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

Fifth Wellington Workshop
in
Probability and Mathematical Statistics

19 – 20 November 2015

Victoria University of Wellington

Presenters, Titles and Abstracts

(Ordered alphabetically, by presenters' last names)

Long-Term Forecasting of Volcanic Explosivity

Mark **Bebbington**
Massey University, Palmerston North

It has been thirty years since the terms time-predictable (repose length increases with previous eruption size) and size-predictable (eruption size increases with repose length) entered the volcanological lexicon. While much evidence of, and models for, the former have emerged, the latter is still largely unsubstantiated. Statistical tests for size-predictability from individual volcanoes suffer from insufficient power and the inherent non-normality and non-linearity in the relationship. Aggregating data from several volcanoes is difficult due to the different temporal and size scales involved. The only broadly available measure of eruption size is given by the Volcanic Explosivity Index (VEI), a 'logarithmic' scale on $\{0, 1, \ldots, 8\}$. We consider characterizing the VEI distribution by parameters which are themselves influenced by the length of the previous repose, the state of the conduit (open or closed) and possibly other factors. Dependency between the parameters for different eruptions at the same volcano is introduced using a multilevel (hierarchical) Bayesian formulation. Using data from Indonesia, largely since AD 1800, we find that there is a significant probability (> 0.999) that the VEI of the next eruption from closed conduit volcanoes increases with increasing repose length. For example, a further 10-year wait for the next eruption from Kelut increases the probability of a VEI > 2 by approximately 11%. On the other hand, open conduit volcanoes show no evidence of an increase in VEI with repose length. The results are insensitive to the details of the VEI distribution, prior distributions or number of levels in the Bayesian structure.

On Explicit Form of the Stationary Distributions
for a Class of Bounded Markov Chains

Kostya **Borovkov**
University of Melbourne
(Joint work with S McKinlay)

We consider a class of discrete time Markov chains with state space $[0, 1]$ and the following dynamics. At each time step, first the direction of the next transition is chosen at random with probability depending on the current location. Then the length of the jump is chosen independently as a random proportion of the distance to the respective end point of the unit interval, the distributions of the proportions being fixed for each of the two directions. Chains of that kind were subjects of a number of studies and are of interest for some applications. Under simple broad conditions, we establish the ergodicity of such Markov chains and then derive closed form expressions for the stationary densities of the chains when the proportions are beta distributed with the first parameter equal to 1. Examples demonstrating the range of stationary distributions for processes described by this model are given, and an application to a robot coverage algorithm is discussed.

Bootstrap Random Walks

Kais **Hamza**
Monash University, Clayton
(Joint work with Andrea Collevecchio and Meng Shi)

Consider a one dimensional simple random walk $X = (X_n)_{n\geq 0}$. We form a new simple symmetric random walk $Y = (Y_n)_{n\geq 0}$ by taking sums of products of the increments of X and study the two-dimensional walk $(X, Y) = ((X_n, Y_n))_{n\geq 0}$. We show that it is recurrent and when suitably normalised converges to a two-dimensional Brownian motion with independent components; this independence occurs despite the functional dependence between the pre-limit processes. The process of recycling increments in this way is repeated and a multi-dimensional analog of this limit theorem together with a transience result are obtained. The construction and results are extended to include the case where the increments take values in a finite set (not necessarily $\{-1, +1\}$). Time permitting, I will present a number of possible extensions currently being investigated.

Log-Density Estimation with Application to Approximate Likelihood Inference

Martin **Hazelton**
Massey University, Palmerston North

This talk concerns kernel estimation of the logarithm of a probability density function at a given evaluation point. The work is motivated primarily by approximate likelihood inference, in which we seek to construct an estimate of the log-likelihood function for a complex model by estimating the constituent log-densities from simulated realizations of the model. The properties of the kernel log-density estimator are heavily influenced by the unboundedness of the log function at zero. In particular, standard asymptotic expansions can provide a poor guide to finite sample behaviour for this estimator, with consequences for the choice of methodology for bandwidth selection. In response we develop a new approximate cross-validation bandwidth selector. We explore its theoretical properties and describe its finite sample behaviour through numerical experiments. This bandwidth selection methodology is then applied to estimation of log-likelihoods from model simulations. Our implementation of approximate likelihood inference is illustrated through an example in which a complex genetic model is used in estimating migration rates between village communities on the Indonesian island of Sumba.

Profile Likelihood Approach to a Large Sample Distribution of Estimators in a Joint Mixture Model of Survival and Longitudinal Ordered Data

Yuichi **Hirose**
Victoria University of Wellington
(Joint work with Ivy Liu and Kemmawadee Preedalikit)

The joint mixture model of survival and longitudinal ordered data is a mixture of semi-parametric models. We consider a general mixture of semi-parametric models and use the EM algorithm to derive asymptotic normality of the parameter of interest. Then the result is applied to the joint model under consideration.

Funky CLTs for Critical Interacting Particle Systems in High Dimensions

Mark **Holmes**
University of Auckland

We will discuss recent and ongoing work involving the proof of "functional central limit theorems for measure-valued processes", relevant to some well-known interacting particle systems (such as the voter model) in high dimensions.

The Accurate Computation of the Key Properties of Markov Chains and Markov Renewal Processes

Jeffrey J. **Hunter**
Auckland University of Technology

Based upon the Grassman, Taksar and Heyman (1985) algorithm and the equivalent Sheskin (1985) State Reduction algorithm for finding the stationary distribution of a finite irreducible Markov chain, Kohlas (1986) developed a procedure for finding the mean first passage times (MFPTs) in Markov renewal processes. The method is numerically stable as it doesn't involve subtraction. It works well for focusing on the MFPTs from any state to a fixed state but it is not ideally suited for a global expression for the MFPT matrix. We present some refinements to the Kohlas algorithm that we specialize to the case of Markov chains. A consequence of our procedure is that the stationary distribution does not need to be derived in advance but is found from the MFPT procedure. This also leads to an expression for the group inverse of $I - P$ where P is the transition matrix of the embedded Markov chain. We utilise MATLAB to compute expressions for the MFPTs. Some comparisons, using some test problems from the literature, with other techniques using generalized matrix inverses and perturbation techniques are also presented.

Keywords: Markov chain, Markov renewal process, stationary distribution, mean first passage times, generalized inverse, group inverse, perturbations.

References

Grassman, W. K., Taksar, M. I. and Heyman, D. P. (1985) Regenerative analysis and steady state distributions for Markov chains, *Oper. Res.* 33, 1107-1116.

Sheskin, T. J. (1985) A Markov partitioning algorithm for computing steady state probabilities, *Oper. Res.* 33, 228-235.

Kohlas, J. (1986) Numerical computation of mean first passage times and absorption probabilities in Markov and semi-Markov models, *Zeit fur Oper Res* 30, 197-207.

On the Function-Parametric Empirical Processes and Unitary Operators: Current Theory and Extensions

Estáte V. **Khmaladze**
Victoria University of Wellington

Suppose $v_F(\phi), \phi \in L_2(F)$, is a function-parametric Brownian bridge $\ldots$. However, having said that, we did not say enough – apart from the Brownian bridge that everybody knows, there are many other "bridges" occurring as limiting objects in different statistical problems. They all are Brownian motions with various orthogonality constraints. The process $v_{F,h}(\phi), \phi \in L_2(F)$, is h-projected F-Brownian bridge, if it has the distribution of a Brownian motion w_F on $L_2(F)$, which satisfies the orthogonality constraint $w_F(h) = 0$, where h is a given function from $L_2(F)$, or a vector of such functions.

Given two distributions F and G, let now $\mathbb{U}$ be a unitary operator from $L_2(F)$ to $L_2(G)$. For two processes $v_{F,h}(\phi)$ and $v_{G,q}(\phi)$, with relatively simple form of the operator $\mathbb{U}$ we can map one onto another:

$$v_{F,h}(\phi) = v_{G,q}(\mathbb{U}\phi),$$

with a relatively strange, but useful consequence: the two statistical problems, which look different, and which asymptotically lead to the processes $v_{F,h}(\phi)$ and $v_{G,q}(\phi)$, respectively, can be mapped into each other in one-to-one transformations. Therefore these different statistical problems are not different, but asymptotically equivalent. For example, the problems of testing any m-dimensional discrete distributions are equivalent to each other. Moreover, testing for parametric families of these distribution are also asymptotically equivalent.

This is what we know so far. However, extension to the class of statistical problems with covariates, we think, is also possible. Another possibility is an extension to Markov chains. This will allow statistical inference about different Markov chains to be made standard and free from a particular form of their transition matrices.

Bibliography

Khmaladze, E. V. (2013) Note on distribution free testing for discrete distributions, *The Annals of Statistics* 41(6), 2979-2993.

Khmaladze, E. V. (2015) Unitary transformations, empirical processes and distribution free testing, *Bernoulli Journal*, pp. 1-27, available online (under Forthcoming Papers): `http://www.bernoulli-society.org/index.php/publications/bernoulli-journal`

Associated Sequences and Demimartingales

B.L.S. **Prakasa Rao**
CR Rao Advanced Institute of Mathematics,
Statistics & Computer Science, Hyderabad

The concept of association of random variables was introduced by Esary, Proschan and Walkup (1967). In several situations, for example, in reliability and survival analysis, the random variables of lifetimes involved are possibly not independent but are generally associated. In the classical case of statistical inference, the observed random variables of interest are generally assumed to be independent and identically distributed. However, in several real life situations, the random variables need not be independent. In reliability studies, there are structures in which the components share the load, so that the failure of one component results in increased load on each of the remaining components. Failure of one component will adversely effect the performance of all the minimal path structures containing it. In such a model, the random variables are not independent but are possibly "associated". We give a short review of probabilistic properties of associated sequences of random variables and the related notion of demimartingales and their properties.

Bibliography

Esary, J. D., Proschan, F. and Walkup, D. W. (1967) Association of random variables, with applications, *The Annals of Mathematical Statistics* 38(5), 1466-1474.

Prakasa Rao, B. L. S. (2012) *Associated Sequences, Demimartingales and Nonparametric Inference*, xii+272pp. Series: Probability and its Applications, Springer Basel AG.

Distribution-Free Goodness of Fit Testing of Categorical Data

Leigh **Roberts**
Victoria University of Wellington

As a by-product of elucidating the relationship between Brownian bridges and Brownian motions, Khmaladze has recently illuminated a general theory of distribution-free goodness of fit testing for both simple and compound hypotheses. For discrete distributions, the elegance of his use of unitary operators to rotate vectors to distribution-free forms is striking. We illustrate this methodology by fitting the logistic distribution to simulated categorical data, testing for goodness of fit on the one hand, and comparing the power of the test against alternative forms of data generation on the other.

Ancestries of a Recombining Diploid Population

Raazesh **Sainudiin**
University of Canterbury
(Joint work with Bhalchandra Thatte and Amandine Véber)

I will convey by pictures the ideas in a recent theoretical unification of previously disparate population genetic processes that underpin modern statistical methods in population genomics: (1) Kingman's coalescent trees, (2) Chang's population pedigrees and (3) Griffiths' and Hudson's ancestral recombination graphs. Using a single parameter in the unit interval we build a family of combinatorial stochastic processes over biologically meaningful discrete random graphs and study their weak limits. The talk will aim to present the definitions needed to appreciate the five main propositions in this work. The basic ideas in the proofs will also be presented, if time permits.

Bibliography

Sainudiin, R., Thatte, B. and Véber, A. (2015) Ancestries of a recombining diploid population, *Journal of Mathematical Biology*, pp. 1-46, DOI: 10.1007/s00285-015-0886-z.

Current Trends in Random Matrix Analysis

Peter **Smith**
Victoria University of Wellington

The communications research literature relies heavily on random matrix theory for system analysis and design. Throughout the 1990s and 2000s this work was mainly focused on small Gaussian matrices which are the traditional models for wireless communication. Since 2010, two major changes have occurred. First, the potential for large systems to be practically deployed has been demonstrated. Mathematically, this means that simpler results for large dimensional matrices are now of genuine interest. Secondly, the deployment scenarios for large arrays tend to make non-Gaussian models of more interest. Hence, the recent developments have made analysis both easier (asymptotic results) and harder (non-Gaussian matrices). In this talk, I will explain these developments and give some specific examples of system analysis under these new regimes.

Some Identities Concerning Excursion Below Zero with Fixed Duration of Spectrally Negative Levy Process

Budhi **Surya**
Victoria University of Wellington
(Joint work with Ronnie Loeffen and Zbigniew Palmowski)

Motivated by some financial problems, we present some identities concerning excursion below zero with fixed duration of spectrally negative Levy process. Our main object of interest is the first time the excursion exceeds the specified duration. It is a stopping time and is usually referred to as Parisian ruin-time. We first derive the Laplace transform of this stopping time. Secondly, taking account of the result, we derive using Esscher transform of measure the joint Laplace transform of the stopping time and the corresponding position of the Levy process. Thirdly, we derive the $q-$resolvent kernel of the Levy process, which corresponds to the total discounted occupation time of the Levy process within an interval prior to the stopping time. The results have semi-explicit expressions in terms of the $q-$scale function. We show in the limit as the excursion duration goes to zero that the joint Laplace transform leads to the known Emery's fluctuation identity for first exit below zero of spectrally negative Levy process, whereas the $q-$resolvent kernel leads to the fluctuation identity obtained by Bertoin.

A Genetic Algorithm Approach to Network Inference

Matthieu **Vignes**
Massey University, Palmerston North

In this talk, I shall briefly present my motivating problem, the reconstruction of gene regulatory networks. I would then introduce genetic and evolutionary algorithms before combining the two themes in an original approach we proposed to explore the space of network topologies and perform inference.

Unbiased Estimates for Products of Moments and Cumulants for Samples from Finite Populations

Kit **Withers**
Wellington

Let F_N be the distribution of a finite real population of size N. Let F_n be the empirical distribution of a sample of size n drawn from the population without replacement. We prove the following remarkable inversion principle for obtaining unbiased estimates.

Let $T(F_N)$ be any product of the moments or cumulants of F_N.

Set $T_{nN}(F_N) = E\ T(F_n)$, then $E\ T_{Nn}(F_n) = T(F_N)$.

On the Ginibre Point Process
and its Applications

Aihua **Xia**
University of Melbourne

The Ginibre point process has attracted considerable attention recently because of its wide use in modelling mobile networks. The Ginibre point process is a special class of the Gibbs point process family and it exhibits a repulsion between the points, hence in applications it often fits better than the Poisson point process.

In this talk, I will focus on the analytical properties of the Ginibre point process and two applications: one is in mobile networks and the other is the Ginibre-Voronoi tessellation. The examples are taken from joint works with P. Keeler, N. Ross, L.H.Y. Chen and A. Roellin.