington, New Zealand*

The space-time epidemic-type aftershock sequence model is a stochastic branching process in which earthquake activity is classified into background (immigrants)and clustering components and each earthquake produces other earthquakes independently according to certain rules. This paper gives some limit properties of probability distributions associated with the largest event in a cluster and their properties for all three cases when the process is subcritical, critical, and supercritical. These probability distributions reveals that the Båth law, which declares that the largest aftershock has a magnitude about 1.2 lower than the mainshock magnitude, can be expressed as an asymptotic property of these distributions. That is, $M_1 = \sigma M_0 - \delta$, where M_1 is the medians of the magnitude of the largest aftershock, M_0 is the mainshock magnitude, and σ and δ are constants. Our results shows, under the assumption of the ETAS model, σ is determined by the productivity function and the magnitude distribution, and, under the different criticality cases of the process, σ takes different values.

471 On the distributions concerning expansions of real numbers

[CS 71,(page 53)]

T. M. ZUPAROV, *National University of Uzbekistan, Tashkent, Uzbekistan*

Sh. A. ISMATULLAEV, *Institute of Mathematics and Information Technologies, Tashkent, Uzbekistan*

For given natural **m** we shall denote **q** as a positive solution of the equation $x^{m+1} - x^m = 1$, and consider the expansion of $x \in [0;1]$:

$$x = \frac{e_1(x)}{q} + \frac{e_2(x)}{q^2} + ... + \frac{e_n(x)}{q^n} + ...;$$

where $e_k(x)$ takes values 0 and 1. It is known, that $e_k = e_k(x)$ can be interpreted as random variables on the probability space $(\Omega, \mathcal{F}, P)$, with $\Omega = [0;1]$, $\mathcal{F}$ - σ-algebra of all Borel subsets of [0; 1], and P - the Lebeque measure on [0; 1]. This talk is devoted to some results (in praticular, limit theorems) received by T. Abdullaev, J. Barsukova and the authors and concerning distributions of $e_1, e_2, ..., e_n$ and $s_n = e_1 + e_2 + .. + e_n$; among them:

Theorem 1. For $n \geq m+2$ the following formula is true

$$P(s_n = k) = q^{-n} C^k_{n-mk+m-1} + q^{-(n+m)} C^{k-1}_{n-mk+m-1}.$$

Theorem 2. The sequence $e_1, e_2, ..., e_n, ...$ is the ψ-mixing sequence, with exponential rate of mixing.