

10-th International workshop on  
Bayesian inference in stochastic processes

*Detailed Program*

June 13-15, 2017  
Bocconi University, Milan



# Program

## TUESDAY, JUNE 13

### Invited Session

<b>Bayesian inference and prediction in dynamic systems</b> . . . . .	1
Incorporating unobserved heterogeneity in Weibull survival models: A Bayesian approach (MARK F.J. STEEL, <i>Catalina A. Vallejos</i> ) . . . . .	1
Bayesian spectral modeling for multiple time series (ANNALISA CADONNA) . . . . .	1
Generalising rate heterogeneity across sites in statistical phylogenetics (SARAH HEAPS) . . . . .	2

### Contributed Session

<b>Stochastic processes and applications</b> . . . . .	2
Bayesian nonparametric analysis of Kingman’s coalescent (STEFANO FAVARO, <i>Shui Feng, Paul A. Jenkins</i> ) . . . . .	2
Hierarchical hidden Markov models for response time data (MARIO PERUGGIA, <i>Zhifei Yan, Peter Craigmile, Trisha Van Zandt</i> ) . . . . .	3
Monitoring for anomalous behaviour in massive traffic time series ( <i>David Rios Insua, ROI NAVEIRO FLORES, Simon Rodríguez Santana</i> ) . . . . .	3
Timing foreign exchange markets ( <i>Samuel W. Malone, Robert B. Gramacy, ENRIQUE TER HORST</i> ) . . . . .	4
Failure processes driven by a selfcorrecting model. Application to earthquake sequences (ELISA VARINI, <i>Renata Rotondi</i> ) . . . . .	4
A Nonparametric estimator of covariance function for parameterized family of locally self-similar processes (YASAMAN MALEKI) . . . . .	4

### Invited Session

<b>Methods and computations for complex data structures</b> . . . . .	5
Smoothing, clustering, and benchmarking for small area estimation: an application to household rental prices in Berlin (REBECCA STEORTS) . . . . .	5
Machine Learning for Healthcare Data (KATHERINE HELLER) . . . . .	5
Analysis of the Gibbs Sampler for hierarchical and crossed-effect Gaussian models via multi-grid decomposition (GIACOMO ZANELLA, <i>Omiros Papaspiliopoulos, Gareth Roberts</i> ) . . . . .	5

### Poster Session

Bayesian semiparametric change point analysis and short-term forecasting of time series (ANDREA ARFÈ, <i>Stefano Peluso, Pietro Muliere</i> ) . . . . .	6
Bayesian Nonparametric Poisson factorization with completely random measures (FADHEL AYED and <i>François Caron</i> ) . . . . .	6
A characterization of product-form exchangeable feature probability functions (MARCO BATTISTON, <i>Stefano Favaro, Daniel M. Roy, Yee Whye Teh</i> ) . . . . .	6
Determinantal point process mixtures with dependence on covariates (ILARIA BIANCHINI, <i>Alessandra Guglielmi, Fernando A. Quintana</i> ) . . . . .	7
Particle methods for inference on random graph bridges (BENJAMIN BLOEM-REDDY, <i>Peter Orbanz</i> ) . . . . .	7
Filtering of time-homogeneous pure jump Markov processes with noise-free observation and applications. (ALESSANDRO CALVIA) . . . . .	7
On some properties of “Bayesian” recursive procedures (LORENZO CAPPELLO, <i>Sandra Fortini, Sonia Petrone</i> ) . . . . .	8
Bayesian nonparametric testing of Hubbell’s neutral hypothesis: expected heterozygosity decomposition under mainland-island community structure (ANNALISA CERQUETTI) . . . . .	8
Monitoring influenza epidemics from severe case data (ALICE CORBELLA, <i>X-S. Zhang, P. J. Birrell, N. Boddington, R. Pebody, A. M. Presanis, D. De Angelis</i> ) . . . . .	9
Probabilistic preference learning with the Mallows rank model ( <i>Valeria Vitelli, Oystein Sorensen, MARTA CRISPINO, Arnaldo Frigessi, Elja Arjas</i> ) . . . . .	10

Non-exchangeable random partition model for microclustering (GIUSEPPE DI BENEDETTO, <i>François Caron, Yee Whye Teh</i> ) . . . . .	10
Well tempered Hamiltonian Monte Carlo (RITABRATA DUTTA, <i>Antonietta Mira</i> ) . . . . .	11
Approximate Bayesian Computation for Dynamic Queueing Networks (ANTHONY EBERT, <i>Paul Wu, Kerrie Mengersen, Fabrizio Ruggeri</i> ) . . . . .	11
Mixture Data-Dependent Priors for Standard Models (LEONARDO EGIDI, <i>Francesco Pauli, Nicola Torelli</i> ) . . . . .	11
The patient-zero problem: a comparison between the Monte Carlo method and Belief Propagation (AUGUSTO FASANO, <i>Alfredo Braunstein</i> ) . . . . .	12
Gaussian process regression with generalized extreme value probabilistic model (MARCELO HARTMANN) . . . . .	13
Bayesian Matrix Regression ( <i>Monica Billio, Roberto Casarin, MATTEO IACOPINI</i> ) . . . . .	13
Prediction for Stochastic Growth Processes in Fatigue Experiments (KATJA ICKSTADT, <i>Simone Hermann, Christine Mueller</i> ) . . . . .	14
Bayesian nonparametric modelling of trended monotonic functional time series (GUILLAUME KON KAM KING, <i>Matteo Ruggiero, Antonio Canale</i> ) . . . . .	14
Kinetic energy choice in Hamiltonian/hybrid Monte Carlo (SAMUEL LIVINGSTONE, <i>Michael Faulkner, Gareth Roberts</i> ) . . . . .	15
Markov Chain Monte Carlo calibration of spot-price models in Electricity markets ( <i>Alice Guerini, ANDREA MARZIALI, Giuseppe De Nicolao</i> ) . . . . .	15
Bayesian Inference for directional data through homogeneous scoring rules ( <i>V. Mameli, MONICA MUSIO, E. Ruli, L. Ventura</i> ) . . . . .	16
Inference for intractable-likelihoods using a likelihood-free version of the stochastic approximation EM algorithm (UMBERTO PICCHINI) . . . . .	16
Using Higher Order Approximations for Transition Densities in Bayesian Inference for Diffusions (SUSANNE PIESCHNER, <i>Christiane Fuchs</i> ) . . . . .	17
A rare event approach to high dimensional ABC (DENNIS PRANGLE, <i>Richard G. Everitt, Theodore Kypraios</i> ) . . . . .	17
Hierarchical Bayesian estimation of the dependence in a compound Poisson risk process (GIOVANNI RABITTI) . . . . .	18
Bayesian semiparametric modelling of contraceptive behavior in India via sequential logistic regressions (TOMMASO RIGON, <i>Daniele Durante, Nicola Torelli</i> ) . . . . .	18
A Bayesian hierarchical model for group fMRI and fMRI meta-analysis neuroimaging data (PANTELIS SAMARTSIDIS, <i>Thomas E. Nichols, Timothy D. Johnson</i> ) . . . . .	18
Statistical inference for multivariate partially observed stochastic processes with application to neuroscience ( <i>Adeline Leclercq-Samson, MASSIMILIANO TAMBORRINO</i> ) . . . . .	19
Bayesian inference for modeling of protein folding data (SAMUEL WIQVIST, <i>Umberto Picchini, Julie Lyng Forman</i> ) . . . . .	19

## WEDNESDAY, JUNE 14

### Invited Session

<b>Stochastic Programming</b> . . . . .	21
Optimisation simulation for stochastic programming ( <i>Tahir Ekin, NICHOLAS G. POLSON, Refik Soyer</i> ) . . . . .	21
Quantifying and reducing uncertainties on sets under Gaussian Process priors (DAVID GINSBOURGER) . . . . .	21
Probabilistic Programming, Bayesian Nonparametrics, and Inference Compilation (FRANK WOOD) . . . . .	22

### Contributed Session

<b>Methods and computations</b> . . . . .	22
Bayesian analysis of a general class of multivariate exponential family of state space models (TEVFIK AKTEKIN, <i>Nick Polson, Refik Soyer</i> ) . . . . .	22
Correlated pseudo marginal schemes for partially observed diffusion processes (ANDREW GOLIGHTLY, <i>Emma Bradley</i> ) . . . . .	23
A dual Markov model for filtering problems (FREDDY PALMA MANCILLA, <i>Ramsés H. Mena</i> ) . . . . .	23

Towards unified Bayesian estimation of diffusions (FRANK VAN DER MEULEN, <i>Moritz Schauer, Omiros Papaspiliopoulos, d Harry van Zanten</i> ) . . . . .	23
<b>Invited Session</b>	
<b>Stochastic Processes in Paleo Climate Science (invited special session sponsored by the Past Earth Network)</b> . . . . .	24
Gaussian process emulators with applications to Paleo-climate simulations (JOCHEN VOSS, <i>Dario Domingo</i> ) . . . . .	24
How fast is sea level rising? (ANDREW PARNELL) . . . . .	24
Quantifying uncertainties in paleoclimate reconstruction: A case study over the past 800 kyr (JAKE CARSON) . . . . .	25
<b>Invited Session</b>	
<b>Bayesian methods for network data</b> . . . . .	25
Interlocking directorates in Irish companies using a latent space model for bipartite networks (NIAL FRIEL, <i>Riccardo Rastelli, Jason Wyse, Adrian E. Raftery</i> ) . . . . .	25
Bayesian nonparametric modeling of dynamic networks (DANIELE DURANTE) . . . . .	26
Sampling and Estimation for (Sparse) Exchangeable Graphs (VICTOR VEITCH) . . . . .	26
 <b>THURSDAY, JUNE 15</b>	
<b>Invited Session</b>	
<b>Industrial methods</b> . . . . .	27
Bayesian approaches to analyzing competing and semi-competing risks data (ANANDA SEN) . . . . .	27
Bayesian modelling of virtual age in repairable systems (DIDEM EGEMEN, <i>Fabrizio Ruggeri, Refik Soyer</i> ) . . . . .	27
Sequential design of experiments and multi-objective optimization for the solutions of complex problems (MATTEO BORROTTI, <i>Antonio Pievatolo</i> ) . . . . .	28
<b>Invited Session</b>	
<b>Genomics and biostatistics</b> . . . . .	28
Bayesian inference for latent biologic structure with determinantal point processes (PETER MUELLER, <i>Yanxun Xu, Donatello Telesca</i> ) . . . . .	28
Bayesian response-adaptive designs for basket trials (LORENZO TRIPPA, <i>Steffen Ventz, William T. Barry, Giovanni Parmigiani</i> ) . . . . .	29
Different coalescent modeling resolutions for posterior inference of evolutionary parameters (JULIA PALACIOS, <i>Amandine Veber</i> ) . . . . .	29



# Tuesday, June 13

INVITED SESSION

## Bayesian inference and prediction in dynamic systems

### Incorporating unobserved heterogeneity in Weibull survival models: A Bayesian approach

13 Jun  
9:30 - 11:00

MARK F.J. STEEL<sup>1</sup>, Catalina A. Vallejos<sup>2,3</sup>

<sup>1</sup>University of Warwick, UK; <sup>2</sup>The Alan Turing Institute, UK; <sup>3</sup>University College London, UK

Outlying observations and other forms of unobserved heterogeneity can distort inference for survival datasets. The family of Rate Mixtures of Weibull distributions includes subject-level frailty terms as a solution to this issue. With a parametric mixing distribution assigned to the frailties, this family generates flexible hazard functions. Covariates are introduced via an Accelerated Failure Time specification for which the interpretation of the regression coefficients does not depend on the choice of mixing distribution. A weakly informative prior is proposed by combining the structure of the Jeffreys prior with a proper prior on some model parameters. This improper prior is shown to lead to a proper posterior distribution under easily satisfied conditions. By eliciting the proper component of the prior through the coefficient of variation of the survival times, prior information is matched for different mixing distributions. Posterior inference on subject-level frailty terms is exploited as a tool for outlier detection. Finally, the proposed methodology is illustrated using two real datasets, one concerning bone marrow transplants and another on cerebral palsy.

---

### Bayesian spectral modeling for multiple time series

13 Jun  
9:30 - 11:00

ANNALISA CADONNA<sup>1</sup>

<sup>1</sup>University of California at Santa Cruz, USA

The problem of modeling multiple time series in the spectral domain arises naturally in fields where information about frequency behavior is relevant and several signals are recorded concurrently. For example, multichannel electroencephalography (EEG) records measurements of electrical potential fluctuations at multiple locations on the scalp of a subject. I will present a hierarchical Bayesian modeling approach to spectral density estimation for multiple time series, where the log-periodogram of each series is modeled as a mixture of Gaussian distributions with frequency-dependent weights and mean functions. The implied model for each log-spectral density is a mixture of mean functions with frequency-dependent weights. In addition to accommodating flexible spectral density shapes, a practically important feature of the proposed formulation is that it allows for ready posterior simulation through a Gibbs sampler with closed form full conditional distributions for all model parameters. I will show results for multichannel electroencephalographic recordings, which provide the key motivating application for the proposed methodology. I will then present some extensions for non-stationary time series.

13 Jun  
9:30 - 11:00

## Generalising rate heterogeneity across sites in statistical phylogenetics

SARAH HEAPS<sup>1</sup>

<sup>1</sup>Newcastle University, UK

In phylogenetics, alignments of molecular sequence data for a collection of species are used to learn about their phylogeny – an evolutionary tree which places these species as leaves and ancestors as internal nodes. Sequence evolution on each branch of the tree is generally modelled using a continuous time Markov process, characterised by an instantaneous rate matrix. Early models assumed that the same rate matrix governed substitutions at all sites of the alignment, ignoring the variation in evolutionary constraints. Substantial improvements in phylogenetic inference and model fit were achieved by augmenting these models with a set of multiplicative random effects that allowed different sites to evolve at different rates which scaled the baseline rate matrix. Motivated by this pioneering work, we consider an extension which allows quadratic, rather than linear, site-specific transformations of the baseline rate matrix.

We present properties of the resulting process and show that when combined with a particular class of non-stationary models, we obtain one that allows sequence composition to vary across sites of the alignment, as well as across taxa. Formulating the model in a Bayesian framework, a Markov chain Monte Carlo algorithm is used to explore the posterior distribution. We consider two applications to alignments concerning the tree of life, fitting both stationary and non-stationary models. In each case we compare inferences obtained under our site-specific quadratic transformation, with those under linear and site-homogeneous models.

---

CONTRIBUTED SESSION

### Stochastic processes and applications

13 Jun  
11:30 - 13:30

## Bayesian nonparametric analysis of Kingman's coalescent

STEFANO FAVARO<sup>1,2</sup>, Shui Feng<sup>3</sup>, Paul A. Jenkins<sup>4</sup>

<sup>1</sup>University of Torino, Italy; <sup>2</sup>Collegio Carlo Alberto, Torino, Italy; <sup>3</sup>McMaster University, Ontario, Canada; <sup>4</sup>University of Warwick, UK

Kingman's coalescent is one of the most popular models in population genetics. It describes the genealogy of a population whose genetic composition evolves in time according to the Wright-Fisher model, or suitable approximations of it belonging to the broad class of Fleming-Viot processes. Ancestral inference under Kingman's coalescent has had much attention in the literature, both in practical data analysis, and from a theoretical and methodological point of view. Given a sample of individuals taken from the population at stationarity, most contributions have aimed at making frequentist or Bayesian parametric inference on quantities related to the genealogy of the observed sample. In this talk we propose a Bayesian nonparametric predictive approach to ancestral inference. That is, under the prior assumption that the composition of the population evolves in time according to a neutral Fleming-Viot process, and given the information contained in an initial sample of  $m$  individuals taken from the population at stationarity, we estimate quantities related to the genealogy of an additional unobservable sample of size  $m' \geq 1$ . For instance, how many non-mutant lineages would I expect a time  $t$  ago if I enlarged my initial sample by  $m' \geq 1$ ? How many of these non-mutant lineages are associated to the rare genetic types in the initial samples? In the context of ancestral inference, these questions are of great interest because they relate directly to the speed of evolution via the rate of turnover of alleles. As a by-product of our analysis we introduce a class of Bayesian nonparametric estimators which can be thought of as Good-Turing type estimators for ancestral inference. The proposed approach is illustrated through an application to genetic data.



## Hierarchical hidden Markov models for response time data

13 Jun  
11:30 - 13:30

MARIO PERUGGIA<sup>1</sup>, Zhifei Yan<sup>1</sup>, Peter Craigmile<sup>1</sup>, Trisha Van Zandt<sup>1</sup>

<sup>1</sup>The Ohio State University, USA

Psychological data, particularly those measurements obtained sequentially in experiments designed to test theories and models of human cognition, are often treated as independent and identically distributed samples from a single distribution that describes the random behavior of the cognitive process of interest. This assumption is made for mathematical and analytic convenience; it is widely appreciated that such data are in fact mixtures from two or more processes, a subset of which are associated with the cognitive process of interest. There is a long history of trying to determine the components of psychological data mixtures and estimate the relative contributions of each (see, e.g., Luce, 1986 for a review), and the mixture construct has been an important tool in investigating memory phenomena such as spreading activation and response preparation (e.g., Meyer et al., 1985; Yantis & Meyer, 1988). Our own work (Kim et al., 2017) has demonstrated the importance of including components to describe fast (subcognitive) and slow (supracognitive) processes that contribute to the measurements derived from the cognitive process of interest. In this project, we build on classic studies that attempt to distinguish the separate components of a psychological mixture process (Falmagne, 1965, 1968; Ollman, 1966; Yellott, 1971). Our modeling framework for response time data makes use of a hierarchical hidden Markov structure. The hidden states of the model are intended to capture the three putative processes (subcognitive, cognitive, and supracognitive) and to describe possibly varying levels of attention within a process. Appropriate parameter specifications allow the processes to evolve over time. The fit of the model is demonstrated on experimental data.

---

## Monitoring for anomalous behaviour in massive traffic time series

13 Jun  
11:30 - 13:30

David Rios Insua<sup>1</sup>, ROI NAVEIRO FLORES<sup>1</sup>, Simon Rodríguez Santana<sup>1</sup>

<sup>1</sup>ICMAT-CSIC, Madrid, Spain

We describe a system to monitor large amounts of Internet traffic time series so as to forecast anomalous behaviour for safety and security systems. The system essentially obtains performance readings from numerous IP connected devices in a nonintrusive manner and displays graphically the readings per device. A forecasting module is incorporated to issue:

- Short-term forecasts to detect whether critical values will be likely reached in the near future or detect anomalous behaviour.
- Long-term forecasts to detect when critical values would be likely reached.

The system covers discrete and continuous performance measures. In the continuous case, the generic model considered is a trend + seasonal DLM combined with an outburst process. In the discrete case, the generic model considered is a non-homogeneous Markov chain.

We describe various modeling issues and outline implementation details to cope with large amounts of time series.

---

13 Jun  
11:30 - 13:30

## Timing foreign exchange markets

Samuel W. Malone<sup>1</sup>, Robert B. Gramacy<sup>2</sup>, ENRIQUE TER HORST<sup>3</sup>

<sup>1</sup>University of the Andes School of Management, Colombia; <sup>2</sup>Chicago Booth School of Business, USA;

<sup>3</sup>Colegio de Estudios Superiores de Administración, Colombia and Instituto de Estudios Superiores de Administración, Venezuela

Priced carry, dollar, and volatility risk factors recently proposed in the finance literature help explain long-standing puzzles related to the cross-section of carry trade returns. We show that conditional linear and Bayesian treed Gaussian process (BTGP) models with perfect foresight of these factors substantially outperform the random walk with respect to accuracy and market timing statistics out-of-sample at one-month horizons. Simple directional trading strategies based on conditional BTGP forecasts achieve superior average Sharpe ratios and profitability compared to those based on random walk forecasts in simulated real-time trading exercises for individual currencies.

---

13 Jun  
11:30 - 13:30

## Failure processes driven by a selfcorrecting model. Application to earthquake sequences

ELISA VARINI<sup>1</sup>, Renata Rotondi<sup>1</sup>

<sup>1</sup>CNR - IMATI, Milano, Italy

Earthquake occurrence is usually modelled by point processes. The longterm recurrence of strong earthquakes is often modelled by Poisson process or selfcorrecting point processes with nondecreasing hazard. In a shortterm time scale, selfexciting point processes with nonincreasing hazard are especially suitable to describe the general tendency to occur in clusters.

In order to provide a unified framework for analyzing earthquake catalogs, we consider a seismic sequence as the union of two disjoint subsets of events, hereinafter named the *leaders* and the *subordinates*. Based on the empirical evidence that earthquakes are typically clustered in time, leaders correspond to main events with magnitude exceeding a threshold magnitude which characterizes destructive quakes in Italy. The remaining events are labelled as subordinates.

We introduce a new point process aimed to jointly model the evolution over time of these two components of the earthquake process. The leaders are assumed to be generated according to a selfcorrecting point process and, conditionally to the occurrence of the leaders, the subordinates follow a failure process which admits a bathtub hazard function. Specifically, we consider the generalized Weibull distributions, a wide family of distributions having hazard functions that admit different shapes (e.g. increasing, decreasing, bathtubshaped, upsidedown bathtubshaped).

We analyze some Italian sequences of earthquakes drawn from the new ParametricCatalogue of Italian Earthquakes (CPTI15).

---

13 Jun  
11:30 - 13:30

## A Nonparametric estimator of covariance function for parameterized family of locally self-similar processes

YASAMAN MALEKI<sup>1</sup>

<sup>1</sup>Alzahra University, Iran

The optimal covariance function estimate, in the sense of mean-square of errors (MSE), for the class of discrete-time locally self-similar processes is accomplished from one observed realization by weighting observations with a kernel function. This paper investigates the MSE optimal kernel function for any parameterized family of locally self-similar processes by solving a system of linear equations. Furthermore, it is shown that the optimal kernel is close to optimal for all members of the family.

---

## Methods and computations for complex data structures

### Smoothing, clustering, and benchmarking for small area estimation: an application to household rental prices in Berlin

13 Jun  
15:00 - 16:30

REBECCA STEORTS<sup>1</sup>

<sup>1</sup>Duke University, USA

We develop constrained Bayesian estimation methods for small area problems: those requiring smoothness with respect to similarity across areas, such as geographic proximity or clustering by covariates; and benchmarking constraints, requiring (weighted) means of estimates to agree across levels of aggregation. We develop methods for constrained estimation decision-theoretically and discuss their geometric interpretation. Our constrained estimators are the solutions to tractable optimization problems and have closed-form solutions. Mean squared errors of the constrained estimators are calculated via bootstrapping. Our techniques are free of distributional assumptions and apply whether the estimator is linear or non-linear, univariate or multivariate. We illustrate our methods to the estimation of house rental prices from the Empirica-Systeme GmbH database in Berlin.

### Machine Learning for Healthcare Data

13 Jun  
15:00 - 16:30

KATHERINE HELLER<sup>1</sup>

<sup>1</sup>Duke University, USA

We will present the healthcare settings we have been working on at Duke University, and the machine learning methods which we have recently developed to improve patient outcomes. These include: 1) Modeling disease trends and other predictions from electronic health record (EHR) data, including for sepsis and chronic kidney disease. The use of a Gaussian process framework for modeling time series health data is explored. 2) Mobile apps and devices for improving the granularity of recorded health data and 3) The combination of mobile app and social network information, and the development of Graph-coupled hidden Markov models for predicting the spread of influenza in student dormitories.

### Analysis of the Gibbs Sampler for hierarchical and crossed-effect Gaussian models via multigrid decomposition

13 Jun  
15:00 - 16:30

GIACOMO ZANELLA<sup>1</sup>, Omiros Papaspiliopoulos<sup>2</sup>, Gareth Roberts<sup>3</sup>

<sup>1</sup>Bocconi University, Italy; <sup>2</sup>Universitat Pompeu Fabra, Spain; <sup>3</sup>University of Warwick, UK

We study the convergence properties of the Gibbs Sampler in the context of Gaussian hierarchical and crossed-effect models. We develop a novel methodology based on multi-grid decompositions to derive analytic expressions for the convergence rates of the algorithm, extending significantly the class of conditionally Gaussian models amenable to direct analysis. In the hierarchical context, our work gives a rather complete understanding of the Gibbs Sampler behavior for symmetric models (with arbitrary depth), while providing approximations and bounds for the non-symmetric cases. The theoretical results give rise to simple and easy-to-implement guidelines to optimize practical implementations of the Gibbs samplers on such models. While the good performances of the Gibbs Sampler in hierarchically-structured models is renowned, the context of crossed-effect models is drastically different. Here hierarchical centering is not possible and the convergence of commonly implemented Gibbs Sampler strategies deteriorates as the data-size increases, resulting in super-linear computational complexity (potentially even quadratic) in the number of data-points. We show how to leverage the negatively-correlated structure of crossed-effect models to design easy-to-implement collapsed Gibbs Samplers whose complexity matches the one of hierarchical scenarios.

## Poster Session

### Bayesian semiparametric change point analysis and short-term forecasting of time series

13 Jun  
17:00 - 19:00

ANDREA ARFÈ<sup>1</sup>, Stefano Peluso<sup>2</sup>, Pietro Muliere<sup>1</sup>  
<sup>1</sup>Bocconi University, Italy; <sup>2</sup>Cattolica University, Italy

We illustrate a novel Bayesian semiparametric model for the analysis of time series with an unknown number of structural change points. Times of structural changes and regimes follow the multi-state beta-Stacy process prior. Accordingly, the predictive distribution of the structural change points is characterized by means of a system of urns with reinforcements. Conditionally on change points and regimes, structural parameters and observations within regimes follow standard state-space models. Inference and short-term forecasting are performed by means of an ad-hoc Markov Chain Monte Carlo algorithm based on Kalman filtering and backward smoothing. As an application, we analyse time-series data on cancer incidence from the Surveillance Epidemiology and Endpoints Research (SEER) program of the United States National Cancer Institute.

### Bayesian Nonparametric Poisson factorization with completely random measures

13 Jun  
17:00 - 19:00

FADHEL AYED<sup>1</sup> and François Caron<sup>1</sup>  
<sup>1</sup>University of Oxford, UK

Recommender systems aim at predicting which items a user may like based on a history of ratings or purchases. In this work, we study Bayesian Nonparametric Poisson factor models for recommendation systems. We assume that each user/item has an infinite set of latent features, which are modeled using a Completely Random Measure (CRM). The model is flexible and allows the number of active features to grow unboundedly with the number of user/items. We show how the properties of the CRM relate to the growth of the number of active features. We also derive a Markov chain Monte Carlo algorithm for posterior inference for this class of models.

### A characterization of product-form exchangeable feature probability functions

13 Jun  
17:00 - 19:00

MARCO BATTISTON<sup>1</sup>, Stefano Favaro<sup>2,3</sup>, Daniel M. Roy<sup>4</sup>, Yee Whye Teh<sup>5</sup>  
<sup>1</sup>Bocconi University, Italy; <sup>2</sup>University of Torino, Italy; <sup>3</sup>Collegio Carlo Alberto, Italy; <sup>4</sup>University of Toronto, Canada; <sup>5</sup>University of Oxford, UK

We characterize the class of exchangeable feature allocations assigning probability  $V_{n,k} \prod_{l=1}^k W_{m_l} U_{n-m_l}$  to a feature allocation of  $n$  individuals, displaying  $k$  features with counts  $(m_1, \dots, m_k)$  for these features. Each element of this class is parametrized by a countable matrix  $V$  and two sequences  $U$  and  $W$  of non-negative weights. Moreover, a consistency condition is imposed to guarantee that the distribution for feature allocations of  $n - 1$  individuals is recovered from that of  $n$  individuals, when the last individual is integrated out. In Theorem 1.1, we prove that the only members of this class satisfying the consistency condition are mixtures of the Indian Buffet Process over its mass parameter  $\gamma$  and mixtures of the Beta-Bernoulli model over its dimensionality parameter  $N$ . Hence, we provide a characterization of these two models as the only, up to randomization of the parameters, consistent exchangeable feature allocations having the required product form.

## Determinantal point process mixtures with dependence on covariates

13 Jun  
17:00 - 19:00

ILARIA BIANCHINI<sup>1</sup>, Alessandra Guglielmi<sup>1</sup>, Fernando A. Quintana<sup>2</sup>

<sup>1</sup>Politecnico di Milano, Italy; <sup>2</sup>Pontificia Universidad Católica de Chile

Bayesian mixture models, both nonparametric or parametric with a prior on the number of components, are frequently employed when dealing with density estimation and clustering. However, these models often leads to an overestimation of the number of clusters: the latent parameters associated to each group may be very similar due to the a-priori independence assumption, thus generating redundant mixture components and worthless complex models.

For this reason, we consider mixture models where latent parameters are a priori encouraged to separate from each other (repulsion), thus inducing well separated clusters. In particular, we induce repulsion using stationary Determinantal Point Processes (DPPs) that are defined via spectral density approach.

In addition, we consider a covariate-dependent clustering assignment induced by the model, so that data points with a similar covariate are a-priori more likely to belong to the same group. Posterior inference is carried out using MCMC and reversible jumps methods. We illustrate the goodness of the model on synthetic and real data.

---

## Particle methods for inference on random graph bridges

13 Jun  
17:00 - 19:00

BENJAMIN BLOEM-REDDY<sup>1,2</sup>, Peter Orbanz<sup>1</sup>

<sup>1</sup>Columbia University, U.S.A.; <sup>2</sup>University of Oxford, UK

Consider a model for a sequence of growing random graphs, and the problem of performing inference when observing a single graph. The model defines a stochastic process on the space of graphs; assuming a seed graph (for simplicity, a single edge) and conditioning on the observed graph, the problem becomes one of performing inference on a stochastic bridge. Building upon techniques designed for stochastic bridges in continuous spaces, we develop a sequential Monte Carlo algorithm that approximates the bridge distribution. We also propose an unbiased estimator of the marginal bridge likelihood, which enables parameter inference using particle Markov Chain Monte Carlo methods. The resulting inference algorithms can be used for any model satisfying a Markov property and a monotonicity property, such as preferential attachment or vertex copying models. To demonstrate and to highlight practical challenges, we devise a particle Gibbs sampler to fit a preferential attachment model, and a model based on random walks.

---

## Filtering of time-homogeneous pure jump Markov processes with noise-free observation and applications.

13 Jun  
17:00 - 19:00

ALESSANDRO CALVIA<sup>1</sup>

<sup>1</sup>Università degli Studi di Milano-Bicocca, Italy

In this talk I am going to address a particular stochastic filtering problem, mainly characterized by observations not corrupted by noise. Let  $X$  and  $Y$  be a given couple of stochastic processes, with values in two Borel spaces  $(I, \mathcal{I})$  and  $(O, \mathcal{O})$  respectively. The unobserved (or signal) process  $X = (X_t)_{t \geq 0}$  is a time-homogeneous pure jump Markov process, whose initial distribution and rate transition measure are known. The observed process  $Y$  is defined as  $Y_t = h(X_t)$ ,  $t \geq 0$ , where  $h : I \rightarrow O$  is a known surjective and measurable function. This model can be viewed as an instance of a Hidden Markov Model and only sporadic and partial results are available in the literature for this specific setting.

The first aim is to provide an explicit equation for the filtering process  $\pi = (\pi_t)_{t \geq 0}$ , satisfying  $\pi_t(\varphi) = \mathbb{E}[\varphi(X_t) | \mathcal{Y}_t]$ , where  $\varphi : I \rightarrow \mathbb{R}$  is a bounded measurable function and  $\mathcal{Y}_t = \sigma(Y_s : 0 \leq s \leq t)$  denotes the natural filtration of  $Y$ . The problem is tackled with the aid of known results from marked point processes theory and a martingale approach.

The second goal is to characterize the filtering process  $\pi$  as a *Piecewise Deterministic Markov Process*, in the sense of Davis. This provides an explicit description of the dynamics of the filtering process and allows to use known results on this class of stochastic processes in applications. If time permits, I will briefly discuss one of these applications: an optimal control problem for a continuous-time Markov chain with noise-free partial observation.

---

13 Jun  
17:00 - 19:00

## On some properties of “Bayesian” recursive procedures

LORENZO CAPPELLO<sup>1</sup>, Sandra Fortini<sup>1</sup>, Sonia Petrone<sup>1</sup>

<sup>1</sup>Bocconi University, Milan, Italy

Some recursive procedures have been suggested in the literature where the update of a Bayesian nonparametric model is the building block of the algorithm (Newton et al. 1998, Hahn et al. 2017). These algorithms provide a nonparametric estimate of an unknown distribution. Asymptotic properties have been studied under the assumption of observing i.i.d. data. How these algorithms relate to the Bayesian models that have inspired them, has not been well understood yet. In this work, we explore two different directions in which the connection might be explored. First, we study asymptotic properties under the assumption of observing exchangeable data. This result is of interest because it might suggest that some of these procedures could be used to approximate Bayesian quantities not available in closed form, such as posterior predictive distributions. Alternatively, if the recursive algorithm is taken as the data-generating mechanism, the sequence of observation can be shown for specific algorithms to be conditional identically distributed, hence asymptotically exchangeable (Berti et al. 2004).

---

13 Jun  
17:00 - 19:00

## Bayesian nonparametric testing of Hubbell’s neutral hypothesis: expected heterozygosity decomposition under mainland-island community structure

ANNALISA CERQUETTI

The neutral hypothesis originates in mathematical population genetics (Kimura, 1968) as opposite to selective Darwinian theory. The theory asserts that a large fraction of observed genetic variation is nonselective but occurs purely by chance. In a community ecology transposition abundances within a neutral population fluctuate and diversity arises as a balance between the immigration of new species and local extinction. In 2001 Hubbell extends this theory to multiple sites populations using a mainland-island community structure: local communities governed by neutral dynamics are connected through migration to a neutral metacommunity where diversity is generated through speciation. Interest in testing Hubbell’s neutral hypothesis is huge, being suspected to play a role in a wide set of interacting living systems, spanning from tropical rain forests to human gut microbiome. The current approach is by a Monte Carlo significance test based on Etienne’s (2007) sampling formula. Just recently it has been shown (Harris et al. 2015) that a large class of neutral models with mainland-island structure converges, in the large population limit, to the Hierarchical Dirichlet process (Teh et al. 2006). Moving from these findings, in a Bayesian nonparametric perspective, we investigate an alternative approach to neutrality testing based on Jost (2007) diversity differentiation. Relying on properties of structural distributions of infinite random discrete distributions, we focus on prior and posterior expectations of alpha and beta heterozygosities under Hierarchical Dirichlet priors and study their possible role both in an hybrid Bayesian-frequentist approach and in the construction of a proper Bayesian solution.

---

Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature*, 217, 624-62.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, NJ.

Etienne, R. (2007) A neutral sampling formula for multiple samples and an exact test of neutrality. *Ecol Lett* 10, 608-618

Harris et al. (2015) Linking statistical and ecological theory: Hubbell’s unified neutral theory of biodiversity as a Hierarchical Dirichlet process. *Proceedings of the IEEE*, 99.

Teh, YW et al. (2006) Hierarchical Dirichlet processes. *JASA* 101, 1566-1581

Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, 88, 2427-2439

# Monitoring influenza epidemics from severe case data

13 Jun  
17:00 - 19:00

ALICE CORBELLA<sup>1</sup>, X-S. Zhang<sup>2</sup>, P. J. Birrell<sup>1</sup>, N. Boddington<sup>2</sup>, R. Pebody<sup>2</sup>, A. M. Presanis<sup>1</sup>, D. De Angelis<sup>1,2</sup>

<sup>1</sup>Medical Research Council, Biostatistics Unit, University of Cambridge, School of Clinical Medicine, UK;  
<sup>2</sup>Public Health England, UK

**Background:** Seasonal and pandemic influenza remains a significant burden on health systems (Pitman et al., 2007). Knowledge of the size and timing of an influenza outbreak, together with the probability of experiencing severe events, are crucial for planning vaccination and other healthcare services, mitigation strategies or, more generally, to inform health policy decisions. The information on infected cases is often sparse, not representative of the population and difficult to access. Our work aims at inferring the parameters governing the underlying transmission dynamics and the outbreak severity from very few, readily available, information.

**Data:** The UK Severe Influenza Surveillance System (USISS) is a mandatory scheme through which Public Health England (PHE) collects information on the weekly count of Intensive Care Unit (ICU) admissions with confirmed influenza. These data are collected over each influenza season and are readily available. Other data of different type may also be available: these include, for example, individual data on the time from hospitalization to ICU admission and to death or discharge, count data on the deaths in all ICUs and count data on hospitalizations, ICU admissions and deaths in a subset of hospitals.

**Methods:** Epidemic models are a specific case of Markov processes where the transition rates depend on the compartment sizes (Keeling and Rohani, 2008). The process of severity can also be viewed from the perspective of a partially observed Markov model. We build a model to describe the spread of the disease in the population and the progression of the individual over various severity states (e.g. from hospitalization to ICU admission and to death or discharge). Within a Bayesian framework, we merged information from external data sources (e.g. individual-level data on a sub-sample of ICU admissions and knowledge on the distribution of some parameters) with the count of weekly ICU admissions reported in USISS. We obtained the posterior distribution of transmission and severity parameters. Moreover we simulated the inference as the epidemic is starting, assessing the potential for prediction.

**Results:** High correlation among a subgroup of the parameters is diagnosed, highlighting properties of the system considered, such as the dependencies of the parameters on the initial states of the system. Plausible estimates are obtained for the effective reproduction number, the probability of ICU admission given infection and the attack rate. Analysis of the forecasting power performs well when data up to the peak of the epidemic are available.

**Discussion:** Routinely collected data on severe cases carry powerful information to infer and forecast influenza epidemics in terms of both transmission and severity. Bayesian methods grant a framework for different sources and kinds of data to be combined. These methods allow both the estimation of the parameters and the identification of sub-processes where information is lacking.

---

R. J. Pitman, A. Melegaro, D. Gelb, M. R. Siddiqui, N. J. Gay, and W. J. Edmunds. “Assessing the burden of influenza and other respiratory infections in England and Wales”. In: *J. Infect.* 54.6 (2007), pp. 530-538.

HPA. *UK Severe Influenza Surveillance System (USISS) Protocol for all NHS Acute Trusts 2011-12*. 2011. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/332482/USISS\\_2011\\_2012\\_Protocol\\_Mandatory.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/332482/USISS_2011_2012_Protocol_Mandatory.pdf).

M. J. Keeling and P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.

13 Jun  
17:00 - 19:00

## Probabilistic preference learning with the Mallows rank model

Valeria Vitelli<sup>1</sup>, Oystein Sorensen<sup>1</sup>, MARTA CRISPINO<sup>2</sup>, Arnaldo Frigessi<sup>1,3</sup>, Elja Arjas<sup>1</sup>

<sup>1</sup>Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Norway;

<sup>2</sup>Department of Decision Sciences, Bocconi University, Milan, Italy; <sup>3</sup>Oslo University Hospital, Norway

Ranking and comparing items is crucial for collecting information about preferences in many areas, including marketing and politics.

The Mallows rank model is among the most successful approaches to analyse rank data, but its computational complexity has limited its use to a particular form based on Kendall distance.

We develop computationally tractable methods for Bayesian inference in Mallows models with any right-invariant distance function.

Our method performs Bayesian inference on the consensus ranking of the items, also when based just on partial rankings, such as top-k items or pairwise comparisons.

In these cases, we also perform inference on the individual latent rankings of each assessor, which are of enormous interest, for example when performing personalized recommendations, or to study how individual preferences change with user related covariates.

We prove that items which none of the assessors has ranked, do not influence the maximum a posteriori consensus ranking, and can therefore be ignored. When assessors are many or heterogeneous, we propose a mixture model for clustering them in homogeneous subgroups, with cluster-specific consensus rankings, without any preprocessing.

We develop approximate stochastic algorithms that allow a fully probabilistic analysis, leading to coherent quantifications of uncertainties. We make probabilistic predictions on the class membership of the assessors, based on their ranking of just some items, and predict missing individual preferences, as needed in recommendation systems.

We provide extensive experiments on simulated values, and compare our results with some of the existing approaches. Finally, we test our method using several experimental and benchmark datasets.

---

13 Jun  
17:00 - 19:00

## Non-exchangeable random partition model for microclustering

GIUSEPPE DI BENEDETTO<sup>1</sup>, François Caron<sup>1</sup>, Yee Whye Teh<sup>1</sup>

<sup>1</sup>University of Oxford, UK

The task of clustering data is one of the most popular applications of Bayesian Nonparametrics. This problem can be addressed by specifying a prior distribution on the partition of the data. The most commonly used priors, such as the Chinese Restaurant Process and its two-parameters generalization, rely on some exchangeability assumption; while this assumption may be reasonable for some applications, it has strong implications on the asymptotic properties of the cluster sizes. In fact, as proved in [Kingman – The Representation of Partition Structures – 1978] and stressed in [Miller et al. – Microclustering: When the Cluster Sizes Grow Sublinearly with the Size of the Data Set – 2015], exchangeable random partitions imply the linear growth of the cluster sizes, which is not suitable for several applications. We will present a flexible non-exchangeable random partition model which is able to generate partitions whose growth of the clusters sizes is almost surely sublinear. Along with this result, we provide the asymptotic behaviour of the number of clusters of a given size. Both these rates can be tuned through a parameter and the base measure of a homogeneous completely random measure and inference on the parameters is based on a Sequential Monte Carlo algorithm.



## Well tempered Hamiltonian Monte Carlo

13 Jun  
17:00 - 19:00

RITABRATA DUTTA<sup>1</sup>, Antonietta Mira<sup>1</sup>

<sup>1</sup>InterDisciplinary Institute of Data Science, Università della Svizzera italiana, Switzerland

For performing Bayesian inference in probabilistic model of continuous variables, Hamiltonian Monte Carlo (HMC) has proved to be a widely used and very powerful Markov chain Monte Carlo method. However while applying HMC for complex models as in the case of stochastic processes, they perform poorly when the distribution has multiple isolated modes. Based on a well studied approach used for simulation of molecular dynamics, we present here a Monte Carlo scheme for which the Hamiltonian dynamics is driven by an adaptively biased potential energy function. We call this sampling scheme as well tempered Hamiltonian Monte Carlo (wt-HMC). wt-HMC helps us in exploring modes of a multi-modal target distribution and can very simply be implemented with existing HMC code. We study the asymptotic properties of wt-HMC and provide a way to sample from the true distribution using a simple importance re-weighting.

---

## Approximate Bayesian Computation for Dynamic Queueing Networks

13 Jun  
17:00 - 19:00

ANTHONY EBERT<sup>1,2</sup>, Paul Wu<sup>1,2</sup>, Kerrie Mengersen<sup>1,2</sup>, Fabrizio Ruggeri<sup>1,2,3</sup>

<sup>1</sup>Queensland University of Technology (QUT), Australia; <sup>2</sup>Australian Research Council (ARC) Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), Australia; <sup>3</sup>CNR-IMATI, Italy

Important real-world systems such as airport terminals, manufacturing processes and hospitals are modeled with networks of queues. To estimate parameters, restrictive assumptions are placed on these models. For instance arrival and service distributions are assumed to be time-invariant. Violating this assumption are so-called dynamic queueing networks which are more realistic but do not allow for likelihood-based parameter estimation.

We present a new type of queueing simulation, called queue departure computation (QDC), with computational speed-ups of more than two orders of magnitude. We propose to embed this queueing simulation within an approximate Bayesian computation (ABC) algorithm to estimate parameters for dynamic queueing networks in a straight forward manner.

---

## Mixture Data-Dependent Priors for Standard Models

13 Jun  
17:00 - 19:00

LEONARDO EGIDI<sup>1</sup>, Francesco Pauli<sup>2</sup>, Nicola Torelli<sup>2</sup>

<sup>1</sup>Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Italy; <sup>2</sup>Dipartimento di Scienze Economiche, Aziendali, Matematiche e Statistiche "Bruno de Finetti", Università degli Studi di Trieste, Italy

Priors elicitation is the core of every Bayesian analysis. In the last decades a huge amount of scientific work has been developed in order to clarify the information role of a prior distribution. Nowadays the notion of informative and weakly informative priors (Gelman, et al. (2008)) is relevant in many applications and many attempts for overcoming the noninformative choice have been proposed. Despite their computational attractiveness, improper and noninformative priors result in fact to be inadequate in model selection and model comparison, since proper priors are required to compute the Bayes factors.

Asking how much informative should be a prior distribution and how we may use data information for our elicitation process represent well posed questions. Regarding the first one, Morita, et al. (2008) proposed the so called prior effective sample size (ESS), an index which provides the amount of information contained in a given prior distribution for the parameter  $\theta$ , computed with respect to a posterior  $q(\theta|y)$  arising from a baseline prior  $\pi_b$ , with  $\pi_b$  less informative than  $\pi$ . The larger is

---

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 13601383.

Morita, S., Thall, P. F., and Müller, P. (2008). Determining the effective sample size of a parametric prior. *Biometrics*, 64 (2), 595602.

this value, and the greater is the prior’s information. Even if the model is simple, when the sample size is small it is not trivial to specify a prior which should be informative and at the same time shouldn’t dominate the inference. In fact, when fitting a Bayesian model to a dataset consisting of 10 observations, an effective sample size of 1 is reasonable, whereas a value of 20 implies that the prior, rather than the data, dominates the inference: with few data, there is the risk of being “too much informative”. Motivated by this consideration and partially by the power prior approach of Ibrahim et al. (1998) and Ibrahim and Chen (2000), our proposal moves to the second question and attaches directly the priors construction for small samples within the standard models. By adopting the idea of the posterior matching in Morita et al. (2008), we develop an algorithm which uses data for driving the priors elicitation. We consider the Hellinger distance  $H_1$  between the data at hand and those generated under the informative prior  $\pi$ , and the Hellinger distance  $H_2$  between the two posteriors resulting from choosing  $\pi$  and  $b$ . We generate  $m$  additional data until the observed value of  $H_2$  is below a fixed threshold. The corresponding observed value  $H_{1,m}$  at the  $m$ -th iteration is plugged in a two-components mixture of  $\pi_b$  and  $\pi$  and used for weighting the two prior distributions. The greater is  $H_{1,m}$  in this new mixture prior, the farther are the data (simulated and real) from  $\pi$ , and consequently the stronger is the influence of the baseline prior. We prove that the so obtained class of priors satisfies some nice properties. Among these, they always have a closed form in standard models and preserve the conjugacy; under mild conditions, they yield a lower effective sample size than that provided by  $\pi$ , i.e. they carry less information; in the extreme case of perfect dissimilarity between the data and those generated under the prior, they coincide with the general power prior of Ibrahim and Chen (2000) when the power  $\alpha_0 = 0$ .

---

## The patient-zero problem: a comparison between the Monte Carlo method and Belief Propagation

13 Jun  
17:00 - 19:00

AUGUSTO FASANO<sup>1</sup>, Alfredo Braunstein<sup>2,3,4</sup>

<sup>1</sup>Bocconi University, Milan, Italy; <sup>2</sup>Politecnico di Torino, Turin, Italy; <sup>3</sup>Collegio Carlo Alberto, Moncalieri, Italy; <sup>4</sup>Human Genetics Foundation, Turin, Italy

The patient-zero problem is an inferential challenging problem: given an observation of an epidemic spread over a graph at a certain instant of time  $T$ , the goal is to infer which node was the source of the epidemic. From a Bayesian perspective, one needs to compute which node is the one with the highest *a posteriori* probability of being the source.

This problem can be tackled with (at least) two different approaches: the first one uses repeated sampling as typical of Monte Carlo (MC) method, while the second one is based on Belief Propagation (BP) algorithm, which is a message-passing algorithm that allows us to compute marginal probability distributions, in our case the *a posteriori* probability of each node being the source of the epidemic, when we know the expression for the joint probability distribution. These marginals are correct when the factor graph corresponding to the joint distribution is a tree, otherwise they are approximations.

In particular, the epidemic spread is modeled through a stochastic Markov process: the SI or the SIR. The two algorithms were compared when the epidemic is spreading according to the SI dynamics, and a natural generalization holds for the SIR in both cases.

Given a realization, the idea behind the MC method is to take an uninformative prior distribution on the source and reduce ourselves to a likelihood maximization problem: for each of the infected nodes, the likelihood of him being the source of the epidemic is computed, and then these nodes are ordered by decreasing likelihood value. On the other hand, BP algorithm allows to compute for each infected node the *a posteriori* probability of him being the source and then the nodes are ranked accordingly.

Comparing the results, in the studied cases BP performs quite better than MC: it gives very accurate estimates when converging, while some tricks can be used to improve performances also in case it is not converging.

---

Ibrahim, J. G., Ryan, L. M., and Chen, M.-H. (1998). Using historical controls to adjust for covariates in trend tests for binary data. *Journal of the American Statistical Association*, 93 (444), 1282-1293.

Ibrahim, J. G., and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, 4660.

Among the studied cases, when BP algorithm is not converging the probability of making perfect inference is lower than in the case convergence is met, but the true source is ranked anyway at the very top of the list of possible sources ordered by decreasing probability of being the source. On the other hand, MC methods are slower and usually less precise, meaning that, even in cases in which the true source is ranked at the top of the list, the proportion of times in which the true source is exactly the first node in the list is not very high.

---

## Gaussian process regression with generalized extreme value probabilistic model

13 Jun  
17:00 - 19:00

MARCELO HARTMANN<sup>1</sup>

<sup>1</sup>Department of Mathematics and Statistics, University of Helsinki, Finland

Extreme value theory has been extensively applied in risk management and environmental statistics among other variety of fields. It also has been subject of much practical work due to its complex analytical models. In statistical modelling it is common to study the extreme value behaviour of maxima (or minima) under the presence of covariates setting a regression function to the location parameter of generalized extreme value distribution. In this work we treat the location parameter of the generalized extreme value distribution as a random function following a Gaussian process. This particular configuration leads to a posterior distribution with non-closed form expression that may be high-dimensional in practical applications and so impose time constrain for full Bayesian inference via Monte Carlo Markov chains methods. For such reason methods of analytical approximation, such as the Laplace method, can be a good and fast alternatives. The Laplace approximation with generalized extreme value model is challenging once the density function has a restricted sample space and it is not log-concave, which in turn brings instability in the Newton-Raphson algorithm when trying to find the maximum a posterior estimate (MAP). We then proposed the use of the natural gradient which uses the Fisher information matrix (the Riemannian metric) and improves the stability of the algorithm. We also compare the Laplace approximation with MCMC in order to evaluate the quality of the approximation.

---

## Bayesian Matrix Regression

13 Jun  
17:00 - 19:00

Monica Billio<sup>1</sup>, Roberto Casarin<sup>1</sup>, MATTEO IACOPINI<sup>1</sup>

<sup>1</sup>Ca' Foscari University of Venice, Italy

We propose a new linear regression model with tensor variate response and covariates, that encompasses univariate and multivariate regression models as special cases. For dealing with the over-parametrization and over-fitting issues due to the curse of dimensionality, we exploit a suitable parametrization which enables to achieve both parameter parsimony and to incorporate sparsity effects. Inference is carried out in the Bayesian framework combined with Monte Carlo Markov Chain (MCMC). We show the efficiency of the MCMC procedure on simulated data.

---

13 Jun  
17:00 - 19:00

## Prediction for Stochastic Growth Processes in Fatigue Experiments

KATJA ICKSTADT<sup>1</sup>, Simone Hermann<sup>1</sup>, Christine Mueller<sup>1</sup>

<sup>1</sup>TU Dortmund University, Germany

In constructional engineering, experiments of material fatigue are expensive and, therefore, seldom. In our research project on *Statistical methods for damage processes under cyclic load* of the Collaborative Research Centre 823 at the TU Dortmund University, the engineers conducted an experiment, in which they set several pre-stressed concrete beams under cyclic load, starting with initial cracks. We aim to find a stochastic model describing the process in order to predict the development of the crack width curve. To extract as much information as possible from existing experiments, we will, in addition, model and do prediction for the famous data set of Virkler et al. (1979), in which the number of cycles that lead to a of fixed crack lengths are observed.

In this talk we propose a general Bayesian approach for stochastic versions of deterministic growth models to provide predictions for crack propagation in an early stage of the growth process. To improve the prediction, the information of other crack growth processes is used in a hierarchical model. Two stochastic versions of a deterministic growth model are considered. One is a nonlinear regression model, in which the trajectory is assumed to be the solution of an ordinary differential equation with additive errors. The other is a diffusion model defined by a stochastic differential equation. The predictive distributions will be approximated simulating from the posterior distributions resulting from a corresponding Markov chain Monte Carlo algorithm.

---

## Bayesian nonparametric modelling of trended monotonic functional time series

GUILLAUME KON KAM KING<sup>1,2</sup>, Matteo Ruggiero<sup>1,2</sup>, Antonio Canale<sup>2,3</sup>

<sup>1</sup>Collegio Carlo Alberto, Italy; <sup>2</sup>Università Degli Studi di Torino, Italy; <sup>3</sup>University of Padua, Italy

Functional time series naturally appear in contexts where phenomena are measured regularly. Examples include the income distribution over time, the evolution of molecular size distribution during polymerisation, or daily demand/offer curves in an exchange market. Trends are common in these series: higher incomes might tend to increase while lower incomes stagnate or decrease, polymerisation increases molecule sizes globally, and prices commonly show rising or falling trends. The functional nature of the data raises a challenge for the inference and indeed, the likelihood can be intractable in the case of fully observed functions. We present a likelihood-free approach for functional data forecast with a trend phenomenon. We develop a Bayesian nonparametric model based on a dependent process. It builds on particle system models, which originate from population genetics. This construction provides a means to flexibly specify the correlation of the dependent process. We take advantage of the expressiveness of interacting particle models to embed a local and transient trend mechanism. To this aim, we draw inspiration from interaction potentials between physical particle systems in molecular dynamics.

We perform the likelihood-free inference by means of Approximate Bayesian Computation (ABC). We discuss the elicitation of informative summary statistics for stochastic processes building on the idea of semi-automatic summaries. Coupled with a population ABC, this results in a very versatile inference method. We show the increased robustness of the trended model and comment on the generality of our approach for building functional forecast models.

---

Virkler, D. A., B. M. Hillberry and P. K. Goel (1979). The Statistical Nature of Fatigue Crack Propagation. *Journal of Engineering Materials and Technology* 101, 148-153.

# Kinetic energy choice in Hamiltonian/hybrid Monte Carlo

13 Jun  
17:00 - 19:00

SAMUEL LIVINGSTONE<sup>1</sup>, Michael Faulkner<sup>1</sup>, Gareth Roberts<sup>2</sup>

<sup>1</sup>University of Bristol, UK; <sup>2</sup>University of Warwick, UK

We consider how different choices of kinetic energy in Hamiltonian Monte Carlo affect algorithm performance. We introduce two quantities which can be easily evaluated, the composite gradient and the implicit noise, as indicators of algorithm performance. Results are established on geometric convergence, and we show that choices of kinetic energy that result in heavy-tailed momentum distributions can exhibit an undesirable negligible moves property, which we define. A general efficiency-robustness trade off is outlined, and implementations which rely on approximate gradients are also discussed. Two numerical studies illustrate our theoretical findings, showing that the standard choice which results in a Gaussian momentum distribution is not always optimal in terms of either robustness or efficiency.

---

# Markov Chain Monte Carlo calibration of spot-price models in Electricity markets

13 Jun  
17:00 - 19:00

Alice Guerini<sup>1,2</sup>, ANDREA MARZIALI<sup>1,2</sup>, Giuseppe De Nicolao<sup>1</sup>

<sup>1</sup>Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli studi di Pavia, Italy;

<sup>2</sup>Modeling & Pricing, A2A S.p.A., Italy

In recent years, due to the liberalization of the energy market, the stochastic behavior of energy prices is playing an increasingly important role in financial models. In particular, there is an extensive literature dealing with electrical spot prices, where a wide variety of models have been proposed. Several authors describe spot prices by one-factor models. A limitation of these models is that the geometric Brownian motion does not take into account the mean reverting behavior, typical of energy prices. This is an issue that is not satisfactorily recovered by resorting to mean-reverting one-factor models, because their forward volatility tends asymptotically to zero. In order to overcome all these problems, the use of a two-factor mean-reverting model has been advocated. The key improvement of this model is that the mean of the system is itself a stochastic process. The deployment of these models requires a calibration step in which the model parameters are estimated based on available price data. A common approach relies on Maximum Likelihood (ML) estimation, which is relatively easy in the one-factor case, whereas efficient ML calculations for the two-factor model can be based on a Kalman Filter (KF) algorithm.

As an alternative to ML estimation, one may adopt a Bayesian paradigm and resort to Markov Chain Monte Carlo (MCMC) methods, that rely on stochastic sampling algorithms. Given their capability to provide the posterior densities of the model parameters, MCMC methods may offer some advantages in the assessment of parameters uncertainties and model robustness, compared to ML methods that rely on approximations based on the Fisher information matrix. In this work, parameter calibration of one- and two-factor models of electricity spot prices is addressed by an MCMC algorithm. In particular, the results of this new calibration procedure are compared with those provided by standard ML calibration, so as to assess the reliability and robustness of the estimation techniques.

---

## Bayesian Inference for directional data through homogeneous scoring rules

V. Mamei<sup>1</sup>, MONICA MUSIO<sup>2</sup>, E. Ruli<sup>3</sup>, L. Ventura<sup>3</sup>

<sup>1</sup>Dept. of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy;

<sup>2</sup>Dept. of Mathematics and Computer Science, University of Cagliari, Italy; <sup>3</sup>Dept. of Statistical Sciences, University of Padova, Italy

Observations that inherit a direction occur in many scientific fields where data are recorded as directions or angles relative to a system with a fixed orientation. There are several parametric distributions for modelling such kind of data. In general, let  $X$  be a random vector with density

$$f(x; \theta) = \frac{1}{c(\theta)} \exp\{h(x; \theta)\}, \quad (1)$$

where  $\theta \in \Theta \subseteq \mathcal{R}^q$ ,  $q \geq 1$  and  $h(x; \theta)$  is a known function. Typically, the normalizing constant  $c(\theta)$  is parameter-dependent and, in the general case, is not available in closed form. Therefore inference for models for directional data is challenging in either a frequentist and Bayesian settings because of the intractable nature of such normalising constant involved in the expression of the likelihood. Distributions with densities that can be written as (1) are commonly encountered not only when working with directional data, but also in spatial statistics (Cressie, 1991, Sect. 7.2). A remarkable example of (1) is given by the  $q$ -dimensional Bingham distribution (Bingham, 1974, Mardia and Jupp, 2000), given by

$$f(x; \theta) = \frac{1}{c(\theta)} \exp\{x^T \theta x\}, \quad x^T x = 1, x \in \mathcal{R}^q$$

In this work we propose a Bayesian approach for inference on  $\theta$  based on proper scoring rules (see e.g. Dawid and Musio (2014), and references therein), thus avoiding the calculation of the likelihood function. In particular we use the fact that proper scoring rules supply unbiased estimating equations, thus form a special case of M-estimation (see e.g. Huber and Ronchetti, 2009). We focus on the Hyvärinen scoring rule, which has the property of homogeneity, i.e. the quoted distribution need only to be known up to normalisation. Using results in Ruli et al. (2016), the scoring rule estimating function will be used as a summary statistic in ABC in order to obtain an accurate approximation to the posterior distribution. Applications for the Bingham distribution and the von Mises-Fisher distribution on the sphere are discussed.

## Inference for intractable-likelihoods using a likelihood-free version of the stochastic approximation EM algorithm

UMBERTO PICCHINI<sup>1</sup>

<sup>1</sup>Centre for Mathematical Sciences, Lund University, Sweden

Inference for intractable likelihoods is an important topic that has received enormous attention, especially in the Bayesian framework. Here we present an approximate maximum likelihood methodology for the parameters of incomplete data models. A likelihood-free version of the stochastic approximation expectation-maximization (SAEM) is constructed to maximize the likelihood function of model parameters. While SAEM is best suited for models having a tractable complete likelihood function, its application to moderately complex models is difficult, and results impossible for models having so-called intractable likelihoods. The latter are typically treated using approximate Bayesian computation (ABC) algorithms or synthetic likelihoods, where information

Cressie, N. A. C. (1991). *Statistics for Spatial Data*. Wiley.

Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *Annals of Statistics*, **2**, 12011225.

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley.

Dawid, A.P., Musio, M. (2014). Theory and applications of proper scoring rules. *Metron*, **72**, 169183.

Huber, P.J., Ronchetti, E.M. (2009). *Robust Statistics*. Wiley.

Ruli, E., Sartori, N. and Ventura, L. (2016). Approximate Bayesian Computation with composite score functions, *Statistics and Computing*, **26**, 3, 679692

from the data is carried by a set of summary statistics. While ABC is considered the state-of-art methodology for intractable likelihoods, its algorithms are often difficult to tune. On the other hand, synthetic likelihoods (SL) is a more recent methodology which is less general than ABC, it requires stronger assumptions but also less tuning. By exploiting the Gaussian assumption set by SL on data summaries, we can construct a likelihood-free version of SAEM. Our SAEM-SL method is completely plug-and-play, the ability to simulate realizations from the model being the only requirement. Although SAEM-SL is constructed in a non-Bayesian context, we show its potential to return inference for intractable likelihoods which are usually dealt with approximate Bayesian methods.

---

## Using Higher Order Approximations for Transition Densities in Bayesian Inference for Diffusions

SUSANNE PIESCHNER<sup>1</sup>, Christiane Fuchs<sup>1</sup>

<sup>1</sup>Institute of Computational Biology, Helmholtz Zentrum München, Germany

13 Jun  
17:00 - 19:00

Diffusion processes are a powerful tool in many areas of sciences to model continuous time dynamical systems which are subject to random fluctuations. They can be equivalently described by stochastic differential equations (SDEs). For the parameter estimation of SDEs from discrete observations, Markov chain Monte Carlo techniques have been developed that introduce auxiliary data points if required. These methods use numerical approximations of the transition densities of the process, both for the calculation of the posterior densities and for proposing auxiliary data points. In both contexts, the Euler-Maruyama scheme is the standard approximation technique in literature. In order to speed up the computationally expensive method, many authors suggest that higher order approximations may be employed instead. However, the specific implementation and benefit remain unclear. We investigate and demonstrate the utilization and usefulness of such higher order approximations on the example of the Milstein scheme.

---

## A rare event approach to high dimensional ABC

DENNIS PRANGLE<sup>1</sup>, Richard G. Everitt<sup>2</sup>, Theodore Kypraios<sup>3</sup>

<sup>1</sup>Newcastle University, UK; <sup>2</sup>University of Reading, UK; <sup>3</sup>University of Nottingham, UK

13 Jun  
17:00 - 19:00

Approximate Bayesian computation (ABC) methods permit approximate inference for intractable likelihoods when it is possible to simulate from the model. However they perform poorly for high dimensional data, and in practice must usually be used in conjunction with dimension reduction methods, resulting in a loss of accuracy which is hard to quantify or control. We propose a new ABC method for high dimensional data based on rare event methods which we refer to as RE-ABC. This uses a latent variable representation of the model. For a given parameter value, we estimate the probability of the rare event that the latent variables correspond to data roughly consistent with the observations. This is performed using sequential Monte Carlo and slice sampling to systematically search the space of latent variables. In contrast standard ABC can be viewed as using a more naive Monte Carlo estimate. We use our rare event probability estimator as a likelihood estimate within the pseudo-marginal Metropolis-Hastings algorithm for parameter inference.

We summarise asymptotics showing that RE-ABC has a lower computational cost for high dimensional data than standard ABC methods. We also illustrate our approach empirically, on a Gaussian distribution and an application in infectious disease modelling.

## Hierarchical Bayesian estimation of the dependence in a compound Poisson risk process

13 Jun  
17:00 - 19:00

GIOVANNI RABITTI<sup>1</sup>

<sup>1</sup>Bocconi University, Milan, Italy

In classical Risk Theory the aggregate amount of claims is typically modeled through a compound Poisson process, which relies on the assumption of independence between the claims arrival process and the associated random claim severities. In actuarial literature, not many contributions have studied the general issue of incorporating some degree of dependence between the two processes, using copula methods or the Sarmanov distribution.

However, still is missing a procedure to make fully Bayesian inference on this degree of dependence, i.e. to find relevant posterior quantities useful to test whether a posteriori the independence assumption is satisfied and, in this case, to calculate a more precise collective premium.

In this paper, we propose a hierarchical Bayesian model which allows these possibilities and we perform the analysis on a dataset from an Italian insurance company.

---

## Bayesian semiparametric modelling of contraceptive behavior in India via sequential logistic regressions

13 Jun  
17:00 - 19:00

TOMMASO RIGON<sup>1</sup>, Daniele Durante<sup>2</sup>, Nicola Torelli<sup>3</sup>

<sup>1</sup>Bocconi University, Milan, Italy; <sup>2</sup>University of Padova, Italy; <sup>3</sup>University of Trieste, Italy

In analyzing contraceptive behavior it is fundamental to flexibly model the impact of a set of covariates on the probabilities of selecting among alternative methods. Most of the literature focuses on generalized linear mixed models, with Gaussian random effects accounting for the hierarchical structure in the observed data. Inappropriately assuming the random effects to be normally distributed and enforcing parametric representations for the linear predictor can have a major impact on the quality of inferences. Motivated by recent large scale survey studies on the contraceptive behavior in India, we propose a Bayesian semiparametric model relying on sequential logistic regressions which account for the nested decision process characterizing the contraceptive choices. The model allows a subset of covariates to enter the predictor via Bayesian penalized splines and exploits the Dirichlet process formulation to flexibly represent uncertainty in the distribution of the State-specific random effects. Employing the Pólya-gamma data augmentation we develop an efficient Gibbs sampler, and apply the model to explore the determinants of the sequential decision process underlying the contraceptive behavior in India.

---

## A Bayesian hierarchical model for group fMRI and fMRI meta-analysis neuroimaging data

13 Jun  
17:00 - 19:00

PANTELIS SAMARTSIDIS<sup>1</sup>, Thomas E. Nichols<sup>2</sup>, Timothy D. Johnson<sup>3</sup>

<sup>1</sup>MRC Biostatistics Unit, University of Cambridge, UK; <sup>2</sup>University of Warwick, UK; <sup>3</sup>University of Michigan, Ann Arbor, U.S.A

A typical neuroimaging study will produce several 3D brain statistic images, one for each participant, that summarise the evidence for activation during the experiment. However, for practical reasons those images are rarely published; instead, authors only report the  $(x, y, z)$  locations of local maxima in the statistic images. Coordinate data from multiple studies can be used to find areas of consistent activation across the human brain. Neuroimaging meta-analysis is an area of growing interest in statistics. Several issues have been solved but there are also many still-open problems. For example, current methods do not account for study characteristics as in meta-regression. Also, it is still unclear how meta-analysis data can be used as prior in new fMRI studies.

In this work, we propose a Bayesian hierarchical model based on spatial point processes for the joint analysis of image data from new fMRI studies and coordinate data from previously conducted studies. The model is structured in 3 levels. At the bottom level we have the image data for each



participant. At each voxel (volume element) of the images we assume that the observed values arise from a mixture, where the mixing components are due to noise or due to the presence of some latent activation center. At the middle level we have the latent activation centers of each participant, the locations in the brain that are activated during the experiment. Those are modelled as spatial Poisson processes and cluster around the unobserved study centers which we find on the top level. Study centers are the locations of activation that are shared among the participants. We assume that study centers are a realisation of a log-Gaussian Cox process and use the meta-analysis data (study centers of previous studies) to estimate the intensity function. That way, the new study can borrow information from the meta-analysis. Note that we further use study characteristics as covariates to model the intensity.

Since a typical brain atlas consists of roughly 200,000 voxels, our model requires the estimation of several thousands of parameters. We describe a Markov Chain Monte Carlo algorithm that can be used for posterior estimation and discuss its mixing properties. Scalar hyperparameters are updated with either Gibbs or Metropolis steps, the spatial birth-and-death algorithm is used to simulate the latent study and activation centers, whereas we choose Hamiltonian Monte Carlo for the high dimensional intensity function. Computational considerations are also discussed. In particular, we find that application of the problem in a graphics processing unit (GPU) vastly reduces the total running time. Finally, we examine sensitivity of the model to prior specifications. Results are shown on both simulated and real data.

---

## Statistical inference for multivariate partially observed stochastic processes with application to neuroscience

Adeline Leclercq-Samson<sup>1</sup>, MASSIMILIANO TAMBORRINO<sup>2</sup>

<sup>1</sup>Laboratoire Jean Kuntzmann, CNRS, Université Grenoble-Alpes, France; <sup>2</sup>Institute for Stochastics, Johannes Kepler University Linz, Austria

13 Jun  
17:00 - 19:00

In many signal-processing applications, it is of primary interest to decode/reconstruct the unobserved signal based on some partially observed information. Some examples are automatic speech, face, gesture and handwriting recognition, and neuroscience (ion channels modeling). From a mathematical point of view, this corresponds to estimate model parameters of an unknown coordinate based on discrete observations of one or more other coordinates. Here we consider a partially observed bivariate stochastic process and discuss it in the framework of stochastic modelling of single neuron dynamics. None of the two components is directly observed: the available observations correspond to hitting times of the first component to the second component and/or measurements at discrete time of the first coordinate. Our aim is to provide statistical inference of the underlying model parameters, as well as developing suitable numerical algorithms. This is particularly difficult since the considered process does not fit into the well-known class of hidden Markov models, requiring the investigation of new ad-hoc mathematical and statistical techniques to handle it.

---

## Bayesian inference for modeling of protein folding data

SAMUEL WIQVIST<sup>1</sup>, Umberto Picchini<sup>1</sup>, Julie Lyng Forman<sup>2</sup>

<sup>1</sup>Centre for Mathematical Sciences, Lund University, Sweden; <sup>2</sup>Department of Biostatistics, University of Copenhagen, Denmark

13 Jun  
17:00 - 19:00

Modelling of protein folding is an important topic in biophysics. The inference problem for the model parameters is however often challenging, because of the size of the data, the nonlinear dynamics and the presence of local perturbations requiring computer intensive Monte Carlo methods. It is therefore of interest to construct a simple model for protein folding, useful for both frequentist and Bayesian inference. We introduced a double-well potential stochastic differential equation model with additive red-noise that has a decreased computational complexity, allowing us to generate data from the model more rapidly compared to an earlier model. But the inference problem for this model is still challenging since the likelihood function is intractable. However, by utilizing

recent developments for Monte Carlo based algorithms we could resort to particle Monte Carlo based methods to sample from the posterior distribution of the parameters or from a surrogate model of the likelihood function based on Gaussian processes.

Simulation studies show that the proposed double-well potential stochastic differential equation model fits the protein folding data well. However, the inference problem still constitutes a challenging problem due to the computational complexity of the particle Monte Carlo based algorithms and the properties of the proposed model.

---

# Wednesday, June 14

INVITED SESSION

## Stochastic Programming

### Optimisation simulation for stochastic programming

14 Jun  
9:30 - 11:00

Tahir Ekin<sup>1</sup>, NICHOLAS G. POLSON<sup>2</sup>, Refik Soyer<sup>3</sup>

<sup>1</sup>McCoy College of Business, Texas State University, U.S.A; <sup>2</sup>Booth School of Business, University of Chicago, U.S.A; <sup>3</sup>School of Business, The George Washington University, U.S.A

We develop a simulation-based approaches for stochastic optimisation. Using augmented probability models and nested sampling, we provide methods for solving complex decision problems with stochastic shocks. Our methods apply to two-stage stochastic programming with recourse, semi-Markov decision processes and deep Q-learning. We illustrate our methodology with a two stage newsvendor model with unimodal and bimodal continuous uncertainty. Finally, conclude with directions for future research.

### Quantifying and reducing uncertainties on sets under Gaussian Process priors

14 Jun  
9:30 - 11:00

DAVID GINSBOURGER<sup>1</sup>

<sup>1</sup>IDIAP, Martigny, Switzerland

Gaussian Process models have been used in a number of problems where an objective function  $f$  needs to be studied based on a drastically limited number of evaluations.

A number of approaches have been recently introduced in such set up for the estimation of sets implicitly defined by  $f$ , including level sets and also non-dominated sets in the multi-objective case.

Algorithms have been developed for decreasing uncertainties by evaluating  $f$ , sequentially or batch-sequentially, at points carefully chosen using dedicated sampling criteria.

In this talk, we will give an overview of recent results and applications, and also state some identified challenges on this emerging research topic.

Based on a series of joint works primarily with Dario Azzimonti, François Bachoc, Julien Bect, Mickaël Binois, Clément Chevalier, Ilya Molchanov, Victor Picheny, Yann Richet, and Emmanuel Vazquez.

14 Jun  
9:30 - 11:00

## Probabilistic Programming, Bayesian Nonparametrics, and Inference Compilation

FRANK WOOD<sup>1</sup>

<sup>1</sup>University of Oxford, UK

Probabilistic programming uses programming language techniques to make it easy to denote and perform inference in the kinds of probabilistic models that inform decision-making, accelerate scientific discovery, and underlie modern attacks on the problem of artificial intelligence.

Higher order probabilistic programming languages were inspired by Bayesian nonparametric modeling. Evaluators for such languages make prototyping and performing inference in a wide variety of Bayesian nonparametric models easy and feasible.

Deep learning uses programming language techniques to automate supervised learning of program parameter values by gradient-based optimization.

What happens if we marry all these things?

This talk will review higher-order probabilistic programming and show how many Bayesian nonparametric models can be straightforwardly represented. It will also introduce inference compilation and address how linking deep learning and probabilistic programming is leading to powerful new AI techniques while also opening up significant new research questions.

---

CONTRIBUTED SESSION

### Methods and computations

#### Bayesian analysis of a general class of multivariate exponential family of state space models

14 Jun  
11:30 - 13:10

TEVFIK AKTEKIN<sup>1</sup>, Nick Polson<sup>2</sup>, Refik Soyer<sup>3</sup>

<sup>1</sup>University of New Hampshire, U.S.A; <sup>2</sup>University of Chicago Booth School of Business, U.S.A; <sup>3</sup>George Washington University, U.S.A

In this paper, we propose a general class of multivariate exponential family of state space models and consider their Bayesian analysis using particle learning methods. Our proposed model can be considered to be a direct multivariate extension of the class of state space models developed in Gamerman et al. (2013) in the Journal of Time Series Analysis. Unlike most state space time series models in the literature, our proposed model is non-Gaussian that is typically observed in real time series applications in engineering, sciences, finance, and economics. One advantage of state space models is that past correlations are captured by the time evolution of the state parameter, which we refer to as the random common environment. Correlations among multiple series are induced by this random common environment which is assumed to follow a Markovian evolution. For instance, in reliability engineering, several components of the same machine might be exposed to the same common environment (dependence across series) and to the same wear and tear effects of time (dependence over time). In estimating the model parameters, we consider both fully adapted and importance sampling based particle learning methods. In doing so, we provide performance comparisons in terms of both speed of estimation and model fit across several candidates. We argue that for large scale estimation our proposed model provides significantly faster estimation with respect to commonly used Markov chain Monte Carlo (MCMC) methods. To show the implementation of the proposed models, we use several simulation studies and discuss applications in reliability analysis.

# Correlated pseudo marginal schemes for partially observed diffusion processes

14 Jun  
11:30 - 13:10

ANDREW GOLIGHTLY<sup>1</sup>, Emma Bradley<sup>1</sup>

<sup>1</sup>School of Mathematics & Statistics, Newcastle University, UK

Stochastic differential equations (SDEs) provide a natural framework for modelling intrinsic stochasticity inherent in many continuous-time physical processes. Performing fully Bayesian inference for such models, using discrete time data that may be incomplete and subject to measurement error is a challenging problem. One widely used approach is to replace unavailable transition densities with an Euler-Maruyama approximation, made sufficiently accurate through the introduction of intermediate times between observations. Integrating over the uncertainty in the process at these intermediate times typically necessitates the use of Markov chain Monte Carlo (MCMC). Pseudo marginal MCMC schemes are increasingly used, since for a given discretisation level, the observed data likelihood can be unbiasedly estimated using a particle filter. When observations are particularly informative, a diffusion bridge construct can be used to drive the particle filter. Recent work in state-space settings has shown how the pseudo marginal approach can be made much more efficient by correlating the underlying pseudo random numbers used to form the estimate of likelihood at the current and proposed values of the unknown parameters. We extend this approach to the discretised diffusion process framework by correlating the Gaussian innovations that drive the diffusion bridge construct used inside particle filter. We find that the resulting approach offers substantial gains in efficiency over a standard implementation.

---

## A dual Markov model for filtering problems

14 Jun  
11:30 - 13:10

FREDDY PALMA MANCILLA<sup>1</sup>, Ramsés H. Mena<sup>1</sup>

<sup>1</sup>Institute of Research in Applied Mathematics and Systems (IIMAS), National Autonomous University of Mexico (UNAM), Mexico

Within a hidden Markov framework, the law driving the signal process is of major interest in a wide variety of fields. Indeed, in such a context, the role played by the transition probability associated such a process turns out to be crucial for the computation of the filters. Here, we study a construction of continuous-time reversible Markov process with arbitrary but given invariant distribution, that, when assumed for the signal, leads us to computable and tractable expressions for the filters. Also, it allows us to calculate statistics associated to the filters. Furthermore, the construction guarantees the existence of a dual to the signal, which can be used to derive an alternative filter recursions. Some well know models fall within our construction.

---

## Towards unified Bayesian estimation of diffusions

14 Jun  
11:30 - 13:10

FRANK VAN DER MEULEN<sup>1</sup>, Moritz Schauer<sup>2</sup>, Omiros Papaspiliopoulos<sup>3</sup>,d Harry van Zanten<sup>4</sup>

<sup>1</sup>Delft University of Technology, Netherlands; <sup>2</sup>University of Leiden, Netherlands; <sup>3</sup>Universitat Pompeu Fabra, Spain; <sup>4</sup>University of Amsterdam, Netherlands

Bayesian estimation of discretely observed diffusions has received much attention over the past two decades. If the data generating SDE is given by

$$dX_t = b(t; X_t)dt + \sigma(t; X_t)dW_t;$$

then it is commonly assumed that the matrix  $a = \sigma\sigma'$  is non-degenerate. In applications, this uniform ellipticity assumption is not always met. As an example, consider the two dimensional system where the first component is the integral of the second:

$$dX_t = \begin{bmatrix} X_{t2} \\ g(t; X_t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \gamma(t; X_t) \end{bmatrix} dW_t.$$

This is an example of an integrated diffusion, or more generally, of a hypo-elliptic diffusion. To obtain samples from the posterior, it is common to use data-augmentation, where latent diffusion bridges in between discrete time observations are simulated. Unfortunately, many methods for simulating bridges crucially depend on the uniform ellipticity assumption. While hypo-elliptic diffusions are interesting from an application perspective, no general solution for Bayesian estimation has appeared in the literature. We explain how the approach of Schauer et al. can be adapted and generalised for this purpose. Although various non-trivial mathematical problems pop-up compared to the uniformly elliptic case, it turns out that with some modifications, the approach of Schauer et al. can be used for hypo-elliptic diffusions as well. This results in a unified method for Bayesian estimation in the sense that it is applicable to both the uniformly- and hypo-elliptic diffusions.

---

INVITED SESSION

## **Stochastic Processes in Paleo Climate Science (invited special session sponsored by the Past Earth Network)**

14 Jun  
14:30 - 16:00

### **Gaussian process emulators with applications to Paleo-climate simulations**

JOCHEN VOSS<sup>1</sup>, Dario Domingo<sup>1</sup>

<sup>1</sup>University of Leeds, UK

Gaussian process emulators are an increasingly popular tool to interpolate between (costly) runs of climate models. We will discuss mathematical aspects of this method, with an eye towards applications in paleo-climate modelling.

---

14 Jun  
14:30 - 16:00

### **How fast is sea level rising?**

ANDREW PARNELL<sup>1</sup>

<sup>1</sup>University College Dublin, Ireland

I will review the statistical approaches used to calculate rates of sea level change which are currently making the news worldwide. Many of the methods use Bayesian inference, require the merging of large, complex and messy data sets, and accounting for and presenting uncertainty is crucial. The statistical processes involved include errors-in-variables regression, smoothing approaches, derivative estimation, prior elicitation, and multinomial regression. I will outline some of the remaining challenges and how increased collaboration across subject boundaries might help solve them.

---

---

Schauer, M. and Van der Meulen, F. H. and Van Zanten, J. H. (2016) *Guided proposals for simulating multi-dimensional diffusion bridges*. Accepted for publication in Bernoulli, ArXiv e-prints 1311.3606.

## Quantifying uncertainties in paleoclimate reconstruction: A case study over the past 800 kyr

14 Jun  
14:30 - 16:00

JAKE CARSON<sup>1</sup>

<sup>1</sup>Imperial College, London, UK

Our understanding of the paleoclimate is based on proxy data taken from climate archives, and models of the climate that mathematically formulate hypotheses of long-term climate dynamics. Utilizing these sources of information is non-trivial. Proxy data are sampled at a sequence of depths that must be converted into ages through the use of an age model, and the climate models need to be calibrated using suitable proxy data. Each of these undertakings involves numerous sources of uncertainty, such as uncertainties in the age and parameter estimates, discrepancies between models and real-world system dynamics, and in how proxy measurements relate to the state of the system. Accurately quantifying these uncertainties, and in particular propagating uncertainties through the entire analysis, is essential if we are to trust in the inferences from these investigations. Performing these investigations in two distinct stages of analysis can make this difficult, particularly when strong dependencies exist between stages. In this talk we demonstrate that Bayesian methodology is now sufficiently advanced for a single joint analysis to be undertaken, which characterizes each of these sources of uncertainty.

Our motivating example is the glacial-interglacial cycle over the past 800 kyr. Over this period the climate exhibits oscillations between cold periods in which glaciers extended, and warm periods in which the glaciers retreated. This is clear in, for example, benthic records of  $^{18}O$ , which is a measurement of the ratio between  $^{18}O$  and  $^{16}O$  taken from calcite shells embedded in deep-sea sediment cores, and is primarily a function of global temperature and ice volume at the time the calcite shell was deposited. Models of the glacial-interglacial cycle are frequently characterized as either ODEs or SDEs that explicitly model only several climate variables. These are termed phenomenological models, as they are consistent with the underlying dynamics of the system, but not derived from the physical processes. The tasks we aim to undertake are fitting an age model to the sediment cores (age estimation), reconstructing components of the climate over time (climate reconstruction), estimating the parameters of a phenomenological model (model calibration), and determining which models are more supported by the data (model comparison).

---

INVITED SESSION

### Bayesian methods for network data

#### Interlocking directorates in Irish companies using a latent space model for bipartite networks

NIAL FRIEL<sup>1</sup>, Riccardo Rastelli<sup>1</sup>, Jason Wyse<sup>2</sup>, Adrian E. Raftery<sup>3</sup>

<sup>1</sup>University College Dublin, Dublin, Ireland; <sup>2</sup>Trinity College Dublin, Dublin, Ireland; <sup>3</sup>University of Washington, Seattle, US

14 Jun  
16:30 - 18:00

We analyze the temporal bipartite network of the leading Irish companies and their directors from 2003 to 2013, encompassing the end of the Celtic Tiger boom and the ensuing financial crisis in 2008. We focus on the evolution of company interlocks, whereby a company director simultaneously sits on two or more boards. We develop a statistical model for this dataset by embedding the positions of companies and directors in a latent space. The temporal evolution of the network is modeled through three levels of Markovian dependence: one on the model parameters, one on the companies' latent positions, and one on the edges themselves. The model is estimated using Bayesian inference. Our analysis reveals that the level of interlocking, as measured by a contraction of the latent space, increased before and during the crisis, reaching a peak in 2009, and has generally stabilized since then.

<http://www.pnas.org/content/113/24/6629.full.pdf>

14 Jun  
16:30 - 18:00

## Bayesian nonparametric modeling of dynamic networks

DANIELE DURANTE<sup>1</sup>

<sup>1</sup>University of Padova, Italy

Real world networks are often associated with a dynamic component, and the development of statistical methodologies to learn how the connectivity patterns among interacting actors are wired across time is a fundamental goal in many fields. The accurate characterization of these processes allows deeper insights in many complex phenomena, while providing inference and prediction strategies in different dynamical systems, covering social and epidemiological processes. Although the number of available contributions in statistical modeling of dynamic networks has registered an exponential growth in the recent years, current proposals still raise open questions about inference, flexibility and computational tractability. Motivated by time-varying data on face-to-face social interactions, I will present recent Bayesian nonparametric models for dynamic networks which characterize the edge probabilities as a function of actors' positions in a latent space, with these positions changing in time via Gaussian processes or nested Gaussian processes. These formulations have theoretical justification and incorporate flexibility along with adaptive dimensionality reduction, allowing improvements in posterior computation, inference and prediction of future contacts.

---

14 Jun  
16:30 - 18:00

## Sampling and Estimation for (Sparse) Exchangeable Graphs

VICTOR VEITCH<sup>1</sup>

<sup>1</sup>University of Toronto, Canada

Many popular statistical models for network valued-datasets fall under the remit of the graphon framework, which (implicitly) assumes the networks are densely connected. However, this assumption rarely holds for the real-world networks of practical interest. Recent work has introduced sparse exchangeable graphs, and the associated graphex framework, as a natural generalization of the graphon framework to the sparse graph regime. These new models are formally defined as the set of random measures satisfying a particular notion of exchangeability. We develop the graphex framework as a tool for statistical network analysis by identifying the sampling scheme that is naturally associated with the models of the framework, and by introducing a general consistent estimator for the parameter underlying these models.

---



# Thursday, June 15

INVITED SESSION

## Industrial methods

### Bayesian approaches to analyzing competing and semi-competing risks data

ANANDA SEN<sup>1</sup>

<sup>1</sup>University of Michigan, Ann Arbor, USA

15 Jun  
9:30 - 11:00

The topic of analyzing time-to-event data where individuals are subjected to multiple risks for the event occurrence has been well-studied for decades. In this framework a particular case that has received lion's share of attention is when the event is caused by the earliest onset of a cause, known as the case of competing risks. Earlier work in competing risks analysis utilized a series system (observing the minimum of several lifetimes) formulation in terms of latent event times. It is well known that such a formulation is fraught with the issue of identifiability, unless one can assume the different causes to act independently. In recent times, substantial efforts have been made to formulate a model that has direct links to the observables and avoids imposing a dependence structure on the causes. We present two scenarios arising in the context of analyzing competing risks data. The first focuses on the premise where the exact cause of the event may only be known partially, necessitating methodologies appropriate for handling missing data. The second deals with dependent censoring that acts as a competing risk to the main event of interest that is recurrent in nature. We shall conclude with some highlight of the difficulties of dependent competing risks formulation in general.

---

### Bayesian modelling of virtual age in repairable systems

DIDEM EGEMEN<sup>1</sup>, Fabrizio Ruggeri<sup>2</sup>, Refik Soyer<sup>1</sup>

<sup>1</sup> The George Washington University, Washington D.C., USA; <sup>2</sup> CNR-IMATI, Milano, Italy

15 Jun  
9:30 - 11:00

In this study, repairable system models, which are subject to minimal, perfect or imperfect repairs upon each failure, are discussed and a unifying model covering all these type of models is presented. Moreover, some extensions of this general model are proposed. These models are generally marked point processes,  $(T_1, Z_1), (T_2, Z_2), \dots, (T_n, Z_n)$  where  $T_i$ 's are failure times and  $Z_i$ 's are repair choices. The marks of this marked point process, i.e. repair actions, are assumed to be unknown and unobservable so modeled as latent variables. According to the dependence structure of the latent variables various models are developed. For the statistical analysis of these models, Bayesian framework is presented and posterior distributions are obtained through Markov Chain Monte Carlo methods.

## Sequential design of experiments and multi-objective optimization for the solutions of complex problems

15 Jun  
9:30 - 11:00

MATTEO BORROTTI<sup>1</sup>, Antonio Pievatolo<sup>2</sup>

<sup>1</sup>Energia Crescente S.r.l., Milano, Italy; <sup>2</sup>, CNR-IMATI Milano, Italy

The rapid growth of technology has allowed experimenters and practitioners to develop and test complex solutions to their problems. An increasing number of variables and system responses characterizes optimal solutions. In many real-life problems, system responses conflict with each other, and optimizing a particular solution with respect to a single system response can lead to unacceptable results. Multi-objective formulations are a realistic approach to the solutions of these problems. In this work, we develop a multi-objective sequential design of experiments approach (DOE) based on the Pareto optimality concept and the Bayesian framework. Different theoretical aspects are investigated and the proposed approach is compared with state-of-the-art methods on a simulation study.

---

INVITED SESSION

### Genomics and biostatistics

## Bayesian inference for latent biologic structure with determinantal point processes

15 Jun  
11:30 - 13:00

PETER MUELLER<sup>1</sup>, Yanxun Xu<sup>2</sup>, Donatello Telesca<sup>3</sup>

<sup>1</sup>University of Texas at Austin, U.S.A; <sup>2</sup>Johns Hopkins University, U.S.A; <sup>3</sup>UCLA, U.S.A

We discuss the use of the determinantal point process (DPP) as a prior for latent structure in biomedical applications, where inference often centers on the interpretation of latent features as biologically or clinically meaningful structure. Typical examples include mixture models, when the terms of the mixture are meant to represent clinically meaningful subpopulations (of patients, genes, etc.). Another class of examples are feature allocation models. We propose the DPP prior as a repulsive prior on latent mixture components in the first example, and as prior on feature-specific parameters in the second case. We argue that the DPP is in general an attractive prior model for latent structure when biologically relevant interpretation of such structure is desired. We illustrate the advantages of DPP prior in three case studies, including inference in mixture models for magnetic resonance images (MRI) and for protein expression, and a feature allocation model for gene expression using data from The Cancer Genome Atlas. An important part of our argument are efficient and straightforward posterior simulation methods. We implement a variation of reversible jump Markov chain Monte Carlo simulation for inference under the DPP prior, using a density with respect to the unit rate Poisson process.

---

## Bayesian response-adaptive designs for basket trials

15 Jun  
11:30 - 13:00

LORENZO TRIPPA<sup>1</sup>, Steffen Ventz<sup>2</sup>, William T. Barry<sup>1</sup>, Giovanni Parmigiani<sup>1</sup>

<sup>1</sup>Harvard University, School of Public Health, U.S.A; <sup>2</sup>University of Rhode Island, U.S.A

We develop a general class of response-adaptive Bayesian designs using hierarchical models, and provide open source software to implement them. Our work is motivated by recent master protocols in oncology, where several treatments are investigated simultaneously in one or multiple disease types, and treatment efficacy is expected to vary across biomarker-defined subpopulations. Adaptive trials such as I-SPY-2 (Barker et al., 2009) and BATTLE (Zhou et al., 2008) are special cases within our framework. We discuss the application of our adaptive scheme to two distinct research goals. The first is to identify a biomarker subpopulation for which a therapy shows evidence of treatment efficacy, and to exclude other subpopulations for which such evidence does not exist. This leads to a subpopulation-finding design. The second is to identify, within biomarker-defined subpopulations, a set of cancer types for which an experimental therapy is superior to the standard-of-care. This goal leads to a subpopulation-stratified design. Using simulations constructed to faithfully represent ongoing cancer sequencing projects, we quantify the potential gains of our proposed designs relative to conventional non-adaptive designs.

---

## Different coalescent modeling resolutions for posterior inference of evolutionary parameters

15 Jun  
11:30 - 13:00

JULIA PALACIOS<sup>1</sup>, Amandine Veber<sup>2</sup>

<sup>1</sup>Stanford University, U.S.A; <sup>2</sup>CMAP, École Polytechnique, France

Sophisticated inferential tools coupled with the coalescent model have recently emerged for estimation of evolutionary parameters, such as past population sizes, from present-day samples of genomic data. Coalescent-based methods rely on the Kingman-coalescent genealogy to model the sample's ancestry. Unfortunately, the state space of genealogies grows superexponentially with the number of samples and hence, inference is computationally challenging for large number of samples. Here, we present a new Bayesian approach that relies on lower resolution coalescent processes with drastically smaller hidden state spaces. We provide new algorithms for efficient and exact likelihood calculations exploiting a new representation on a directed acyclic graph. We compare the performance of our algorithms with state-of-the-art algorithms in population genetics.

---



# Presenting Authors

Aktekin  
Tevfik, 22

Arfè  
Andrea, 6, 11

Ayed  
Fadhel, 6

Battiston  
Marco, 6

Bianchini  
Ilaria, 7

Bloem-Reddy  
Benjamin, 7

Borrotti  
Matteo, 28

Cadonna  
Annalisa, 1

Calvia  
Alessandro, 7

Cappello  
Lorenzo, 8

Caron  
Francçois, 6

Carson  
Jake, 25

Cerquetti  
Annalisa, 8

Corbella  
Alice, 9

Crispino  
Marta, 10

Di Benedetto  
Giuseppe, 10

Durante  
Daniele, 26

Dutta  
Ritabrata, 11

Egemen  
Didem, 27

Egidi  
Leonardo, 11

Fasano  
Augusto, 12

Favaro  
Stefano, 2

Friel  
Nial, 25

Ginsbourger  
David, 21

Golightly  
Andrew, 23

Hartmann  
Marcelo, 13

Heaps  
Sarah, 2

Heller  
Katherine, 5

Iacopini  
Mateo, 13

Ickstadt  
Katja, 14

Kon Kam King  
Guillaume, 14

Livingstone  
Samuel, 15

Maleki  
Yasaman, 4

Marziali  
Andrea, 15

Mueller  
Peter, 28

Musio  
Monica, 16

Naveiro Flores  
Roi, 3

Palacios  
Julia, 29

Palma Mancilla  
Freddy, 23

Parnell  
Andrew, 24

Peruggia  
Mario, 3

Picchini  
Umberto, 16

Pieschner  
Susanne, 17

Plson  
Nicholas G., 21

Prangle  
Dennis, 17

Rabitti  
Giovanni, 18

Rigon  
Tommaso, 18

Samartsidis  
Pantelis, 18

Sen  
Ananda, 27

Steel  
Mark F.J., 1

Steorts  
Rebecca, 5

Tamborrino  
Massimiliano, 19

ter Horst  
Enrique, 4

Trippa  
Lorenzo, 29

van der Meulen  
Frank, 23

Varini  
Elisa, 4

Veitch  
Victor, 26

Voss  
Jochen, 24

Wiqvist  
Samuel, 19

Wood  
Frank, 22

Zanella  
Giacomo, 5