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

Fourth Wellington Workshop
in
Probability and Mathematical Statistics

2 – 3 December 2013

Victoria University of Wellington

Presenters, Titles and Abstracts

(Ordered alphabetically, by presenters' last names)

Couplings for Locally Branching Epidemic Processes

Andrew **Barbour**
Universitaet Zuerich

The asymptotic behaviour of many locally branching epidemic models can, at least to first order, be deduced from the limit theory of two branching processes. The first is the 'Whittle' branching approximation to the early stages of the epidemic, the phase in which approximately exponential growth takes place. The second is the 'susceptibility', or backward branching process, that approximates the contact history that would lead to an individual becoming infected. The simplest coupling arguments for demonstrating the closeness of the branching process approximations do not keep the processes identical for quite long enough, so that arguments showing that the differences are unimportant are also needed. In this talk, we show that, for some models, couplings can be constructed that are sufficiently accurate for this extra step to be dispensed with. Two such are the Reed–Frost and Markovian SIR models.

Asymptotically Distribution-Free Goodness-of-Fit Testing for Tail Copulas

Sami Umut **Can**
University of Amsterdam
(Joint work with John Einmahl, Estate Khmaladze and Roger Laeven)

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d. sample from a bivariate distribution function F. The asymptotic joint distribution of the standardized component-wise maxima $\bigvee_{i=1}^n X_i$ and $\bigvee_{i=1}^n Y_i$ is characterized by the marginal extreme value indices and the tail copula R associated with F. The extreme value indices specify the asymptotic marginal distributions of the standardized maxima, and the tail copula specifies the dependence structure. We propose a procedure for constructing asymptotically distribution-free goodness-of-fit tests for the tail copula R. The procedure is based on a transformation of a suitable empirical process derived from a semi-parametric estimator of R. The transformed empirical process converges weakly to a standard Wiener process, paving the way for a multitude of asymptotically distribution-free goodness-of-fit tests. In a simulation study we show that the limit theorems provide good approximations for finite samples and that tests based on the transformed empirical process have high power.

Modelling and Approximations: Dengue Fever Epidemics in Singapore

Daryl J. **Daley**
University of Melbourne

The talk describes the annual notifications of dengue fever (DF) in Singapore. DF was the major childhood disease in the 1960s, but a larval source reduction campaign, targeting the *Aedes aegypti* mosquito that spreads the viral disease from one human to another, reduced incidence of DF to almost nothing for over 15 years. The disease returned in the later 80s, and it is now again endemic. Can epidemic models help us understand the data? What approximations might help?

Dynamics of Gamma Bursts in Local Field Potentials

Priscilla (Cindy) **Greenwood**
University of British Columbia
(Joint work with Mark McDonnell and Lawrence Ward)

We provide a simple linear stochastic model of the generation of gamma bursts in local field potential (LFP) recordings from interacting populations of excitatory and inhibitory neurons. The stochastic model can be approximated in terms of a rotation multiplied by a two-dimensional Ornstein-Uhlenbeck process. Properties of the approximation are mirrored in LFP data simulated from the original model and may explain gamma bursts in real LFP data.

On some Brownian Filtrations

Kais **Hamza**
Monash University
(Joint work with Fima Klebaner and Jie Yen Fan)

Let B be a Brownian motion and f be a (smooth) function. The aim of this talk is to determine the natural filtration of the process $f(B)$, for as large a set of functions as possible. In particular we ask when such a filtration is that of a Brownian motion. As an example we show that the filtration generated by $h_n(B)$, where h_n is the Hermite polynomial of degree n, is that of B if n is odd, and that of $|B|$ if n is even. As a consequence we show that the martingales $H_n(B_t, t)$, where $H_n(z, a) = a^{n/2} h_n(z/\sqrt{a})$ and $H_n(z, 0) = z^n$, are not Markovian for $n \geq 3$. This fact is used when attempting to mimic $H_n(B_t, t)$ in the sense of constructing martingales whose marginal distributions match those of $H_n(B_t, t)$.

Network Tomography for Integer-Valued Traffic

Martin **Hazelton**
Massey University

Volume network tomography concerns making inference about relatively high dimensional properties of network flow (e.g. origin-destination traffic volumes) based on lower dimensional data – in particular, traffic counts on the network links. Inference is possible in principle for very general classes of traffic model using stochastic EM and MCMC methods. However, implementation requires that we sample route flows conditional on the traffic counts. These conditional route flows are constrained to lie in a convex polytope. Development of an efficient method of sampling from this polytope has challenged researchers for 20 years. We describe a new sampler with excellent properties, assuming that the network routing matrix to be totally unimodular. We discuss the implications of this assumption.

The Importance of Being Urn-est

Mark **Holmes**
University of Auckland

We'll describe some interesting generalisations of Polya's urn, which is the standard example of a random process with reinforcement. In particular we will see models with multiple phase transitions as we vary the parameter governing the strength of the reinforcement.

Unitary Transformations, Empirical Processes and Distribution Free Testing

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

The main message in the talk is that there are surprisingly many different Brownian bridges, some of them – familiar, some of them – less familiar. Many of these Brownian bridges are surprisingly close to Brownian motions. Somewhat loosely speaking, all the bridges can be conveniently mapped onto each other, and, hence, to one 'standard' bridge. Thus a unified theory of distribution free testing in $\mathbb{R}^d$, both for discrete and continuous cases, can be built without too much effort.

On Sustainability of Stochastic Systems

Jacek **Krawczyk**
Victoria University of Wellington

Scientists and politicians seek to understand what it means, practically and conceptually, to be sustainable. The aim of this talk is to introduce viability theory, a relatively young branch of mathematics which provides a conceptual framework that is very well suited to such problems. Viability theory can be used to answer important questions about the sustainability of systems, including those studied in environmental management and macroeconomics, and can be used to determine sustainable policies for their management. The principal analytical tool of viability theory is the viability kernel which, for deterministic dynamic systems, is the set of all state-space points in a constrained system starting from which it is possible to remain within the systems constraints indefinitely. For stochastic systems, this possibility becomes probabilistic, and if the system is perturbed by arbitrarily large disturbances it will be impossible to claim that any given point is viable with certainty. In this case, we speak about viability for a bounded "strength" of disturbance, determined by choosing some finite interval on the probability distribution of the disturbance. (Intuitively, the larger the disturbance interval, the smaller the viability kernel.) This talk provides an outline of the core concepts of viability theory and an overview of the numerical approaches available for computing approximate viability kernels. VIKAASA, a specialised software application developed and designed to compute such approximate viability kernels is presented along-side examples of viability theory in action in the spheres of bio-economics.

Robustness of Efficiency in Nearly True Models for Incomplete Data

Thomas **Lumley**
University of Auckland

Consider estimates of a possibly-misspecified probability model fitted to a complete cohort of data as defining the target of inference. Under unequal-probability sampling from the cohort it is possible to construct design-based estimators that are consistent for this target of inference. If in addition the probability model is assumed to be correct, it is typically possible to construct more efficient model-based estimators. I examine how badly the model has to be misspecified in order for the best design-based estimator to have lower mean squared error than the best model-based estimator, and find it is not very far.

Control of Markov Chains with Constraints based on Incomplete Information

Boris **Miller**
Monash University

Complex systems like broadband networks, power networks, water storage systems, operating in abruptly changing environments can be considered as hybrid systems driven by controlled Markov chains (MC), subjected on one side to different constraints and on the other side are not completely observable and such that their states are subject to abrupt changes. Generally the solution of the optimal control problem for such systems may be reduced to the dynamic programming equation (DPE) which is hard to solve even numerically. In our previous work where the states of the MC were completely observable we suggested the procedure for solving the problem with constraints with the aid of deterministic maximum principle and for obtain the Markov type optimal control. Moreover, if the generator of controlled MC does not depend on control one can obtain the optimal control as a function of conditional probabilities of the states. In this case there is no direct way to evaluate the optimal cost without solution of the dynamic programming equation; however, the value of the optimal cost corresponding to given initial condition may be evaluated by Monte-Carlo testing. Confidence level may be estimated by t-distribution, so for given optimal control one can obtain the confidence intervals either for the cost function or for the constraints values. It permits to solve the optimal control problem with constraints at a given confidence level. An important characterization of the attainable set of constrained criteria is that its closure is convex which permits to use the dual approach and to reduce the problem to the maxmin one which may be solved numerically.

Can One Hear the Shape of a Population History?

Nathan **Ross**
University of Melbourne
(Joint work with Junhyong Kim, Elchanan Mossel and Miklos Racz)

Kingman's coalescent is a mathematical object used in population genetics to model certain genealogies for a sample of individuals from a present day population. For given parameters, the model produces a random set of branch lengths between each pair in the sample which represent the number of generations in the past until the most recent common ancestor of a pair. A parameter of the model is the past population size as a function of the number of generations back in time from the present. And so a random sample of branch lengths from a present day population can be used for statistical inference of past population size. We discuss this topic and more generally how to estimate past population size and structure from present day genetic data using the coalescent, a topic that has received recent attention due to its implications for human evolutionary history.

Multi-Player Stopping and Quitting Games with Redistribution of Payoffs

Marek **Rutkowski**
University of Sydney
(Joint work with Ivan Guo and Tianyang Nie)

In the seminal paper by Dynkin (1969), he introduced the concept of a zero-sum, optimal stopping game between two players, where each player can stop the game for a payoff observable at that time. An abundant research was subsequently done on stopping games and their applications. In recent works by Guo and Rutkowski, a novel class of multi-period multi-player competitive stopping games with redistribution of payoffs is constructed. Each player can either exit the game for a fixed payoff, determined a priori, or stay and receive an adjusted payoff depending on the decision of other players. The single-period case is shown to be *weakly unilaterally competitive* under some assumptions on redistribution of payoffs. We present an explicit construction of the unique value at which optimal equilibria are attained. The multi-period stochastic extension of the game is also studied and solved by the backward induction. The game has interpretations in economic and financial contexts, for example, as a consumption model with bounded resources or a starting point to the construction of multi-person financial game options (for two-person game options, see Kifer (2000)). In the latter case, the concept of optimal equilibria becomes pivotal in the pricing of the multi-person game option through no-arbitrage arguments. Deterministic multi-period quitting games are also examined as an alternative to stopping games.

References

Dynkin, E. B. (1969) Game variant of a problem on optimal stopping. *Soviet Math. Dokl.* 10, 270–274.

Guo, I. and Rutkowski, M. (2012) A zero-sum multi-player game. *Demonstratio Math.* 45, 415–433.

Guo, I. and Rutkowski, M. (2013) Discrete-time multi-player stopping and quitting games with redistribution of payoffs. Working paper, University of Sydney.

Guo, I. and Rutkowski, M. (2013) Multi-person affine stochastic stopping games in discrete time. Working paper, University of Sydney.

Kifer, Y. (2000) Game options. *Finance and Stochastics* 4, 443–463.

Nie, T. and Rutkowski, M. (2013) Multi-player stopping games with redistribution of payoffs and multi-dimensional BSDEs with oblique reflection. Working paper, University of Sydney.

Rank Tests for Survey Data

Alastair **Scott**
University of Auckland
(Joint work with Thomas Lumley)

Rank tests are widely used for exploratory and formal inference in the health and social sciences. With the widespread use of data from complex surveys, where the standard exchangeability assumptions break down, there is increasing demand for versions of rank tests that account for the sampling design. In the absence of design-based rank tests, naive unweighted rank tests are being used in survey analyses even by researchers who otherwise use inferential methods appropriate for the sampling design. We propose a general approach to constructing design-based rank tests when comparing groups within a complex sample and when using a national survey as a reference distribution, and illustrate both scenarios with examples. We show that the tests have asymptotically correct level and that the relative power of different rank tests is not greatly affected by complex sampling.

Ancestral Reconstruction, Lateral Gene Transfer, and the Joys of Leaping Between Trees

Mike **Steel**
University of Canterbury

In part 1, I will present some recent results on reconstructing discrete ancestral states in a pure-birth tree under a Markovian process of character evolution. In part 2, I will describe recent work on the problem of inferring an evolutionary 'species tree' from gene trees, under a simple model of random lateral gene transfer (LGT). A typical question biologists ask is: 'could we reconstruct a species tree on (say) 200 species from lots of gene trees, if each gene has been laterally transferred into other lineages, on average, ten times?' Another is 'can LGT lead to statistically inconsistent tree estimation?' Our analysis involves a curious connection to random walks on cyclic graphs.

The Multivariate Kaplan-Meier Estimator

Winfried **Stute**
Justus Liebig University Giessen
(Joint work with Arusharka Sen)

Finding the multivariate extension of the famous Kaplan-Meier (KM) estimator is an important question which has attracted many researchers over the last thirty years. Since, for higher-dimensional data, the distribution function is no longer specified by the hazard function, traditional analysis focusing on the product limit approach is not feasible. Rather we propose a so-called mass-shifting method, by which we are able to uniquely solve an associated multivariate Volterra integral equation and obtain the multivariate extension of KM, both by solving an appropriate eigenvalue problem respectively computing the Neumann series of a Volterra operator. We study our new estimator for small and large sample sizes and derive its efficiency.