

BISP12 ONLINE WORKSHOP

The Twelfth Workshop on Bayesian Inference in Stochastic Processes

27-28 May 2021



PROGRAM – DAY 1

27 May 2021

CEST time
13.55-14.00

Opening

Chair: Fabrizio Ruggeri, CNR-IMATI, Italy

Invited talks

14.00-14.40

Talk 1 - Approximate Bayesian Conditional Copulas

Clara Grazian, Luciana Dalla Valle, Brunero Liseo, Alexander McInner

Discussant: Maria Concepcion Ausin, Universidad Carlos III, Spain

Chair: Sonia Petrone, Università Bocconi, Italy

14.45-15.25

Talk 2 - Bayesian Dynamic Fused LASSO

Kaoru Irie

Discussant: Sylvia Frühwirth-Schnatter, Vienna University of Economics and Business, Austria

15.30-16.10

Talk 3 - Multiscale Bayesian Survival Analysis

Stéphanie van der Pas, Ismaël Castillo

Discussant: Antonio Lijoi, Università Bocconi, Italy

Chair: David Rios Insua, CSIC-ICMAT, Spain

16.15-16.55

Talk 4 - Bayesian Modeling of Power Outages and Their Consequences

Atilla Ay, Suleyman Ozekici, Refik Soyer

Discussant: Melike Baykal-Gursoy, Rutgers University, USA

17.00-17.40

Talk 5 - From viral evolution to spatial contagion: a biologically modulated Hawkes model

Andrew J. Holbrook, Xiang Ji, Marc A. Suchard

Discussant: Katja Ickstadt, Technical University Dortmund, Germany

Virtual poster presentations

Virtual posters are available at the link: bisp12.imati.cnr.it/virtual_poster.php

Chair: Elisa Varini, CNR-IMATI, Italy

- 17.45-18.35** **Poster 1 - A matrix-variate t model for networks**
Monica Billio, Roberto Casarin, Michele Costola, [Matteo Iacopini](#)
- Poster 2 - Forecasting count series in retail**
[Bruno Flores](#)
- Poster 3 - Latent Branching Trees: A Novel Class of Semi-Parametric Time Series Models**
[Theodore Kypraios](#), [Paul Fearnhead](#), [Gareth O. Roberts](#)
- Poster 4 - Predictive inference with Fleming–Viot-driven dependent Dirichlet processes**
[Filippo Ascolani](#), [Antonio Lijoi](#), [Matteo Ruggiero](#)
- Poster 5 - More for less: predicting and maximizing genomic variant discovery via Bayesian nonparametrics**
[Lorenzo Masoero](#), [Federico Camerlenghi](#), [Stefano Favaro](#), [Tamara Broderick](#)
- Poster 6 - Local Exchangeability**
[Trevor Campbell](#), [Saifuddin Syed](#), [Chiao-Yu Yang](#), [Michael Jordan](#), [Tamara Broderick](#)
- Poster 7 - Spatially meshed Markov chain Monte Carlo**
[Michele Peruzzi](#), [David B. Dunson](#)
- Poster 8 - Efficient Bayesian Model Selection for Coupled Hidden Markov Models with Weak Inter-Chain Dependencies**
[Jake Carson](#), [Trevelyan J McKinley](#), [Peter Neal](#), [Simon EF Spencer](#)

Invited talk 1 – 27 May 2021 | 14:00-14:40 CEST

Approximate Bayesian Conditional Copulas

Clara Grazian ⁽¹⁾, Luciana Dalla Valle ⁽²⁾, Brunero Liseo ⁽³⁾, Alexander McInner ⁽¹⁾

(1) University of New South Wales, Australia, (2) Plymouth University, United Kingdom, (3) Sapienza Università di Roma

c.grazian@unsw.edu.au

Abstract: Many proposals are now available to model complex data, in particular thanks to the recent advances in computational methodologies and algorithms which allow to work with complicated likelihood functions in a reasonable amount of time.

However, it is, in general, difficult to analyse data characterized by complicated forms of dependence. Copula models have been introduced as probabilistic tools to describe a multivariate random vector via the marginal distributions and a copula function which captures the dependence structure among the vector components, thanks to the Sklar's theorem, which states that any d -dimensional absolutely continuous density can be uniquely represented as the product of the marginal distributions and the copula function. Major areas of application include econometrics, hydrological engineering, biomedical science, signal processing and finance. Bayesian methods to analyse copula models tend to be computational intensive or to rely on the choice of a particular copula function, in particular because methods of model selection are not yet fully developed in this setting. We will present a general method to estimate some specific quantities of interest of a generic copula by adopting an approximate Bayesian approach based on an approximation of the likelihood function. Our approach is general, in the sense that it could be adapted both to parametric and nonparametric modelling of the marginal distributions and can be generalised in presence of covariates. It also allows to avoid the definition of the copula function. We will apply the methods to a real-data example, to identify the impact of the regularity of OPEC announcements on energy assets prices.

Keywords: Copulas; Approximate Bayesian computation; OPEC; GARCH; Gaussian processes.

Discussant: Maria Concepcion Ausin, Universidad Carlos III, Spain

Invited talk 2 – 27 May 2021 | 14:45-15:25 CEST

Bayesian Dynamic Fused LASSO

Kaoru Irie⁽¹⁾

(1) The University of Tokyo, Japan

irie@e.u-tokyo.ac.jp

Abstract: We study a new class of Markov processes whose transition density consists of two penalty functions that shrink the current state variable to its previous value and zero, similarly to Bayesian fused lasso in its functional form. The normalizing constant of the transition density, whose closed form is not available in the general Bayesian fused lasso, is shown to be essentially the log-geometric mixture of double-exponential densities, which enables the posterior computation by a Gibbs sampler. As the prior for time-varying regression coefficients of dynamic linear models, this process represents two conflicting shrinkage effects on the magnitude and dynamics of coefficients. In the context of high-dimensional dynamic regression, we compare the proposed process with the dynamic Bayesian lasso and horseshoe priors via the estimation and prediction of simulated data.

Keywords: Bayesian fused LASSO; Dynamic linear models; Horseshoe priors; Shrinkage.

Discussant: Sylvia Frühwirth-Schnatter, Vienna University of Economics and Business, Austria

Invited talk 3 – 27 May 2021 | 15:30-16:10 CEST

Multiscale Bayesian Survival Analysis

Stéphanie van der Pas⁽¹⁾, *Ismaël Castillo*⁽²⁾

(1) Amsterdam UMC, Department of Epidemiology and Data Science, The Netherlands,

(2) Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, France

s.l.vanderpas@amsterdamumc.nl, ismael.castillo@upmc.fr

Abstract: We consider Bayesian nonparametric inference in the right-censoring survival model, where modeling is made at the level of the hazard rate. We derive posterior limiting distributions for linear functionals of the hazard, and then for ‘many’ functionals simultaneously in appropriate multiscale spaces. As an application, we derive Bernstein-von Mises theorems for the cumulative hazard and survival functions, which lead to asymptotically efficient confidence bands for these quantities. Further, we show optimal posterior contraction rates for the hazard in terms of the supremum norm. In medical studies, a popular approach is to model hazards a priori as random histograms with possibly dependent heights. This and more general classes of arbitrarily smooth prior distributions are considered as applications of our theory. A sampler is provided for possibly dependent histogram posteriors. Its finite sample properties are investigated on both simulated and real data experiments.

Keywords: Frequentist analysis of Bayesian procedures; Survival analysis; Nonparametric Bernstein–von Mises theorem; Supremum norm contraction rate.

Discussant: Antonio Lijoi, Università Bocconi, Italy

Invited talk 4 – 27 May 2021 | 16:15-16:55 CEST

Bayesian Modeling of Power Outages and Their Consequences

Atilla Ay⁽¹⁾, Suleyman Ozekici⁽²⁾, Refik Soyer⁽¹⁾.

(1) George Washington University, USA, (2) Koc University, Turkey

aay@gwu.edu; sozekici@ku.edu.tr; soyer@gwu.edu

Abstract: In this paper, we introduce a general framework for doubly stochastic Markov process and present one of its special case, a Markov modulated compound Poisson process, to model the number of households affected by power outages. In doing so, we develop Bayesian inference for the model using a Gibbs sampler based on an extension of the data augmentation algorithm of Fearnhead and Sherlock for Markov modulated Poisson processes. We present an algorithm to predict the number of future outages and the number of households affected by those outages. We consider several extensions of the proposed model including order restrictions to overcome the well-known label switching problem, incorporating covariates, modeling outage hours and non-homogeneous compound Poisson processes. We use simulated, as well as actual power outage data to demonstrate the implementation of our proposed model and the methodology.

Keywords: Markov Modulated Compound Poisson Processes; Bayesian Analysis; Power Outage Modeling.

Discussant: Melike Baykal-Gursoy, Rutgers University, USA

Invited talk 5 – 27 May 2021 | 17:00-17:40 CEST

From viral evolution to spatial contagion: a biologically modulated Hawkes model

Andrew J. Holbrook⁽¹⁾, Xiang Ji⁽²⁾, Marc A. Suchard⁽¹⁾.

(1) University of California, Los Angeles, USA, (2) Tulane University, USA

aholbroo@g.ucla.edu; xji4@tulane.edu; msuchard@ucla.edu

Abstract: Mutations sometimes increase contagiousness for evolving pathogens. During an epidemic, epidemiologists use viral genetics to infer a shared evolutionary history and connect this history to geographic spread. We propose a model that directly relates a pathogen's evolution to its spatial contagion dynamics – effectively combining the two epidemiological paradigms of phylogenetic inference and self-exciting process modeling – and apply this *phylogenetic Hawkes process* to a Bayesian analysis of 23,422 viral cases from the 2014-2016 Ebola outbreak in West Africa. With a mere 1,610 samples providing RNA data, our model is able to detect subsets of the Ebola virus with significantly elevated rates of spatiotemporal propagation.

Keywords: Bayesian phylogeography; Ebola virus; Spatiotemporal Hawkes processes.

Discussant: Katja Ickstadt, Technical University Dortmund, Germany

Poster 1 – 27 May 2021 | 17:45-18:35 CEST

A matrix-variate t model for networks

Monica Billio⁽¹⁾, Roberto Casarin⁽¹⁾, Michele Costola⁽¹⁾, Matteo Iacopini^(2,3)

(1) Ca' Foscari University of Venice, Italy, (2) Vrije Universiteit Amsterdam, The Netherlands, (3) Tinbergen Institute, The Netherlands.

billio@unnive.it; r.casarin@unive.it; michele.costola@unive.it; m.iacopini@vu.nl

Abstract: Networks represent a useful tool to describe relationships among financial firms and network analysis has been extensively used in recent years to study financial connectedness. An aspect, which is often neglected, is that network observations come with errors from different sources, such as estimation and measurement errors, thus a proper statistical treatment of the data is needed before network analysis can be performed. We show that node centrality measures can be heavily affected by random errors and propose a flexible model based on the matrix-variate t distribution and a Bayesian inference procedure to de-noise the data. We provide an application to a network among European financial institutions.

Keywords: Bayesian; financial markets; matrix-variate distributions; networks; t distribution.

Poster 2 – 27 May 2021 | 17:45-18:35 CEST

Forecasting count series in retail

Bruno Flores ⁽¹⁾

(1) ICMAT-CSIC, Spain.

bruno.flores@icmat.es

Abstract: Large-scale dynamic forecasting of non-negative count series is a major challenge in many areas like epidemic monitoring or retail management. We propose Bayesian state-space models that are flexible enough to adequately forecast high and low count series and exploit cross-series relationships with a multivariate approach.

This is illustrated with a large scale sales forecasting problem faced by a major retail company, integrated within its inventory management planning methodology. The company has hundreds of shops each one with thousands of references.

Keywords: Count time series; Sales forecasting; Bayesian analysis; Dynamic generalized linear models; Inventory management.

Poster 3 – 27 May 2021 | 17:45-18:35 CEST

Latent Branching Trees: A Novel Class of Semi-Parametric Time Series Models

Theodore Kypraios⁽¹⁾, Paul Fearnhead⁽²⁾, Gareth O. Roberts⁽³⁾

(1) University of Nottingham, United Kingdom, (2) Lancaster University, United Kingdom, (3) University of Warwick, United Kingdom.

*theodore.kypraios@nottingham.ac.uk; p.fearnhead@lancaster.ac.uk;
gareth.o.roberts@warwick.ac.uk*

Abstract: A novel class of semi-parametric time series models is presented in which the marginal distribution of the observations is specified in advance, and their serial dependence structure is built around them by introducing an underlying stochastic process termed as a “Latent Branching Tree”. We first describe how such a stochastic process is constructed and discuss its properties, as well as how it leads to the development of a general class of time series models. We demonstrate how one can simulate from this stochastic process *exactly* (in a Monte Carlo sense) and how to draw Bayesian inference for the model parameters by exploiting the model’s structure and developing efficient Markov Chain Monte Carlo algorithms. Finally, we illustrate the proposed class of models by fitting them to genomic and internet traffic data to enhance our understanding of the covariance structure of these data.

Keywords: Time series; MCMC; Copulas; Trees.

Poster 4 – 27 May 2021 | 17:45-18:35 CEST

Predictive inference with Fleming–Viot-driven dependent Dirichlet processes

Filippo Ascolani ⁽¹⁾, Antonio Lijoi ⁽¹⁾, Matteo Ruggiero ⁽²⁾

(1) Department of Decision Sciences, Bocconi University, Milano, Italy, (2) Collegio Carlo Alberto and ESOMAS Department, University of Torino, Italy.

filippo.ascolani@phd.unibocconi.it; antonio.lijoi@unibocconi.it; matteo.ruggiero@unito.it

Abstract: We consider predictive inference using a class of temporally dependent Dirichlet processes driven by Fleming–Viot diffusions, which have a natural bearing in Bayesian nonparametrics and lend the resulting family of random probability measures to analytical posterior analysis. Formulating the implied statistical model as a hidden Markov model, we fully describe the predictive distribution induced by these Fleming–Viot-driven dependent Dirichlet processes, for a sequence of observations collected at a certain time given another set of draws collected at several previous times. This is identified as a mixture of Pólya urns, whereby the observations can be values from the baseline distribution or copies of previous draws collected at the same time as in the usual Pólya urn, or can be sampled from a random subset of the data collected at previous times. We characterise the time-dependent weights of the mixture which select such subsets and discuss the asymptotic regimes. We describe the induced partition by means of a Chinese restaurant process metaphor with a *conveyor belt*, whereby new customers who do not sit at an occupied table open a new table by picking a dish either from the baseline distribution or from a time-varying offer available on the conveyor belt. We lay out explicit algorithms for exact and approximate posterior sampling of both observations and partitions, and illustrate our results on predictive problems with synthetic and real data.

Keywords: Chinese restaurant; conveyor belt; random partition; hidden Markov model; generalized Pólya urn; predictive distribution.

Poster 5 – 27 May 2021 | 17:45-18:35 CEST

More for less: predicting and maximizing genomic variant discovery via Bayesian nonparametrics

Lorenzo Masoero⁽¹⁾, Federico Camerlenghi⁽²⁾, Stefano Favaro⁽³⁾, Tamara Broderick⁽¹⁾

(1) Massachusetts Institute of Technology, USA, (2) University of Milano - Bicocca, Italy, (3) University of Torino and Collegio Carlo Alberto, Italy.

lom@mit.edu; federico.camerlenghi@unimib.it; stefano.favaro@unito.it; tbroderick@csail.mit.edu

Abstract: While the cost of sequencing genomes has decreased dramatically in recent years, this expense often remains non-trivial. Under a fixed budget, then, scientists face a natural trade-off between quantity and quality: spending resources to sequence a greater number of genomes (quantity) or spending resources to sequence genomes with increased accuracy (quality). Our goal is to find the optimal allocation of resources between quantity and quality. Optimizing resource allocation promises to reveal as many new variations in the genome as possible. In this paper, we introduce a Bayesian nonparametric methodology to predict the number of new variants in a follow-up study based on a pilot study. We validate our method on cancer and human genomics data. When experimental conditions are kept constant between the pilot and follow-up, we find that our prediction is competitive with the best existing methods.

Unlike current methods, though, our new method allows practitioners to change experimental conditions between the pilot and the follow-up. We demonstrate how this distinction allows our method to be used for more realistic predictions and for optimal allocation of a fixed budget between quality and quantity.

Keywords: Bayesian nonparametric inference; Bernoulli process; genomic variation; gnomAD project; empirical Bayes; new variants discovery; optimal experimental design; sequencing error; three-parameter beta process prior.

Poster 6 – 27 May 2021 | 17:45-18:35 CEST

Local Exchangeability

Trevor Campbell⁽¹⁾, Saifuddin Syed⁽¹⁾, Chiao-Yu Yang⁽²⁾, Michael Jordan⁽²⁾, Tamara Broderick⁽³⁾

(1) University of British Columbia, Canada, (2) University of California, Berkeley, USA, (3) Massachusetts Institute of Technology, USA.

trevor@stat.ubc.ca

Abstract: Exchangeability – in which the distribution of an infinite sequence is invariant to reorderings of its elements – implies the existence of a simple conditional independence structure that may be leveraged in the design of probabilistic models, efficient inference algorithms, and randomization-based testing procedures. In practice, however, this assumption is too strong an idealization; the distribution typically fails to be *exactly* invariant to permutations and de Finetti’s representation theory does not apply. Thus there is the need for a distributional assumption that is both weak enough to hold in practice, and strong enough to guarantee a useful underlying representation. We introduce a relaxed notion of *local exchangeability* – where swapping data associated with nearby covariates causes a bounded change in the distribution. We prove that locally exchangeable processes correspond to independent observations from an underlying measure-valued stochastic process. We thereby show that de Finetti’s theorem is robust to perturbation and provide further justification for the Bayesian modelling approach. Using this probabilistic result, we develop three novel statistical procedures for (1) estimating the underlying process via *local empirical measures*, (2) testing via *local randomization*, and (3) estimating the *canonical premetric* of local exchangeability. These three procedures extend the applicability of previous exchangeability-based methods without sacrificing rigorous statistical guarantees. The talk will conclude with empirical results from these methods and a discussion of popular statistical models that exhibit local exchangeability.

Keywords: Exchangeability, randomization tests, local empirical measures.

Spatially meshed Markov chain Monte Carlo

Michele Peruzzi ⁽¹⁾, David B. Dunson ⁽¹⁾

(1) Department of Statistical Science, Duke University.

michele.peruzzi@duke.edu; dunson@duke.edu

Abstract: Complex relationships between a set of inputs and multiple outcomes at n measurement locations can be characterized in Bayesian hierarchical models via latent Gaussian Processes (GPs). Several solutions to the poor performance of GPs on big data (large n) settings have been tailor-made to work in diverse contexts, but may lack in flexibility, interpretability, or have limited scope for applicability. For example, computing methods for posterior sampling of Bayesian hierarchical models in these contexts are typically restricted to Gaussian first stages thanks to their analytical tractability. In more general cases, approximations of the joint posterior distribution are typically sought. Here, we consider spatial process models constructed on patterned directed acyclic graphs and domain partitioning – a procedure we refer to as spatial meshing. We show that spatial meshing leads to a class of *exact* posterior computing algorithms for general spatial big data regression models, which we label as spatially meshed Markov chain Monte Carlo (MEMCMC). In particular, we show that the Gibbs sampler arising from latent meshed GP regressions with Gaussian first stages can be straightforwardly generalized to a spatially meshed Metropolis adjusted Langevin sampling method defined on the Riemann manifold (MELANGE), which enables efficient posterior sampling of all unknowns (including the covariance hyperparameters) even with data in the millions, as demonstrated by a multivariate regression application on misaligned spatiotemporal satellite imaging data with $n > 2.5 \cdot 10^6$.

Keywords: Bayesian; big data; spatial; directed acyclic graphs; Gaussian processes; Riemann manifold MALA; Markov chain Monte Carlo.

Poster 8 – 27 May 2021 | 17:45-18:35 CEST

Efficient Bayesian Model Selection for Coupled Hidden Markov Models with Weak Inter-Chain Dependencies

Jake Carson⁽¹⁾, *Trevelyan J McKinley*⁽²⁾, *Peter Neal*⁽³⁾, *Simon EF Spencer*⁽¹⁾

(1) University of Warwick, UK, (2) University of Exeter, UK, (3) University of Nottingham, UK.

Jake.Carson@warwick.ac.uk; T.McKinley@exeter.ac.uk; Peter.Neal@nottingham.ac.uk; S.E.F.Spencer@warwick.ac.uk

Abstract: Performing model selection for coupled hidden Markov models (CHMMs) is highly challenging, owing to the large dimension of the hidden state process. Whilst in principle the hidden state process can be marginalized out via forward filtering, in practice the computational cost of doing so increases exponentially with the number of coupled chains, making this approach infeasible in most applications. Monte Carlo methods can be utilized, but despite many remarkable developments in model selection methodology, generic approaches continue to be ill-suited for such high-dimensional problems. Here we develop specialized solutions for CHMMs with weak inter-chain dependencies. Specifically we construct effective proposal distributions for the hidden state process that remain computationally viable as the number of chains increase, and that require little user input or tuning. This methodology is particularly applicable to individual-level infectious disease models characterized as CHMMs, in which each chain represents an individual, and the coupling represents contact between individuals. When this contact is non-restrictive, the resulting CHMMs naturally have low inter-chain dependencies. We demonstrate the utility of our methodology with an application to a study of highly pathogenic avian influenza in chickens.

Keywords: Model Selection; Bayes Factors; Infectious Diseases; Hidden Markov Models.