PROGRAMME AND ABSTRACTS

9th International Conference on

Computational and Financial Econometrics (CFE 2015)

http://www.cfenetwork.org/CFE2015

and

8th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (CMStatistics 2015)

http://www.cmstatistics.org/CMStatistics2015

Senate House & Birkbeck University of London, UK
12 – 14 December 2015

ISBN 978-9963-2227-0-4

Technical Editors: Angela Blanco-Fernandez and Gil Gonzalez-Rodriguez.

EO164 Room Court BAYESIAN SEMI- AND NONPARAMETRIC MODELLING III Chair: Antonio Lijoi

EO0180: **Bayesian inference for intra-tumor heterogeneity in mutations and copy number variation**
Presenter: **Juhee Lee**, University of California Santa Cruz, United States
Co-authors: Peter Mueller, Yuan Ji, Subhajit Sengupta, Kamalakar Gulukota
Tissue samples from the same tumor are heterogeneous. They consist of different subclones that can be characterized by differences in DNA nucleotide sequences and copy numbers on multiple loci. Inference on tumor heterogeneity involves the identification of the subclonal copy number and single nucleotide mutations at a selected set of loci. We estimate such tumor heterogeneity on the basis of a Bayesian feature allocation model. We jointly model subclonal copy numbers and the corresponding allele sequences for the same loci. The proposed method utilizes three random matrices, L, Z and w to represent subclonal copy numbers (L), the number of subclonal variant alleles (Z) and the cellular fractions (w) of subclones in one or more tumor samples, respectively. The unknown number of subclones implies a random number of columns for these matrices and having more than one subclone indicates tumor heterogeneity. We estimate the subclonal structures through inference on these three matrices, using next-generation sequencing data. Using simulation studies and a real data analysis, we demonstrate how posterior inference on the subclonal structure is enhanced with the joint modeling of both structure and sequencing variants on subclonal genomes. R package is available at http://cran.rproject.org/web/packages/ BayClone2/index.html.

EO0927: **Bayesian ordination of species sampling from dependent populations with applications to microbiome studies**
Presenter: **Sergio Bacallado**, Cambridge University, United States
Co-authors: Boyu Ren, Lorenzo Trippa, Stefano Favaro, Susan Holmes
We introduce a Bayesian nonparametric prior for sampling from many discrete distributions. The distributions are correlated through continuous latent factors, whose dimensionality is shrunk by the prior. These factors can be used in exploratory data analysis, and we discuss the visualization of credible regions through multivariate analysis methods. It is also possible to include factors for fixed effects if the distributions are characterized by covariates. The marginal prior on each distribution is a normalized generalised Gamma process, a well-known nonparametric prior. We apply Markov chain Monte Carlo and variational Bayesian inference methods. The procedure is applied to the analysis of microbiome studies, which produce discrete samples from the distribution of bacterial species in several environments. The goal of these studies is to understand the variation of species abundance across environments and their relation to covariates. The analysis deals naturally with the uncertainty in the distribution of species, which is important when the sampling depth is low.

EO1096: **Multi-resolution scanning for distributional variation via the k-group design**
Presenter: **Li Ma**, Duke University, United States
Co-authors: Jacopo Soriano
An inference task that holds key to numerous applications is the comparison of multiple data sets to identify the underlying difference. The most common setting is the k-sample problem, in which the samples are collected under k different experimental settings or treatment/response statuses. A fundamental challenge in modern k-sample problems is the presence of many potential confounders, or extraneous sources of variation, that can contribute to the difference across the sampling distributions, but are difficult to control explicitly. This results in numerous irreplicable findings in many modern applications such as functional genomics. A simple experiment design that allows the intrinsic (i.e. scientifically interesting) variation to be identified from the extraneous ones is the k-group design, under which replicate data sets are collected under each experiment setting. A fully probabilistic multi-resolution scanning framework is presented based on local hierarchical Binomial testing for analyzing k-group studies that effectively identifies actual cross-group variation while adjusting for extraneous ones. It takes into account the spatial heterogeneity of cross-sample variation and is particularly effective for finding highly local cross-group variation. It is also highly computationally efficient, suitable for real-time applications.

EO0758: **On some exchangeable increment processes derived from Bayesian ideas**
Presenter: **Ramses Mena**, UNAM, Mexico
In several applied areas, such as finance and risk theory, the use of stochastic models with the property of having independent increments is predominant. This is due perhaps to the mathematical tractability such a property implies. However, some phenomena require other kind of dependence structure and thus models relaxing such assumption are required. We discuss some models with exchangeable increments that, while relaxing the independent increments assumption, remain mathematically and computationally tractable. Our proposals borrow some Bayesian ideas for their construction. In particular, some applications in risk theory and finance will be presented.

EO330 Room Chancellor's Hall COPULA MODELS AND APPLICATIONS Chair: Claudia Tarantola

EO0237: **Bayesian copula modelling in the presence of covariates**
Presenter: **Julian Stander**, Plymouth University, United Kingdom
Co-authors: Luciana Dalla Valle, Charlotte Taglioni, Mario Cortina Borja
Copula models separate the dependence structure in a multivariate distribution from its univariate marginals, so overcoming many of the issues associated with commonly used statistical modelling methods by allowing, for example, different complex asymmetric dependencies and tail behaviours to be modelled. We discuss the modelling of bivariate data using copulas, of which there are now a rich choice. The parameter or parameters of the copula density are modelled as a function of a covariate using a natural cubic spline. Working in the Bayesian framework, we perform inference on the natural cubic spline and an associated smoothing parameter. We also discuss the choice of the copula density itself. We illustrate our approach using data from finance and medicine. We outline the extension of our methodology to more than one covariate and to multivariate data.

EO0494: **Tree copula mixture distribution for multivariate dependence analysis: An application to energy data**
Presenter: **Federico Bassetti**, Univeristy of Pavia, Italy
Co-authors: Maria Elena De Giuli, Enrica Nicolino, Claudia Tarantola
Motivated by the study of the energy market, a Bayesian analysis for a panel AR(p) model is proposed where the multivariate distribution of the innovations is described by a mixture of tree copula distributions. We assume that the innovations of the AR(p) model have normal marginal distributions, without assuming their joint normality. The use of the copulas allows us to split the dependence structure from the marginal distributions. Hence, we can represent a multivariate distribution through its univariate marginal distributions and a copula that captures the dependence structure. Although in the multivariate case the Gaussian and Student$-t$ copulas are the most used, they often are not flexible enough to represent the dependence structure of the data. A possible solution is the use of tree copula: particular types of Markov tree distributions that allow to represent the multivariate joint density through a suitable set of bivariate densities. In particular, in a tree copula the bivariate distributions over the edges of the underlying tree structure are specified by bivariate copulas. In order to develop a more flexible model and to take into account more complex dependence structure, we consider finite mixture of tree copulas. We use a MCMC algorithm to obtain posterior distributions.