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

Sixth Wellington Workshop in Probability and Mathematical Statistics

4 – 5 December 2017
Victoria University of Wellington

Presenters, Titles and Abstracts

(**Ordered alphabetically**, by presenters' last names)

Bootstrap Percolation on Random Graphs

Omer **Angel**, University of British Columbia
(Joint work with Brett Kolesnik)
Day 1, Session 3

We consider bootstrap percolation on the Erdos-Renyi graph: given an initial infected set, a vertex becomes infected if it has at least r infected neighbours. The graph is susceptible if there exists an initial set of size r that infects the whole graph. We identify the critical threshold for susceptibility. We also analyse Bollobas's related graph-bootstrap percolation model.

Statistics of Ambiguous Rotations

Richard **Arnold**, Victoria University of Wellington
(Joint work with Peter Jupp)
Day 2, Session 4

Orientations of objects in $\mathbb{R}^p$ with symmetry group K cannot be described unambiguously by elements of the rotation group $SO(p)$, but correspond to elements of the quotient space $SO(p)/K$. Specifications of probability distributions and appropriate statistical methods for such objects have been lacking – with the notable exceptions of axial objects in the plane and in $\mathbb{R}^3$.

We exploit suitable embeddings of $SO(p)/K$ into spaces of symmetric arrays to provide a systematic and intuitively appealing approach to the statistical analysis of the orientations of such objects. We firstly consider the case of *orthogonal r-frames* in $\mathbb{R}^p$, corresponding to sets of $r \leq p$ orthogonal axes in p dimensions. Probability distributions in this setting include the Watson and Bingham distributions. Using the same approach we then treat the three dimensional case of $SO(3)/K$, where K is one of the point symmetry groups.

In both cases the resulting tools include measures of location and dispersion, tests of uniformity, one-sample tests for a preferred orientation and two-sample tests for a difference in orientation. The methods are illustrated using data from earthquake focal mechanisms (fault plane and slip vector orientations) and crystallographic data with orientation measurements of distinct mineral phases in EBSD experiments.

Schistosomiasis: Models and Data

Andrew **Barbour**, Universitaet Zuerich

Day 2, Session 6

Human schistosomiasis is a parasitic disease that is prevalent throughout the tropics, and closely related animal parasites infect agricultural livestock even in temperate climes. The first dynamical model of the transmission of the disease was formulated by Macdonald in 1965, and many variants were explored in the following decade. We briefly discuss the inadequacies of these models when confronted with data, and introduce a very different stochastic model that seems to represent the data much better. Some open problems connected with the analysis of the new model will be discussed.

On the Stability and the Uniform Propagation of Chaos Properties of Ensemble Kalman-Bucy Filters

Pierre **Del Moral**, INRIA (University of Bordeaux)

Day 1, Session 2

The Ensemble Kalman filter is a sophisticated and powerful data assimilation method for filtering high dimensional problems arising in fluid mechanics and geophysical sciences. This Monte Carlo method can be interpreted as a mean-field McKean-Vlasov type particle interpretation of the Kalman-Bucy diffusions. Besides some recent advances on the stability of nonlinear Langevin type diffusions with drift interactions, the long-time behaviour of models with interacting diffusion matrices and conditional distribution interaction functions has never been discussed in the literature. One of the main contributions of the talk is to initiate the study of this new class of models. The talk presents a series of new functional inequalities to quantify the stability of these nonlinear diffusion processes. The second contribution of this talk is to provide uniform propagation of chaos properties as well as $\mathbb{L}_p$-mean error estimates w.r.t. the time horizon.

The Power of Tests for Signal Detection in High Dimensional Data

Marc **Ditzhaus**, Heinrich Heine University Duesseldorf
(Joint work with Arnold Janssen)
Day 1, Session 1

In the literature weak and sparse (or dense) signals within high dimensional data or Big Data are well studied concerning detection, feature selection and estimation of the number of signals. In this talk we focus on the quality of detection tests for signals. Since the pioneering work of Donoho and Jin (2004) it was shown in the literature for different (mainly) parametric models that the detection boundary of the log-likelihood ratio test and Tukey's higher criticism test coincide asymptotically. In contrast to this it is less known about the behavior of tests on the detection boundary, especially for the higher criticism test. We fill this gap in great detail with the analysis on the detection boundary. We give general tools to handle the log-likelihood ratio and higher criticism statistics. To illustrate our results we apply them to a nonparametric model based on the chimeric alternatives of Khmaladze (1998) and get a nonparametric detection boundary for it. In particular, it results that the higher criticism test has no power on the nonparametric detection boundary while the log-likelihood ratio test has nontrivial power there.

Bibliography

Ditzhaus, M. and Janssen, A. (2017) The power of big data sparse signal detection tests on nonparametric detection boundaries. *Submitted* (arXiv:1709.07264).

Ditzhaus, M. (2017) *The Power of Tests for Signal Detection in High Dimensional Data.* PhD-thesis, Heinrich Heine University Duesseldorf.
https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=42808

Donoho, D. and Jin, J. (2004) Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* 32(3), 962–994.

Khmaladze, E. V. (1998) Goodness of fit tests for Chimeric alternatives, *Statist. Neerlandica* 52(1), 90–111.

Two Locality Properties in Two Dimensions

Jesse **Goodman**, University of Auckland
Day 1, Session 3

In two dimensions, many self-interacting processes are described by the Schramm-Loewner Evolution SLE(κ), a family of random fractal paths joining two boundary points of an underlying domain D. These continuous paths arise as the scaling limits of various discrete self-interacting paths, such as loop-erased random walk.

A self-interacting process has the locality property if it does not "feel" the boundary of its domain D until it hits the boundary. Among the two-dimensional processes known as Schramm-Loewner Evolution SLE(κ), it is known that only one, SLE(6), satisfies the locality property. In this talk, I will describe the key properties that identify SLE(6) – the Domain Markov Property, conformal invariance, and the (classical) Locality Property – and introduce a "non-local" form of locality also satisfied by SLE(6), describing the behaviour of the process when it first encloses a target set.

Mexican Hat Coupling of Quasi-Cycle Operators Produces Quasi-Patterns

Priscilla **Greenwood**, University of British Columbia
(Joint work with Lawrence Ward)
Day 1, Session 2

We use "quasi" in the sense "it doesn't work without noise". An ODE system with a fixed point reached by damped oscillations has sustained oscillations when noise is added to the system. The sustained oscillations are called "quasi-cycles". Similarly a PDE system with excitable states may fail to show a Turing-type pattern because of damping. When noise is added, a pattern called a "quasi-pattern" may appear. We replace the Laplacian in a PDE with a Mexican Hat (difference of Gaussians) convolution operator, and study how the resulting spatial and temporal quasi-oscillations interact. Our model is suggested by previous work on the brain basis of visual perception.

Statistical Generalized Derivative Applied to the Profile Likelihood Estimation in a Mixture of Semiparametric Models

Yuichi **Hirose**, Victoria University of Wellington
Day 1, Session 1

There is a difficulty in finding an estimate of variance of the profile likelihood estimator in the joint model of longitudinal and survival data. We solve the difficulty by introducing the "statistical generalized derivative". The derivative is used to show the asymptotic normality of the estimator without assuming the second derivative of the density function in the model exists.

Random Graphs Arising from Strong Reinforcement Models

Mark **Holmes**, University of Melbourne
Day 1, Session 3

We will consider certain reinforcement processes on graphs, and in particular the random subset of edges that are reinforced "infinitely often".

Possibilities for Testing Linear Regression

Estáte V. **Khmaladze**, Victoria University of Wellington
(Joint work with Laura Dumitrescu)
Day 2, Session 6

We all know that in testing the linear regression model,

$$Y_i = X_i\beta + e_i, \quad i = 1, \ldots, n$$

we start with estimated errors,

$$\hat{e}_i = Y_i - X_i\hat{\beta}.$$

All test procedures are to be based on the estimated errors. However, the distribution of these errors depends on covariates (or regressors) $X_i, i = 1, \ldots, n$. Many interesting and useful test statistics based on $\hat{e}_i, i = 1, \ldots, n$, will have distributions that are not easy to find. And so far, one has to find them for every new collection of possible $X_i, i = 1, \ldots, n$.

What we suggest is a one-to-one transformation of the sequence $\hat{e}_i, i = 1, \ldots, n$, into anther sequence, say $\varepsilon_i, i = 1, \ldots, n$, which has a distribution, independent from (or not involving) covariates. The transformation we have in mind is given by an appropriate unitary operator. We think it is a simple transformation and can be "practically" convenient.

The approach carries on to similar problems, like testing $AR(p)$ models in time series.

Distribution-Free Goodness of Fit Tests in Contingency Tables

Thuong **Nguyen**, Victoria University of Wellington
Day 2, Session 6

A new class of asymptotically distribution-free goodness of fit tests for testing independence in contingency tables has been suggested for the classic case when the total number of trials is fixed. These new tests perform reasonably well compared to the conventional chi-square test in several cases. We may also discuss a potential extension of the method to the case when the number of cells is increasing significantly, which leads to the sparsity of the table in which the chi-square test is no longer valid.

Rate of Convergence to Hermite-type Distributions

Andriy **Olenko**, La Trobe University, Melbourne

Day 2, Session 5

The case when the summands/integrands are functionals of a long-range dependent Gaussian process is of great importance in the theory of limit theorems for sums/integrals of dependent random variables. Comparing with the CLT, long-range dependent summands can produce different normalizing coefficients and non-Gaussian limits.

Integral functionals of homogeneous random fields/processes with long-range dependence are investigated. We will present some new results on the rate of convergence to the Hermite-type distributions in non-central limit theorems. The results were obtained under rather general assumptions on the spectral densities of random fields. These assumptions are even weaker than in the known convergence results for the case of Rosenblatt distributions. Additionally, Levy concentration functions for Hermite-type distributions were investigated.

Bibliography

Anh, V., Leonenko, N. and Olenko, A. (2015) On the rate of convergence to Rosenblatt-type distribution, *J. Math. Anal. Appl.* 425(1), 111–132.

Anh, V., Leonenko, N., Olenko, A. and Vaskovych, V. (2017) On rate of convergence in non-central limit theorems, arXiv:1703.05900.

Leonenko, N. and Olenko, A. (2014) Sojourn measures of Student and Fisher-Snedecor random fields, *Bernoulli* 20(3), 1454–1483.

Locally Robust Methods and Near-Parametric Asymptotics

Spiridon **Penev**, University of New South Wales

(Joint work with Kanta Naito)

Day 1, Session 1

Some past papers have demonstrated theoretically and numerically that infusing a little localisation in the likelihood-based methods for regression and for density estimation can actually improve the resulting estimators with respect to suitably defined global risk measures. Thus a variety of local likelihood methods have been suggested. In this paper, we demonstrate that similar effect can also be observed with respect to robust estimation procedures. Localised versions of robust density estimation procedures perform better with respect to suitably defined global risk measures based on minimization of Bregman divergence measures. The talk contains some theoretical statements, as well as numerical examples supporting our claims.

Perturbative Characterization of Ballisticity of Random Walks in i.i.d. Random Environments

Alejandro **Ramirez**, Catholic University of Chile
(Joint work with Santiago Saglietti)
Day 1, Session 3

We provide a sufficient condition for ballisticity for random walks in i.i.d. random environments which are perturbations of the simple symmetric random walk in terms of the expectation and variance of the drift at a single site, extending and giving a sharper version of a result of Sznitman from 2004. Our theorem gives new examples of ballistic random walks in random environments satisfying the polynomial decay condition (P), but which do not satisfy Kalikow's condition.

Distribution-Free Testing of Grouped Bernoulli Trials

Leigh **Roberts**, Victoria University of Wellington
Day 2, Session 4

Recently Khmaladze has shown how to 'rotate' one empirical process to another. We apply this rotation to Bernoulli trials, when the distribution of successive trials depends on covariates. A new result is that subgroups of trials need to be rotated rather than individual trials. For goodness of fit tests, the Kolmogorov Smirnov statistic applied to distribution free rotated processes has high power. Additionally, the rotation is readily computed and possesses good convergence properties.

Keywords: Covariates, binomial trials, Kolmogorov-Smirnov, logistic distribution, rotation, unitary transform.

Robust Hemodynamic Response Function Estimation and Brain Activity Detection from fNIRS Signals

Karim **Seghouane**, University of Melbourne
(Joint work with Davide Ferrari)
Day 1, Session 2

Hemodynamic response function (HRF) estimation in functional near-infrared spectroscopy (fNIRS) plays an important role in characterizing the temporal dynamics of the brain response. Estimation based on semiparametric modelling is very useful for fNIRS signals. However, the Gaussian noise assumption may be too simplistic since the sources of noise and their nature are various in fNIRS. In this study, we consider the problem of HRF estimation using a semiparametric model whose nonparametric part is viewed as a nuisance component used to represent the drift. We describe a new robust HRF estimation procedure to minimize the impact of unexpected noise in fNIRS signals. Within the proposed estimation method, the drift effect is removed by applying a first order difference to the fNIRS signal samples. Consistency and asymptotic normality of the proposed estimator are established and its effectiveness is illustrated on both simulated and real fNIRS data.

Tricks for Trees: Probabilistic Techniques
for Phylogenetic Combinatorics

Mike **Steel**, University of Canterbury, Christchurch
Day 2, Session 4

Phylogenetic trees are widely used in evolutionary biology and related areas (e.g. linguistics, epidemiology, anthropology, medicine). These trees are typically reconstructed from discrete character data (e.g. genes), by methods that assume these characters evolved from a common ancestor under a Markovian process. Accordingly, probability theory plays a central role in the inference and analysis of these trees. In this talk, I will describe how probability theory can also shed light on some important combinatorial properties of phylogenetic trees. By applying classical probability-based techniques (e.g. Pólya urns, the probabilistic method) we obtain shorter and more insightful proofs of several combinatorial results.

Markov Mixture Process with Multi Absorbing States

Budhi **Surya**, Victoria University of Wellington
Day 2, Session 5

In this talk I will discuss a mixture of continuous-time Markov chains with multi absorbing states moving at different speeds on the same state space, where the mixture occurs at random time. A variety of associated distributional properties of the mixture process are discussed. Identities are explicit in terms of the Bayesian updates of switching probability and the intensity matrices of the underlying Markov chains, despite the fact that the mixture process is not Markovian. They form non-stationary function of time and duration and have the appealing features of capturing heterogeneity and path dependence when conditioning on the available information (either full or partial) of the past history of the process. The results extend further the recent work of Surya (2017) to the presence of censoring/competing risks.

Reference

Surya, B. A. (2017) Distributional properties of the mixture of continuous-time absorbing Markov chains moving at different speeds. Accepted by *Stochastic Systems – INFORMS Applied Probability Society.*

Spatial Pattern Formation via Stochastic Neural Field Equations

Lawrence **Ward**, University of British Columbia
(Joint work with Cindy Greenwood)
Day 1, Session 2

The formation of spatial patterns in biological systems may be modeled by a set of reaction-diffusion equations. A diffusion-type coupling operator biologically significant in neuroscience is a difference of Gaussian functions (Mexican Hat operator) used as a spatial-convolution kernel. Here we study the simplest reaction-diffusion system with this type of coupling. We are interested in the difference among behaviors of *stochastic* neural field equations, namely space-time stochastic differential-integral equations, and similar deterministic ones. We explore, quantitatively, how the parameters of our model that measure the shape of the coupling kernel, coupling strength, and aspects of the spatially-smoothed space-time noise, control the pattern in the resulting evolving random field. We find that a spatial pattern that is damped in time in a deterministic system may be sustained and amplified by stochasticity, most strikingly at an optimal spatio-temporal noise level. In addition, we find that spatially-smoothed noise alone causes pattern formation even without spatial coupling by the Mexican Hat operator.

The Distribution of the Minimum of a Positive Sample

Kit **Withers**, Wellington
Day 2, Session 5

I give expansions for the distribution, density and moments of the sample minimum when sampling from any distribution on $[0, u]$ that is nearly analytic at 0. Here $0 < u \leq \infty$. Examples include the gamma distribution.

New Universal Distributions arising from Correlated Random Variables: Examples from Random Matrices

Nicholas **Witte**, Massey University, Palmerston North
Day 2, Session 5

When i.i.d. random variables are strongly correlated then the Gaussian or normal distribution does not appear in any limiting sense as the number of variables grows. Such a situation arises with the singular values of Wishart or double-Wishart matrices. The single and double Wishart matrices under the null hypothesis are also known as the Laguerre and Jacobi Orthogonal Ensembles from the viewpoint of random matrix theory. This is an area where a number of very different disciplines have overlapped ranging from mathematical physics, combinatorics, number theory, analysis and not least of all, statistics.

What one gets instead are new distributions such as the Tracy-Widom distribution for the largest eigenvalue of a random real symmetric matrix, first identified in 1994. These new distributions are believed to be universal in a similar way to the normal distribution, in that they are the limiting form (as the matrix rank grows) of broad categories of population distributions. However since 1994 additional ones have been revealed such as in studies of the largest eigenvalue λ_1 of certain non-null sample covariance matrices, in particular the complex ones. Here one has large values of the sample size n and the number of variables p. It has been shown that the resulting distribution of λ_1 as $n, p \to \infty$ can swing from one behaviour to another depending on their ratio. Even for finite n, p this can mean that the eigenvector of the sample PCA (e.g. associated with eigenvalue λ_1) may exhibit a sharp loss of tracking suddenly losing its relation to the eigenvector of the population PCA. In statistical mechanics language the system exhibits a "phase transition" – the BBP transition after Baik, Ben Arous and Péché (2005).

A review of recent works investigating the current state of affairs beyond the null case, such as the above, will be given. A great deal of detailed information is known about these universal distributions exploiting the fact they are intimately connected with integrable systems, such as the class of Painlevé transcendants.

Reference

Baik, J., Ben Arous, G. and Péché, S. (2005) Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, *The Annals of Probability* 33(5), 1643–1697.