TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA UNIVERSITY OF WELLINGTON

Wellington Workshop
in
Probability Theory and Mathematical Statistics

3 and 4 November 2009

Victoria University of Wellington

Kelburn Campus, Murphy Building, MY 632

Presenters, titles and abstracts

(Ordered alphabetically, by presenters' last names)

Decomposition of mixtures of distributions

Christopher Ball, MSc student
Victoria University of Wellington

The aim of the talk is to explore the idea of decomposition of mixture distributions. Given a mixture of normal densities

$$f(x) = \sum_{i=1}^{m} \alpha_i \phi_{\mu_i,\sigma_i^2}(x)$$

and $\chi(t)$ is the corresponding characteristic function of f, and if $\sigma^2 < \min_i \sigma_i^2$, then $\chi(t)\exp(\frac{t\sigma^2}{2})$ is the characteristic function of the mixture

$$f^*(x) = \sum_{i=1}^{m} \alpha_i \phi_{\mu_i,v_i^2}(x)$$

of normal densities, with smaller variances $v_i^2 = \sigma_i^2 - \sigma^2$. In $f^*(x)$ the individual components of the mixtures are much more pronounced. The same idea can be carried over to other mixtures. The talk will discuss stability of numerical methods to implement this approach.

From estimation of traffic flows to deconvolution of densities: some statistical linear inverse problems

Martin Hazelton, Professor and Chair of Statistics
Massey University Manawatu

Statistical linear inverse problems come in a wide range of guises. Two that are of particular interest to me are traffic flow estimation, and density deconvolution. I briefly discuss the common structure of these problems, and then describe a new semiparametric approach to the second of them.

Analysis of a stochastic difference equation: exit times and invariant distributions

Goran Hognas and Brita Jung
Abo Akademi University, Department of Mathematics
Biskopsgatan 8, FIN-20500 Abo, Finland

Mark Kac proved in 1947 that the mean return time of a discrete Markov chain to a point x is the reciprocal of the invariant probability $\pi(x)$. This result was extended to chains on general measure spaces (subject to some irreducibility conditions) by Cogburn (1975). We revisit this classical theme to investigate certain exit times for stochastic difference equations of autoregressive type. More specifically, we will discuss the asymptotics, as $\epsilon \to 0$, of the first time $\tau\epsilon$ that the d-dimensional process

$$X^{\epsilon}_{n+1} = f(X^{\epsilon}_n) + \epsilon\xi_{n+1}, \quad n = 0, 1, 2, \ldots,$$

(where $\xi_1, \xi_2, \ldots,$ is a sequence of i.i.d. random d-vectors) leaves a given neighborhood of the fixed point of the contraction f.

1. R. Cogburn, A uniform theory for sums of Markov chain transition probabilities, *Annals of Probability* 3, 2 (1975), 191-214.
2. F. Klebaner, R. Liptser, Large deviations for past-dependent recursions, *Problems of Information Transmission* 32, 4 (1996), 320-330. Corrected version 2006.
3. B. Ruths (Jung), Exit times for past-dependent systems, *Surveys of Applied and Industrial Mathematics* (Obozrenie prikladnoy i promyshlennoy matematiki) 15, 1 (2008), 25-30.

Degenerate random environments

Mark Holmes
University of Auckland

Motivated by an interest in random walks in random environments, in joint work with Tom Salisbury we study a class of random directed graphs that are closely related to percolation models. Among other things we study the geometry of: the set of points reachable from the origin; the set of points from which the origin can be reached; the intersection of these two. We observe interesting phase transitions as we vary parameters of some of these models.

Coupling in Markov chains

Jeff Hunter, Professor Emeritus, Massey University Albany
and Adjunct Professor, AUT

The coupling of two discrete time Markov chains is discussed. Key properties are derived leading to the derivation of the expected time to coupling. In general, explicit expressions for these expectations are difficult to obtain. However, simple upper and lower bounds can be obtained.

Some special cases are presented together with a discussion of current research that hopefully will lead to some computational simplifications.

Uniform convergence of autocovariances

Laimonis Kavalieris
University of Otago

When modelling time series we accept that the dimension of candidate models increases with increasing sample size. In particular, for a given sample size n, candidate AR models should include all models up to a maximum order $p \leq H(n)$ where the bound $H(n)$ increases with sample size. In order to develop theory for model selection (using criteria such as AIC, BIC or MDL) estimators should converge uniformly for all $p \leq H(n)$. In order to achieve this we require good uniform convergence rates for sample autocovariances and autocorrelations.

Estimate the autocovariance function $\gamma(h)$ of a stationary zero mean time series as

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} y_t y_{t+h}$$

The first uniform convergence results were established by Ted Hannan and co-workers in a series of papers 1978–1984, where they showed

$$\max_{h<n} |\hat{\gamma}(h) - \gamma(h)| = \mathrm{O}(\log n/n)^{1/2}, \text{ a.s.}$$

where y_t has a MA(∞) representation

$$y_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}$$

where the ϵ's are stationary martingale differences with finite fourth moments, and $\sum j^{1/2}|\theta_j| < \infty$. [So, for example, long memory models are excluded.]

More recent work (Kavalieris, 2008) has required a moment of order α, $2 < \alpha < 4$ but then the ϵ's need to be independent. However the best rate of a.s. convergence is then $\mathrm{O}(n^{\alpha/2-1})$ for both autocovariances and autocorrelations although an in probability rate of $\mathrm{O}(\log n/n)^{1/2}$ is possible.

I want to discuss extensions to these and related results with the idea of dispensing with the annoying condition $\sum j^{1/2}|\theta_j| < \infty$ and replacing it with the ℓ_2 norm $(\sum_j \theta^2)^{1/2}$.

Goodness-of-fit problem for errors in nonparametric regression: distribution free approach

Estáte V. Khmaladze
Victoria University of Wellington

We discuss asymptotically distribution free tests for the classical goodness-of-fit hypothesis of an error distribution in nonparametric regression models. These tests are based on the same martingale transform of the residual empirical process as used in the one sample location model. This transformation eliminates extra randomization due to covariates but not due to the errors, which is intrinsic in the estimators of the regression function.

The results of this paper are applicable as soon as asymptotic uniform linearity of non-parametric residual empirical process at the rate $n^{-1/2}$ is available. In particular they are applicable under the conditions stipulated in recent papers [1] and [2]. The talk is based on joint work with Hira L. Koul [3].

1. Akritas, M.G. and van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.* 28, 549-567.
2. Müller, U.U., Schick, A. & Wefelmeyer, W. (2007). Estimating the error distribution function in semi-parametric regression. *Statistics and Decisions*, 25, 1-18.
3. E.V. Khmaladze, H.L. Koul, Goodness of fit problem for errors in non-parametric regression: distribution free approach. *Annals Statistics*, 2009, v. 37, 3165-3185.

Diversity analysis in multiple choice questionnaires

Giorgi Kvizhinadze, PhD Student
Victoria University of Wellington

The statistical analysis of a large number of rare events, or LNRE, which can also be called statistical theory of diversity, is the subject of acute interests both in statistical theory and in numerous applications. Careful eye will quickly see the presence of a large number of very rare objects almost everywhere: large number of rare species in ecosystems, large number of rare opinions in any opinion pool, large number of small admixtures in any solution and large number of rare words in any text are only few examples. In studying such objects, the difficulty lies in the fact that most of the frequencies are small and, therefore, quite unstable. It is not immediately clear how one should be able to derive consistent and reliable inference from a large number of such frequencies.

We consider a multiple choice questionnaire with q questions, and n individuals asked to fill out this questionnaire, so that we obtain n "opinions". Then how many different opinions will we observe? What will be the proportion of unique opinions? Or of opinions we will see twice or any k times?

One would assume that the answers depend on various properties of the questionnaire, such as the number of possible answers in each question, the probabilities of these answers and so on. However, we show that the asymptotic behaviour of the quantities we are interested in follows the Karlin-Rouault law, which depends only on one parameter that incorporates all information about the properties of the questionnaire.

1. Giorgi Kvizhinadze and Haizhen Wu, Diversity analysis in questionnaries – the general case, *Statistics and Probability Letters* (submitted), 2009.

Testing isotropy and a related random walk problem

Sreenivasa Rao Jammalamadaka
University of California, Santa Barbara, USA

One comes across directions as the observations in a number of situations. The first inferential question that one should answer when dealing with such data, is “Are they isotropic or uniformly distributed?” The answer to this question goes back in history which we shall retrace a bit and provide an exact and approximate solution to this so-called “Pearson’s Random Walk” problem.

An assumption-free two-sample test and confidence interval for the difference in medians

Robin Willink
Industrial Research Ltd

This talk presents a small-sample distribution-free two-sample test of medians that does not require the validity of the location-shift model involved in many other tests. The principle that the null and alternative hypotheses in a test are to be logically complementary is advocated (cf. [1], p. 287), and the familiar Wilcoxon-Mann-Whitney and two-sample-median procedures are discussed in the light of this principle. The location-shift model under which these standard tests are valid is criticised. A simple confidence interval for a difference in medians follows from the new test.

1. Edgington and Onghena, Randomization Tests, 2007, 4th ed.

The distribution of the maximum of two multivariate ARMA processes

Kit Withers
Industrial Research Ltd

We give the distribution function of the (element-wise) maximum of a sequence of n observations from a multivariate AR(1) process. We do the same for a multivariate MA(1) process. Solutions are first given in terms of repeated integrals and then for the case where the marginal distribution of the observations is absolutely continuous. The distribution of the multivariate maximum is then given as a weighted sum of the n-th powers of the eigenvalues of a non-symmetric Fredholm kernel. The weights are given in terms of the left and right eigenfunctions of the kernel.

Martingale limit theorem for divisible statistics in general situation

Haizhen Wu and Giorgi Kvizhinadze, PhD Students
Victoria University of Wellington

For each n, consider a multinomial random vector $(\nu_{n1}, \ldots, \nu_{nN})$ with sample size n and probabilities $(p_{n1}, \ldots, p_{nN})$. Statistics of the form

$$\sum_{i=1}^{N} g_{ni}(\nu_{ni}),$$

which include many popular statistics, such as the likelihood-ratio statistic, χ^2- statistics and many others, are often called *divisible statistics.*

Classical analysis studies the asymptotic behaviour of divisible statistics when the sample size increases but the probabilities are fixed. However, this scheme is not applicable to many interesting applications. From both practical and theoretical points of view, we should consider the case when, as $n \to \infty$ the number of classes N increases and the probabilities $\{p_{ni}\}$ also may change.

In [1] E. Khmaladze proposed an innovative martingale approach to analyze this sort of problems. He considered $n \sim N$ and $\sup Np_{ni} < \infty$ and show that the normalized partial sum process

$$X_{n,N}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \left[g_{ni}(\nu_{ni}) - \mathbb{E} g_{ni}(\nu_{ni}) \right].$$

converge to Ito Gaussian processes in distribution.

We extend this idea to more general situations where both of the two constraints will be removed. For example, when $n/N \to \infty$ or $n \sim N$ but some $Np_{ni} \to \infty$. Surprisingly, with a little modification of type of g_{ni}, the similar limit theorems can still be established.

1. E.V. Khmaladze, Martingale Limit Theorems for Divisible Statistics, *Theory of Probability and Applications,* 1983, 18, 530-549.
2. G. Kvizhinadze and Haizhen Wu, MLT approach to divisible statistics with uneven probabilities of cells, *Work in progress*, 2009.