

**THIRD
BAYESIAN YOUNG
STATISTICIANS MEETING**



**Florence, Italy
June, 19-21, 2016**

Program

Abstracts

Conference Venues

Useful Information

Sun 19	Mon 20	Tue 21
	Registration (8.30-9.15)	Keynote (9-10)
	Keynote (9.15-10.15)	
	Talks (10.30-12.30)	Talks (10.15-13)
	Talks (13.45-15.45)	Keynote (14.30-15.30)
Informal meeting (16.45-20)	Keynote (16-17)	Talks (15.45-17.45)
	Poster session (17-19)	Keynote (18-19)
	Social dinner	Awards (19-20)

SCIENTIFIC PROGRAM

Monday June 20th

Keynote 1 9.15 – 10.15

Causal inference: a Bayesian perspective
Fabrizia Mealli, University of Florence

Session 1 10.30 – 12.30

[Discussant: Fabrizio Ruggeri]

10.30-10.50 Dynamic Bayesian Predictive Synthesis in Time Series Forecasting
Kenichiro McAlinn*, Mike West

10.50-11.10 Bayesian inference for high-frequency data using particle filtering
Andrea Cremaschi*, Jim Griffin

11.10-11.30 Sparse time-varying covariance estimation in many dimensions
Gregor Kastner*

11.30-11.50 Bayesian Risk Evaluation in State Space Models using Importance Sampling
Agnieszka Borowska*, Lennart F. Hoogerheide, Siem Jan Koopman

11.50-12.10 Dynamic clustering for spatiotemporal data
Lucia Paci*, Francesco Finazzi

12.10-12.30 Forecasting patterns and trends in seasonality of all-cause mortality in the USA:
a Bayesian hierarchical spatiotemporal model
Robbie Parks*, Vasilis Kontis, James Bennett, Ralf Toumi, Majid Ezzati

Session 2 13.45-15.45

[Discussant: Daniela Cocchi]

13.45-14.05 A Bayesian Joint Dispersion Model with Flexible Links
Rui Martins*

14.05-14.25 Priors for degrees of freedom in penalized spline regression
Massimo Ventrucci*

14.25-14.45 An application of Bayesian seemingly unrelated regression models with flexible tails
Charles Au*, S.T. Boris Choy

14.45-15.05 Bayesian inference in Longitudinal Regression Discontinuity Designs
Laura Forastiere*

15.05-15.25 Bayesian nonparametric estimation of a Conditional copula density function
Luca Rossini*, Fabrizio Leisen, Luciana Dalla Valle

15.25-15.45 Bayesian spatial model selection for identifying chemical plumes in hyperspectral imagery using data augmentation, mixtures of g-priors and Markov random field priors
Nicole Mendoza*, Abel Rodriguez

Keynote 2 16 – 17

Bayesian models for the analysis of neuroimaging data
Marina Vannucci, Rice University

Poster Session 17 – 19

Bayesian nonparametric sparse seemingly unrelated regression model (SUR)
Luca Rossini*, Roberto Casarin, Monica Billio

Bayesian prediction for diffusion processes based on the Euler approximation scheme
Simone Hermann*

Bayesian estimation on controlled branching processes: Robustness via disparities
Miguel González, Carmen Minuesa*, Inés del Puerto

Localization in High-Dimensional Monte Carlo Filtering
Sylvain Robert*, Hans R. Künsch

Sequential Monte Carlo methods in random intercept models for longitudinal data
Danilo Alves*, Carmen Armero Cervera, Anabel Forte, Nicolas Chopin

Linear inverse problem with range prior on correlations and its Variational Bayes inference
Ondrej Tichy*, Vaclav Smidl

A covariate-dependent nonparametric model for clustering
Ilaria Bianchini*, Raffaele Argiento, Alessandra Guglielmi

Bayesian survival analysis to model plant resistance and tolerance to virus diseases
Elena Lázaro Hervás*, Carmen Armero Cervera, Luis Rubio Miguélez

A Bayesian models for describing and predicting the stochastic demand of emergency calls
Vittorio Nicoletta*, Alessandra Guglielmi, Ettore Lanzarone, Valérie Belangér, Angel Ruiz

Likelihood tempering in dynamic model averaging
Jan Reichl, Kamil Dedecius*

Bayesian consistency of the bivariate extremes spectral measure estimation
Khader Khadraoui*, Pierre Ribereau

Optimizing movement of cooperating pedestrians by exploiting floor-field model
and Markov decision process
Vladimíra Sečkárová*, Pavel Hrabák

Extension in TVP Model Shrinkage
Angela Bitto*

Identification of patient dependent parameters in a kinetic model of fluid
and mass transfer during dialysis
Camilla Bianchi*, Ettore Lanzarone, Giustina Casagrande, Maria Laura Costantino

Parallel Split-Merge MCMC for the HDP
Stefania Perego*, Debora Parisi*

A Bayesian spatio-temporal Markov switching model for the detection of influenza outbreaks
Rubén Amorós Salvador*, David Conesa, Antonio Lopez-Quilez, Miguel Angel Martinez-Beneito

Fast, approximate MCMC for Bayesian analysis of large data sets: A design based approach
Matthias Kaeding*

Randomization Inference and Bayesian Inference in Regression Discontinuity Design:
An application to Italian University grants
Federica Licari*

Identification and Estimation of Principal Causal Effects in Randomized Experiments with
Treatment Switching
Emanuele Gramuglia*

Bayesian methods for microsimulation models
Consuelo Nava*, Cinzia Carota, Ugo Colombino

Closed Form Bayesian Inferences for Binary Logistic Regression via Polynomial Expansions
Kevin Dayaratna*

Reverse Random Walk Prior for Regression Discontinuity Designs
Federico Ricciardi*, Gianluca Baio

The Science behind the Magic: The Relation of Harry Potter “Sorting Hat Quiz” To Measures of
Values and Personality - A Bayesian analysis
Lea Jakob*, Eduardo Garcia-Garzon, Hannes Jarke

Inference for Coupled SDE: Metropolis Algorithms via Density Tracking by Quadrature
Harish Bhat*, R. W. M. A. Madushani, Shagun Rawat

M-quantile bayesian regression in Small Area Estimation
Enrico Fabrizi, Nicola Salvati, Carlo Trivisano, Giovanni Riccardi*

Bayesian autoregressive models for waiting times of recurrent events
Giorgio Paulon*, Maria De Iorio, Alessandra Guglielmi

Multiple membership model for data integration using the hierarchical beta process
Takoua Jendoubi*, Korbinian Strimmer

Tuesday June 21th

Lecture 9 – 10

The Bayesian nonparametric approach to statistics via exchangeability
Alessandra Guglielmi, Politecnico di Milano

Session 3 10.15 – 11.35 **[Discussant: Emanuela Dreassi]**

10.15-10.35 Analysis of Substitution Times in Soccer
Rajitha Silva*, Tim Swartz

10.35-10.55 Bayesian Inference for Continuous Time Animal Movement Based on Steps and Turns
Alison Parton*, Paul Blackwell, Anna Skarin

10.55-11.15 Explaining the Lethality of Boko Haram's Terrorist Attacks in Nigeria, 2009-2014: a Hierarchical Bayesian Approach
Andre Python*, Janine Illian, Charlotte Jones-Todd, Marta Blangiardo

11.15-11.35 Approximate Bayesian Computation Methods in the identification of atmospheric contamination sources for DAPPLE experiment
Piotr Kopka*, Anna Wawrzynczak, Mieczyslaw Borysiewicz

Session 4 11.40-13 **[Discussant: Brunero Liseo]**

11.40-12 On the truncation error of a superposed gamma process
Julyan Arbel*, Igor Prunster

12-12.20 Noninformative Analysis for Mixture Models: A Hierarchical Story
Clara Grazian*, Christian Robert

12.20-12.40 An extension of Clayton Lévy copulas with applications to Bayesian nonparametrics
Alan Riva Palacio*, Fabrizio Leisen

12.40-13 Bayesian mixture modeling for multiple spectral densities
Annalisa Cadonna*, Raquel Prado, Athanasios Kottas

Keynote 3 14.30-15.30

A/B testing with Bayesian multi-armed bandits
Steven Scott, Google

Session 5 15.45-17.45
[Discussant: Antonio Pievatolo]

15.45-16.05 Scalable Joint Modeling of Longitudinal and Point Process Data
Joseph Futoma*, Katherine Heller

16.05-16.25 Bayesian hierarchical model for assessment of climate model biases
Maeregu Woldeyes Arisido*, Carlo Gaetan, Davide Zanchettin, Angelo Rubino

16.25-16.45 A Bayesian nonparametric approach to ecological risk assessment
Guillame Kon Kam King*, Julyan Arbel, Igor Prunster

16.45-17.05 Local posterior concentration rate for multilevel sparse sequences
Nurzhan Nurushev*, Eduard Belitser

17.05-17.25 Inclusion of prior knowledge using hierarchical shrinkage priors to inform genetic
association studies
Miguel Pereira*, John Thompson, Cosetta Minelli

17.25-17.45 Preference learning from non-transitive pairwise preferences
Marta Crispino*, Elja Arjas, Valeria Vitelli, Arnaldo Frigessi

Keynote 4 18-19

Modeling and inference with feature allocation models
Peter Muller, University of Texas

Closing & Awards 19-20

ABSTRACTS

Keynotes

Causal inference: a Bayesian perspective

Fabrizia Mealli, University of Florence

Drawing inferences about causal effects of treatments, interventions and actions is central to decision making in many research disciplines. We review the Bayesian approach to causal inference under the potential outcome approach, which defines a causal effect as the comparison of potential outcomes under different treatment conditions for the same units. Focus will be on settings with intermediate variables, that is, post-treatment variables potentially affected by the treatment and also affecting the response, where Principal Stratification is used to define causal estimands and formally express structural and distributional assumptions. Bayesian inference is natural for causal inference, and PS analysis in particular, inherently a missing data problem under the potential outcome approach. The general structure of Bayesian inference will be presented and specific scientific applications will be illustrated, such as the analysis of clustered encouragement designs and trials with treatment switching. Important open questions will be discussed.

Bayesian models for the analysis of neuroimaging data

Marina Vannucci, Rice University

Functional magnetic resonance imaging (fMRI), a noninvasive neuroimaging method that provides an indirect measure of neuronal activity by detecting blood flow changes, has experienced an explosive growth in the past years. Statistical methods play a crucial role in understanding and analyzing fMRI data. Bayesian approaches, in particular, have shown great promise in applications. A remarkable feature of fully Bayesian approaches is that they allow a flexible modeling of spatial and temporal correlations in the data. In this talk I will describe Bayesian spatiotemporal models that detect task-related activation patterns as well as Bayesian hierarchical models for the estimation of brain connectivity.

A/B testing with Bayesian multi-armed bandits

Steven Scott, Google

An A/B test is really just another name for an experiment that happens to be conducted online. However, the mechanics of an online experiment are dramatically different than those of a medical, agricultural, or manufacturing experiment. In particular, the cost of a type-I error is much lower. The standard playbook for experimental design mandates that you control for type-I errors before doing anything else. This results in experiments that are needlessly conservative, sometimes by multiple orders of magnitude. An alternative approach is to run the A/B test as an optimization problem, where the goal is to minimize regret. A technique known as "Thompson sampling" is a simple, intuitive heuristic based on Bayesian reasoning that also happens to outperform all similar heuristics. Thompson sampling uses "probability matching" to manage the explore-exploit tradeoff in multi-armed bandit problems. Each new observation is assigned to an arm according to the probability of that arm being "the best". The heuristic can be applied across a very broad class of reward distributions, making it easy to incorporate many of the "good" ideas from classical experimental design into an online A/B test.

Modeling and inference with feature allocation models

Peter Muller, University of Texas

We discuss an application of feature allocation models to inference for tumor heterogeneity. We use a variation of Indian buffet process models to facilitate model-based imputation of hypothetical subpopulations of tumor cells, characterized by unique sets of somatic mutations and/or structural variants like copy number variations. Implementing posterior inference in this problem gives rise to several computational challenges. We discuss solutions based on fractional Bayes factors, MAD Bayes small variance asymptotics, and a reversible jump implementation for a determinantal point process.

Lecture

The Bayesian nonparametric approach to statistics via exchangeability

Alessandra Guglielmi, Politecnico di Milano

We will interpret the usual Bayesian approach via the notion of exchangeability. In particular, we will see how the Bayesian nonparametric approach can be understood as the most natural generalization of Bayesian parametric models when the prior does not select finite-dimensional families of distribution (as in the parametric case).

Oral session 1

Dynamic Bayesian Predictive Synthesis in Time Series Forecasting

Kenichiro McAlinn, Mike West*

We discuss model and forecast comparison, calibration, and combination from a foundational perspective. Bayesian predictive synthesis (BPS) defines a coherent theoretical basis for combining multiple forecast densities, whether from models, individuals, or other sources, and extends existing forecast pooling and Bayesian model mixing methods. Time series extensions are implicit dynamic latent factor models, allowing adaptation to time-varying biases, mis-calibration, and dependencies among models or forecasters. Bayesian simulation-based computation enables implementation. A macroeconomic time series study highlights insights into dynamic relationships among synthesized forecast densities, as well as the potential for improved forecast accuracy at multiple horizons.

Bayesian inference for high-frequency data using particle filtering

Andrea Cremaschi, Jim Griffin*

Financial prices are usually modelled as continuous, often involving geometric Brownian motion with drift, leverage and possibly jump components. An alternative modelling approach allows financial observations to take integer values that are multiples of a fixed quantity, the ticksize - the monetary value associated with a single change during the asset evolution. These samples are usually collected at irregularly-spaced time points, as in the case of high-frequency data, exhibiting diverse trading operations in a few seconds. In this context, the observables are modelled via the Skellam process - defined as the difference between two Poisson processes. Volatility modelling is included in the analysis using the class of discretised Gaussian Ornstein-Uhlenbeck AR(1) processes. In particular, the variances of the likelihood process is a function of the AR(1) process. The time series structure of the model makes it suitable for inference via Particle Filtering methods, of which a performance comparison between different methods is provided.

Sparse time-varying covariance estimation in many dimensions*Gregor Kastner**

Dynamic covariance estimation for multivariate time series suffers from the curse of dimensionality; as a consequence, parameter parsimony plays an important role in reliable statistical inference. We address this issue by modeling the underlying dynamics of a time series vector through a lower dimensional collection of latent factors that allow for time-varying stochastic volatilities. Furthermore, we apply a Normal-Gamma prior to the elements of the factor loadings matrix. This hierarchical shrinkage prior is a generalization of the Bayesian lasso and effectively pulls the factor loadings of unimportant factors towards zero, thereby increasing sparsity even more. To guarantee efficiency of the estimation procedure, we employ a fully Bayesian yet computationally feasible approach to obtain draws from the high-dimensional posterior and predictive distributions via Markov chain Monte Carlo (MCMC) samplers. We utilize several variants of an ancillarity-sufficiency interweaving strategy (ASIS) to boost efficiency when sampling the factor loadings as well as the parameters driving the time-varying volatilities. The effectiveness of the approach is demonstrated through extensive simulation studies. Furthermore, the model is applied to a 300-dimensional vector of stock returns to evaluate predictive performance for financial data. Additionally to being a stand-alone tool, the algorithm is designed to act as a "plug and play" extension for other MCMC samplers.

Bayesian Risk Evaluation in State Space Models using Importance Sampling*Agnieszka Borowska*, Lennart F. Hoogerheide, Siem Jan Koopman*

We present a novel approach to Bayesian estimation of two financial risk measures, Value at Risk and Expected Shortfall, in nonlinear, non-Gaussian state space models. In particular, we consider two specifications of the stochastic volatility model: with normal and Student's *t* observation disturbances. The key insight behind our proposed importance sampling based approach is to accurately approximate the optimal importance density, which focuses on the augmented parameter subspace corresponding to high losses. By oversampling the extreme scenarios and punishing them by lower importance weights, we achieve a much higher precision in characterising the properties of the left tail. We report substantial gains in the accuracy of estimates in an empirical study on daily financial data.

Dynamic clustering for spatiotemporal data*Lucia Paci*, Francesco Finazzi*

A finite space-time mixture model is proposed to identify clusters in spatiotemporal data according to the underlying space-time process driving the observations. Spatiotemporally varying mixing weights are introduced to allocate observations collected at near locations and time points with similar cluster membership's probability. Indeed, mixing probabilities are driven by a latent spatial process that evolves over time, leading to a dynamic clustering varying over time and space. For each cluster, a common temporal pattern is estimated from a dynamic process. Fully posterior inference is provided under a Bayesian framework through Monte Carlo Markov Chain algorithms.

Forecasting patterns and trends in seasonality of all-cause mortality in the USA: a Bayesian hierarchical spatiotemporal model*Robbie Parks,* Vasilis Kontis, James Bennett, Ralf Toumi, Majid Ezzati*

All-cause mortality is known to exhibit seasonal variation. In this study, we will use all-cause mortality records of the entire USA from 1982-2010 for forecasting of seasonal age-specific mortality on a state level, analysing differences in trends. The novel approach of this research comes from the Bayesian hierarchical model, which borrows strength by neighbouring location, age group, and time. Our study will be the first systematic analysis of seasonality of the entire USA throughout this time period stratified by gender, age group, and state. Initial results from the in-sample model indicate a distinct difference between seasonal mortality profiles of younger and older age groups. Younger age groups relative mortality peaks in the summer months, while older age groups peak in the winter months. Further analysis over geography will

determine if this variation is constant across locations. We expect the results to improve understanding on how distinct age groups and locations are affected by season, as previous studies have looked at all age groups combined. We also expect the model framework to enable coherent forecasts of patterns and trends of seasonal mortality.

Oral session 2

A Bayesian Joint Dispersion Model with Flexible Links

*Rui Martins**

The objective is to model longitudinal and survival data jointly taking into account the dependence between the two responses in a real HIV/AIDS dataset using a shared parameter approach inside a Bayesian framework. We propose a linear mixed effects dispersion model for the CD4 longitudinal counts with a between-individual heterogeneity in the mean and variance, relaxing the usual assumption of a common variance for the longitudinal residuals. A hazard regression model is considered in addition to model the time since HIV/AIDS diagnostic until failure, where the coefficients accounting for the linking between the processes are time-varying. This flexibility is specified using Penalized Splines and allows the relationship to vary in time. Because residual heteroscedasticity may be related with the survival, the standard deviation is considered as a covariate in the hazard model thus enabling to study the effect of the CD4 counts' stability on the survival.

Priors for degrees of freedom in penalized spline regression

*Massimo Ventrucchi**

Bayesian P-splines assumes an intrinsic Gaussian Markov random field prior on the spline coefficients, conditional on a precision parameter τ . Prior elicitation of τ is difficult, because τ depends on the number of basis functions selected by the user. To overcome this problem we aim to building priors on an interpretable property of the model indicating the complexity of the smooth function to be estimated. These new prior are easy-to-elicited based on prior knowledge about the degree of an equivalent polynomial. Also, their interpretation is invariant to different choices for the number of knots. We present the general ideas behind the construction of priors for degrees of freedom and show an application with simulated data.

An application of Bayesian seemingly unrelated regression models with flexible tails

Charles Au, S.T. Boris Choy*

Seemingly unrelated regression (SUR) models are useful for capturing the correlation structure between different regression equations. While the multivariate normal error distribution is a common choice in an SUR model, the multivariate (t)-distribution is also popular for robustness considerations. However, the usual version of the multivariate (t)-distribution has the limitation that the degrees of freedom of its marginal distributions are identical. In this paper, we consider a version of the multivariate Student-(t) error distribution with flexible shape parameters for the marginal distributions in SUR models. We express this non-elliptical distribution into a scale mixtures of normal form and use Markov chain Monte Carlo (MCMC) algorithms for Bayesian inference. In the empirical study, the modified multivariate (t)-distribution gives a better fit to the excess market return data under a capital asset pricing model (CAPM) than the multivariate normal and Student-(t) distributions.

Bayesian inference in Longitudinal Regression Discontinuity Designs*Laura Forastiere**

Motivated by the evaluation of Italian university grants, we consider a RD design where the treatment is dynamically assigned according to a sequence of cut-off rules, which are based on time-varying multiple forcing variables. We use a probabilistic formulation of the assignment mechanism underlying RD designs by making a local overlap assumption for a subpopulation that is defined by multiple thresholds. We invoke a local latent sequential ignorability assumption to identify and estimate the causal effect of sequences of treatments. We propose a Bayesian approach to select the subpopulation and to draw inference on the target causal estimands.

Bayesian nonparametric estimation of a Conditional copula density function*Luca Rossini*, Fabrizio Leisen, Luciana Dalla Valle*

Copula models have been used in a variate of fields of interest (such as twin and life expectancies literature and financial time series analysis) and in particular the conditional copula models are popular for understanding the dependence between different variables through an underlying covariate. In this work, we focus our analysis on the possible relations between the total score at a high-school level of a couple of twins with respect to different covariate variables (such as the father's level of education or the family's income). To solve this problem, we propose a Bayesian nonparametric approach to the estimation of a conditional copula. Our methodology extends the work of Walker, Wu and Wang (2015), by introducing the dependence from the covariate in the infinite mixture model. The effectiveness of the method is illustrated with simulated data and an example from the National Merit Twin Study.

Bayesian spatial model selection for identifying chemical plumes in hyperspectral imagery using data augmentation, mixtures of g-priors and Markov random field priors*Nicole Mendoza* Abel Rodriguez*

The use of hyperspectral imagery in the remote sensing of gas plumes has proven to be important for a wide variety of military and environmental applications. For example, hyperspectral images can be used to detect a gas cloud invisible to the human eye, and to identify its chemical structure. A hyperspectral image is a massive cube of data consisting of thousands of pixels each with ~ 100 observations over a range of frequencies in the electromagnetic spectrum. Algorithms that use hypothesis testing and assume independence over pixels have shown success in detecting gas clouds, but fail in identifying chemical components. In this paper we explain that identification is a variable/model selection problem, which can be solved robustly by taking advantage of spatial information in the image. We develop a Bayesian spatial model selection algorithm, which uses mixtures of g-priors and incorporates a Gaussian Markov Random Field prior to induce dependence among neighboring pixels. We apply our model to several partially synthesized hyperspectral images, and through the use of confusion matrices and multi-class classification metrics, we show that our method outperforms state-of-the-art algorithms such as the LASSO and fused LASSO.

Oral session 3**Analysis of Substitution Times in Soccer***Rajitha Silva*, Tim Swartz*

This paper considers the problem of optimal substitution times in soccer. An analysis is presented based on Bayesian logistic regression. We find that with evenly matched teams, there is a goal scoring advantage to the trailing team during the second half of a match. We observe that there is no discernible benefit for a team that has substituted more than its opposition. Moreover, this remark entails all possible substitution times taken during the second half of a match.

Bayesian Inference for Continuous Time Animal Movement Based on Steps and Turns

Alison Parton, Paul Blackwell, Anna Skarin*

Although animal locations gained via GPS, etc. are typically observed on a discrete time scale, movement models formulated in continuous time are preferable in order to avoid the struggles experienced in discrete time when faced with irregular observations or the prospect of comparing analyses on different time scales. A class of models able to emulate a range of movement ideas are defined by representing movement as a combination of stochastic processes describing both speed and bearing. A method for Bayesian inference for such models is described through the use of a Markov chain Monte Carlo approach. Such inference relies on an augmentation of the animal's locations in discrete time that have been observed with error, with a more detailed movement path gained via simulation techniques. Real data on an individual reindeer (*Rangifer tarandus*) illustrates the presented methods.

Explaining the Lethality of Boko Haram's Terrorist Attacks in Nigeria, 2009-2014: A Hierarchical Bayesian Approach

Andre Python, Janine Illian, Charlotte Jones-Todd, Marta Blangiardo*

Since 2009, Nigeria has been the scene of numerous deadly terrorist attacks perpetrated by the terrorist group Boko Haram. In response to this threat, stakeholders in the fight against terrorism have deployed various counterterrorism policies, the costs of which could be reduced through efficient preventive measures. Statistical models able to integrate complex spatial dependence structures have not yet been applied, despite their potential for providing guidance to assess characteristics of terrorist attacks. In an effort to address this shortcoming, we use a flexible approach that represents a Gaussian Markov random field through stochastic partial differential equation and model the fine-scale spatial pattern of the lethality of terrorism perpetrated by Boko Haram in Nigeria from 2009 to 2014. Our results suggest that the lethality of terrorist attacks is contagious in space and attacks are more likely to be lethal in higher altitude and far from large cities.

Approximate Bayesian Computation Methods in the identification of atmospheric contamination sources for DAPPLE experiment

Piotr Kopka, Anna Wawrzynczak, Mieczyslaw Borysiewicz*

Sudden releases of harmful material into a densely-populated area pose a significant risk to human health. The apparent problem of determining the source of an emission in urban and industrialized areas from the limited information provided by a set of released substance concentration measurements is an ill-posed inverse problem. When the only information available is a set of measurements of the released substance concentration in urban and industrial areas, it is difficult to determine the source of emission. The Bayesian probability framework provides a connection between model, observational and additional information about the contamination source. The posterior distribution of the source parameters was sampled using an Approximate Bayesian Computation (ABC) algorithm. The stochastic source determination method was validated against the real data set acquired in a highly disturbed flow field in an urban environment. The datasets used to validate the proposed methodology include the dispersion of contaminant plumes in a fullscale field experiment performed within the project 'Dispersion of Air Pollutants and their Penetration into the Local Environment in London (DAPPLE)'.

Oral session 4

On the truncation error of a superposed gamma process

Julyan Arbel, Igor Prunster*

Completely random measures (CRM) are a central building block for countless stochastic models and are also ubiquitous in Bayesian Nonparametrics. A constructive series representation of CRMs, due to Ferguson

and Klass, provides the jumps in decreasing order. This feature is of primary interest when it comes to sampling since it minimizes the truncation error for a fixed truncation level of the series. In order to quantify the quality of the approximation, Arbel and Prünster (2015) introduce a moment-matching criterion which consists in evaluating a measure of discrepancy between actual moments of the CRM and moments based on the simulation output. Here we consider a general class of CRMs, namely superposed gamma process, derive a bound in probability of the truncation error and show the applicability of the moment-matching criterion.

Noninformative Analysis for Mixture Models: A Hierarchical Story

Clara Grazian, Christian Robert*

Mixture models are a flexible tool to represent complicated data, because they may be used for kurtotic, asymmetric or multimodal observations. In a Bayesian setting, defining a prior distribution is challenging, because it is difficult to analyze the influence of the prior on the inferential result. On the other hand, a noninformative analysis is not yet fully developed. For instance, Jeffreys priors usually are well-defined for the parameters of mixtures of distributions, but not available in closed form. We will review the literature about noninformative priors for mixture models and present a novel approach which allow for improper priors, based on a hierarchical representations of mixture models.

An extension of Clayton Lévy copulas with applications to Bayesian nonparametrics

Alan Riva Palacio, Fabrizio Leisen*

The Clayton Lévy, copula has recently been used to build dependent Bayesian nonparametric priors. A nice feature of this copula is that it depends by a parameter which regulates the degree of dependence. In this work we propose a novel two parameter family of Lévy, copulas which includes as special case the Clayton Lévy, copula. We will illustrate applications of this new copulas to modelling dependent data when the following marginals are considered: 1) Compound poisson processes, 2) Stable processes. Moreover, we will show that this copula can generate dependent stable processes which have been recently proposed in the literature (Griffin and Leisen, 2016). This novel dependent stable process will be used to estimate the survival function of dependent data in a Bayesian nonparametric setting.

Bayesian mixture modeling for multiple spectral densities

Annalisa Cadonna, Raquel Prado, Athanasios Kottas*

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, as in neuroscience, econometrics and geoscience. We develop a Bayesian modeling approach to modeling and inference for multiple spectral densities. We propose a local Gaussian mixture approximation to the Whittle log-likelihood. Under this approximation, the implied model for the log-spectral density is a mixture of linear functions with frequency-dependent logistic weights, which allows for general shapes for smooth spectral densities. The approach presented facilitates efficient posterior simulation as it casts the spectral density estimation problem in a mixture modeling framework for density estimation and sets the stage for hierarchical extensions for spectral analysis of multiple time series. The methodology is illustrated with synthetic and real data sets. Specifically, the method is applied to multichannel electroencephalographic recordings.

Oral session 5

Scalable Joint Modeling of Longitudinal and Point Process Data

Joseph Futoma, Katherine Heller*

A major challenge for personalized medicine is to provide individualized predictions about the future

trajectory of a disease. However, for many complex chronic diseases, a patient may also simultaneously have many comorbid conditions. Accurate determination of the risk of developing serious complications associated with the disease or its comorbidities may be more clinically useful than prediction about future trajectory of the disease itself in these cases. We propose a novel probabilistic generative model that jointly provides individualized predictions of future disease progression while also modeling the pattern of related recurrent adverse events. We fit our model using a scalable variational inference algorithm and apply our method to a large dataset of longitudinal patient records. Our model gives good performance in terms of both prediction of future disease trajectories and of future serious events, and our predictions are being incorporated by our local accountable care organization during chart reviews of high risk patients.

Bayesian hierarchical model for assessment of climate model biases

Maeregu Woldeyes Arisido, Carlo Gaetan, Davide Zanchettin, Angelo Rubino*

Assessments of climate change rely on numerical outputs simulated from Global Climate Models coupling the dynamics of ocean and atmosphere (GCMs). GCMs are, however, notoriously affected by substantial systematical errors (biases), whose assessment is essential to assert the accuracy and robustness of simulated climate features. This contribution focuses on constructing a Bayesian hierarchical model for the assessment of climate model biases in a multi-model framework. The method combines information from a multi-model ensemble of GCM simulations to provide a unified assessment of the bias. It further individuates different bias components that are characterized as non-stationary spatial fields accounting for spatial dependence.

A Bayesian nonparametric approach to ecological risk assessment

Guillame Kon Kam King, Julyan Arbel, Igor Prunster*

We revisit a classical method for ecological risk assessment using a Bayesian nonparametric approach. By resorting to nonparametric mixture models it is possible to overcome a historically debated parametric assumption while retaining the ability to deal with small datasets that are typical of ecological risk assessment.

Local posterior concentration rate for multilevel sparse sequences

Nurzhan Nurushev, Eduard Belitser*

We consider empirical Bayesian inference in the many normal means model in the situation when the high-dimensional mean vector is multilevel sparse, that is, most of the entries of the parameter vector are some fixed values. For instance, the traditional sparse signal is a particular case of multilevel sparse sequences. We apply an empirical Bayesian approach, namely we put an appropriate prior modeling the multilevel sparsity and make data-dependent choices of certain parameters of the prior. Next we study the local (i.e., depending on the "true" parameter) contraction rate of the resulting empirical Bayes posterior. As another consequence, we also derive the oracle estimation result. Adaptive minimax results (for the estimation and posterior contraction problems) over sparsity classes follow from our local results if the sparsity level is of polynomial order. We illustrate the results by simulations.

Inclusion of prior knowledge using hierarchical shrinkage priors to inform genetic association studies

Miguel Pereira, John Thompson, Cosetta Minelli*

Genome-Wide Association Studies are classically analysed by estimating single nucleotide polymorphisms (SNP) effects individually and adjusting for multiple testing. However, SNPs identified so far explain a small proportion of the variability of most traits. The integration of prior biological information in the analysis has been proposed as a way to improve SNP detection, which can be achieved in a multitude of ways. We propose the use of external information in a Bayesian hierarchical model with shrinkage priors that estimates the effects of SNPs clustered in genomic regions by applying differential shrinkage to each SNP. We perform a simulation study and apply different combinations of shrinkage parameters, according to prior knowledge, to assess the applicability of this approach. Our results show that the model performs

significantly better than the classical analysis when less shrinkage towards 0 (no effect) is applied to the SNPs in the causal genomic region. Translating several levels of prior knowledge into differential shrinkage by using a linear or an exponential correspondence also resulted in better performance with the bayesian model better ranking the SNPs in the causal region. These results suggest that using differential shrinkage is a possible strategy to integrate prior knowledge in the setting of genetic association studies.

Preference learning from non-transitive pairwise preferences

Marta Crispino, Elja Arjas, Valeria Vitelli, Arnoldo Frigessi*

We develop a full Bayesian probabilistic method to learn preferences from non-transitive pairwise comparison data. Such lack of transitivity can easily arise when the number of pairwise comparisons is large, and they are given in a sequential manner without allowing for consistency check. We develop a Bayesian Mallows model to handle such data, by adding a latent layer of uncertainty which captures the generation of preference misreporting. Inference is based on a MCMC algorithm that allows a fully probabilistic analysis.

Poster Session

Bayesian nonparametric sparse seemingly unrelated regression model (SUR)

Luca Rossini, Roberto Casarin, Monica Billio*

Seemingly unrelated regression (SUR) models are used in studying the interactions among economic variables of interest. In a high dimensional setting and when applied to large panel of time series, these models have a large number of parameters to be estimated and suffer of inferential problems. We propose a Bayesian nonparametric hierarchical model for multivariate time series in order to avoid the overparametrization and overfitting issues and to allow for shrinkage toward multiple prior means with unknown location, scale and shape parameters. We propose a two-stage hierarchical prior distribution. The first stage of the hierarchy consists in a lasso conditionally independent prior distribution of the Normal-Gamma family for the SUR coefficients. The second stage is given by a random mixture distribution for the Normal-Gamma hyperparameters, which allows for parameter parsimony through two components. The first one is a random Dirac point-mass distribution, which induces sparsity in the SUR coefficients; the second is a Dirichlet process prior, which allows for clustering of the SUR coefficients. We provide a Gibbs sampler for posterior approximations based on introduction of auxiliary variables. Some simulated examples show the efficiency of the proposed. We study the effectiveness of our model and inference approach with an application to macroeconomics.

Bayesian prediction for diffusion processes based on the Euler approximation scheme

*Simone Hermann**

In many fields of statistical analysis, one is not only interested in the estimation of model parameters, but in a prediction. For stochastic processes, on the one hand, one can be interested in the prediction for the further development of the current, i.e. observed, series. On the other hand, prediction for a new series can be of interest. This work presents two Bayesian prediction procedures, that include the estimation uncertainty as well as the model variance. In a first algorithm, trajectories will be drawn, in a second, the predictive distribution is calculated. Both methods will be compared and analyzed with respect to their advantages and drawbacks.

Bayesian estimation on controlled branching processes: Robustness via disparities

Miguel González, Carmen Minuesa, Inés del Puerto*

One of the most remarkable models for the description of the growth of populations are the branching processes. Within the class of these stochastic processes, an appealing generalization of the classical model

is the controlled branching process (CBP). The distinctive characteristic of the aforementioned model is that the number of individuals with reproductive capacity in each generation is determined by a random control function. The dynamic of this process can be described as follows: each individual reproduces independently of the others according to the same probability law, called offspring distribution, and when the number of individuals in each generation is known, a random control mechanism determines the number of progenitors that participate in the subsequent reproduction process. The strong relation between the value of the parameters of this process and its behaviour has led recent research on this field to focus on the inferential theory. The aim of this work is to tackle the inference problem arising from this model developing a robust procedure against outliers in a Bayesian outlook. In this sense, it is remarkable that robust procedures against outliers have been barely studied in this field and all the results correspond to the frequentist context; hence, this work presents pioneer results for that problem. For our purpose, we consider the sample given by the entire family tree and assuming that the offspring distribution belongs to a very general one-dimensional parametric family, we make use of disparity measures. More specifically, the proposed method consists of replacing the log likelihood with an appropriately scaled disparity in the expression of the posterior distribution and considering Bayesian point estimators based on the resulting function. Asymptotic properties and robustness measures of these estimators, such as the influence curve, α -influence curves and the asymptotic breakdown point, are studied. Finally, the accuracy of the proposed method is illustrated by the way of simulated examples developed with the statistical software R.

Localization in High-Dimensional Monte Carlo Filtering

Sylvain Robert, Hans R. Künsch*

The high dimensionality and computational constraints associated with filtering problems in large-scale geophysical applications are particularly challenging for the Particle Filter (PF). Approximate but efficient methods such as the Ensemble Kalman Filter (EnKF) are therefore usually preferred. A key element of these approximate methods is localization, which is in principle a general technique to avoid the curse of dimensionality and consists in limiting the influence of observations to neighboring sites. However, while it works effectively with the EnKF, localization introduces harmful discontinuities in the estimated physical fields when applied blindly to the PF. In the present paper, we explore some possible local algorithms based on the Ensemble Kalman Particle Filter (EnKPF), a hybrid method combining the EnKF and the PF. A simulation study in a conjugate normal setup allows to highlight the trade-offs involved when applying localization to PF type of algorithms in the high-dimensional setting.

Sequential Monte Carlo methods in random intercept models for longitudinal data

Danilo Alvares, Carmen Armero Cervera, Anabel Forte, Nicolas Chopin*

Longitudinal modelling is common in the field of Biostatistical research. In some studies, it becomes mandatory to update posterior distributions based on new data in order to perform inferential process on-line. In such situations, the use of posterior distribution as the prior distribution in the new application of the Bayes' theorem is sensible. However, the analytic form of the posterior distribution is not always available and we only have an approximated sample of it, thus making the process "not-so-easy". Equivalent inferences could be obtained through a Bayesian inferential process based on the set that integrates the old and new data. Nevertheless, this is not always a real alternative, because it may be computationally very costly in terms of both time and resources. This work uses the dynamic characteristics of sequential Monte Carlo methods for "static" setups in the framework of longitudinal modelling scenarios. We used this methodology in real data through a random intercept model.

Linear inverse problem with range prior on correlations and its Variational Bayes inference

Ondrej Tichy, Vaclav Smidl*

Regularization in ill-conditioned linear inverse problem has significant impact on the resulting estimates. We consider a linear inverse model with truncated Gaussian prior on the solution and unknown covariance matrix in modified Choleski form. The prior on the covariance matrix is assumed to be Gaussian with

truncated support to expert given range on correlations. This model is motivated by estimation of mixture of radionuclides from gamma dose rate measurements under the prior knowledge of a range of their ratios. Since we aim at high dimensional problems, we use the Variational Bayes inference procedure to derive approximate inference of the model. The method is illustrated and compared on a simple example and on more realistic 6 hours long release of 3 radionuclides.

A covariate-dependent nonparametric model for clustering

Ilaria Bianchini, Raffaele Argiento, Alessandra Guglielmi*

Our aim is to develop a Bayesian nonparametric model that relaxes the hypothesis of exchangeability, due to the presence of covariates. Recently, this problem has been widely studied: the work of Barcella et al. (2015) provides a review on this topic.

Bayesian survival analysis to model plant resistance and tolerance to virus diseases

Elena Lázaro Hervás, Carmen Armero Cervera, Luis Rubio Miguélez*

Viruses constitute a major threat to large-scale production of crops worldwide producing important economical losses and undermining sustainability. We evaluate a new plant variety characterised by its genotype for resistance and tolerance to a specific virus through a comparison with other well-known varieties. The comparison is based on two independent Bayesian accelerated failure time models which assess resistance and tolerance survival times. Variables related to plant genotypes and virus biotype were considered as baseline covariates and error terms are assumed to follow a standard Gumbel distribution.

Bayesian models for describing and predicting the stochastic demand of emergency calls

Vittorio Nicoletta, Alessandra Guglielmi, Ettore Lanzarone, Valérie Belangér, Angel Ruiz*

Emergency Medical Service (EMS) system has the aim to provide immediate medical care to the population. EMS managers constantly deal with the problem of improving the system performance, in particular the response time to emergencies. A careful strategical and planning phase is a major prerequisite for the success of a system. The demand for emergency services is highly variable and stochasticity should not be neglected: in particular, would be useful to predict the future demand for interventions and the expected waiting time between to calls. In this paper, we propose two Bayesian models to represent the demand of interventions' evolution along with the waiting times and to predict it in future periods. Demands and waiting times are described by means of two generalized linear mixed models, whose posterior densities of parameters are obtained through Markov chain Monte Carlo simulation. Moreover, predictions are given in terms of their posterior predictive probabilities. To the best of our knowledge, no Bayesian approaches exist to deal with this problem. Results from the application to a relevant real case show the applicability of the proposed model in the practice and validate the approach.

Likelihood tempering in dynamic model averaging

*Jan Reichl, Kamil Dedecius**

We study the problem of online prediction with a set of candidate models using dynamic model averaging procedures. The standard assumptions of model averaging state, that the set of admissible models contains the true one(s), and that these models are continuously updated by valid data. However, both these assumptions are often violated in practice. The models used for online tasks are often more or less misspecified and the data corrupted (which is, mathematically, a demonstration of the same problem).

Bayesian consistency of the bivariate extremes spectral measure estimation

Khader Khadraoui, Pierre Ribereau*

The tail of a bivariate distribution function F in the max-domain of attraction of an extreme-value distribution function G may be approximated by that of its extreme value attractor. For some specific parametrization, the function G is characterized by a probability measure with expectation equal to 0.5, called the spectral measure, and two extreme-value index. This spectral measure determines the tail

dependence structure of F . We study the asymptotic behavior of a Bayesian nonparametric spectral measure estimation for a bivariate extreme value distribution. We first show that, under mild assumptions, the posterior is weakly consistent for any M -spline spectral measure H_0 on $[0, 1]$ with continuous and bounded Lebesgue density. When the number of knots is truncated, we show that under mild restrictions the posterior concentrates on the set of pseudotrue spectral measures.

Optimizing movement of cooperating pedestrians by exploiting floor-field model and Markov decision process

Vladimíra Sečkárová, Pavel Hrabák*

Optimizing movement of pedestrians is a topic of great importance, providing suggestions and solutions for situations like evacuation. Here, we assume that pedestrians choose their actions in order to move closer to the exit by exploiting the floor-field model or Markov decision process. We also allow them to cooperate and exchange their information about the state of the surrounding environment. The results of suggested approaches are compared on an example from evacuation time and amount of inhaled CO point of view.

Extension in TVP Model Shrinkage

*Angela Bitto**

We investigate shrinkage for time-varying parameter models based on the normal-gamma prior which has already been introduced for standard regression models. Our approach extends previous work in which the Bayesian Lasso prior has been considered. The Bayesian Lasso is a special case of the normal-gamma prior. We show how the normal-gamma prior can easily be extended to the time-varying parameter models and focus on inducing shrinkage on the square root of the variance of the prior of the error term in the non-centered state equation. We present both a univariate and a multivariate application. First we choose EU area inflation modelling based on the generalized Phillips curve, then we draw our attention to a multivariate time series with a time-varying covariance matrix and analyse DAX-30 data. Our findings suggest, that the normal-gamma prior bears advantages over the Bayesian Lasso prior in terms of statistical efficiency and performs significantly better when drawing attention to the predictive performance.

Identification of patient dependent parameters in a kinetic model of fluid and mass transfer during dialysis

Camilla Bianchi, Ettore Lanzarone, Giustina Casagrande, Maria Laura Costantino*

Hemodialysis (HD) is nowadays the most common therapy to treat renal insufficiency. However, despite the improvements made in the last years, HD is still associated with a non-negligible rate of morbidity, which could be reduced by means of appropriate treatment customization. Many multi-compartment models have been developed to describe the solute kinetics during HD, but they often refer to an average uremic patient. On the contrary, the clinical need for customization requires patient-specific multi-compartment models. In this work, while assuming that the customization can be obtained by means of patient-specific model parameters, we propose a tool to determine such parameters and to predict the single patient response to the treatment, in order to prevent intra-dialysis complications. We apply a Bayesian estimation approach, which has been implemented in STAN, to get the entire probability density function of the estimates (parameters and response to the treatment). As for the likelihood function of the model parameters given the observations, we consider a discretized version of a multi-compartment model, where the discretization is in terms of a Runge-Kutta method to guarantee the convergence.

Parallel Split-Merge MCMC for the HDP*Stefania Perego*, Debora Parisi**

We consider problems involving groups of data where each observation within a group is a draw from a mixture model and where it is desirable to share mixture components between groups. We assume that the number of mixture components is unknown a priori and is to be inferred from the data. The HDP model is suitable to solve this clustering problem. In particular, we have used this framework in the topic modeling, which is a model useful to discover the latent structure of topics in a corpus of documents. We have decided to use an algorithm that involves steps of Gibbs sampler and Metropolis Hastings. The Metropolis Hastings' steps propose new clusters, using merge moves to join two clusters and split moves to divide a cluster. Each cluster is characterized by the left and the right subcluster; new clusters are proposed basing on these subclusters. The code is designed to manage big data; indeed, to speed up the computation, we have decided to parallelize the code using OpenMP. We applied the HDP model to the problem of topic modeling, so we have used Categorical likelihood and Dirichlet prior on the latent parameters; nevertheless, the code is easily extendible to other problems and models.

A Bayesian spatio-temporal Markov switching model for the detection of influenza outbreaks*Rubén Amorós Salvador*, David Conesa, Antonio Lopez-Quilez, Miguel Angel Martinez-Beneito*

Influenza dispersion is related with climate variables and spreads person to person, which suggests a spatio-temporal evolution of the incidence. In this work we present a spatio-temporal extension of the Bayesian Markov switching model over the differentiated rates for the detection of influenza epidemic outbreaks presented in Martinez-Beneito et al. 2008. The variable of the Markov switching model represents the epidemic and non-epidemic states. The non-epidemic state differentiated rates only share a common mean for each time. The rates on the epidemic state are spatially and temporally related through Gaussian Markov random fields. This new proposal has been compared with the one in Martinez-Beneito et al. 2008 and offers better scores.

Fast, approximate MCMC for Bayesian analysis of large data sets: A design based approach*Matthias Kaeding**

We propose a fast approximate Metropolis- Hastings algorithm for large data sets embedded in a design based approach. The building block is one single subsample from the complete data set. The subsample is taken via balanced sampling: A sample is balanced if the means in the sample match the means of the complete data set. We balance on loglikelihood ratios. The complete data set does not have to be available for our algorithm. We develop several computationally and statistically efficient approximations of the Metropolis-Hastings update step using design based estimators, which are well-suited for this problem. Our simulation studies show that the approach works well and can lead to results which are almost identical to the use of the complete data set.

Randomization Inference and Bayesian Inference in Regression Discontinuity Design: An application to Italian University grants*Federica Licari**

The entrance in the university world results a radical change in the student's life. It requires financial resources that may obstacle the student's University career, prompting students from low income families to dropout. In order to give equal opportunity to achieve higher education to motivated students, every year Italian universities offer financial aids to a limited number of eligible students who apply for a grant. The assignment rules of these financial aids required that students are eligible and apply for a grant to receive a grant. In this work we focus on first year students, where the eligibility status is only based on an indicator of the economic family situation, i.e., the ISEE indicator. In order to obtain a financial aid for years subsequent the year of enrollment, students have also to realize a certain number of University Credits (CFUs) during the previous academic year, that is, they need to pass a certain number of University exams. Our purpose is to evaluate causal effects of university grants on two outcomes: dropout, a binary variable

equal to one for students who dropout by the end of the first academic year, and zero otherwise; and CFU, a binary variable equal to one for students who realize at least one CFU during the academic year, and zero otherwise. The assignment rules underlying the Italian university grants define a Fuzzy Regression Discontinuity (FRD) design. In FRD design the assignment mechanism depends on the realized value of a so called forcing variable falling above or below a pre-fixed threshold and by individual choices which imply the presence of Non-compliance. In order to do this we provide a Bayesian and Randomization inference approach to our case study.

Identification and Estimation of Principal Causal Effects in Randomized Experiments with Treatment Switching

*Emanuele Gramuglia**

In randomized clinical trials designed to evaluate the effect of a treatment on patients with advanced disease stages, treatment switching is often allowed for ethical reasons. Because the switching is a prognosis-related choice, identification and estimation of the effect of the actual receipt of the treatment becomes problematic. Although an Intention-To-Treat analysis, comparing groups formed by random assignment, is a reliable method of estimation in standard situations, in the presence of switching it does not provide causal information about the effect of the \textit{receipt} of the treatment, the effect of primary interest. Other existing methods in the literature try to reconstruct the ideal situation that would be observed if the switchers had not switched. Rather than focusing on reconstructing the \textit{a-priori} counterfactual outcome for the switchers, had they not switched, we propose to identify and estimate effects for (latent) subgroups of units according to their switching behaviour. The reference framework of the proposed method is the potential outcome approach. In order to estimate causal effects for subgroups of units not affected by treatment, we rely on the principal stratification approach (Frangakis and Rubin, 2002). To illustrate the proposed method and evaluate the maintained assumptions, we analyse a dataset from a randomized clinical trial on patients with asymptomatic HIV infection assigned to immediate (the active treatment) or deferred (the control treatment) Zidovudine (ZDV). We simulate different scenarios according to the relation between outcome and switching. The results, obtained through a full-Bayesian estimation approach, are promising and emphasize the high heterogeneity of the effects for different latent subgroups defined according to the switching behaviour.

Bayesian methods for microsimulation models

Consuelo Nava, Cinzia Carota, Ugo Colombino*

In this article we propose Bayesian methods for microsimulation models. In particular, we present Bayesian Multinomial Logit and Bayesian Multinomial Mixed Logit models applied to EUROMOD single female and single male data. Model performances are observed under Bayesian models and compared to standard methods applied to this problem.

Closed Form Bayesian Inferences for Binary Logistic Regression via Polynomial Expansions

*Kevin Dayaratna**

In many fields, logistic regression has become one of the most widely used tools in statistical modeling. Hierarchical Bayesian models are often used to incorporate individual-level heterogeneity such models. Unfortunately, incorporating individual-level heterogeneity often results in increased computational complexity in estimating these models, especially for large data sets involving high-dimensional parameter spaces. We present an alternative estimation technique for Bayesian binary logistic regression using polynomial expansions. We present a series of simulations as well as an application to child poverty and find that this approach significantly outperforms existing Bayesian estimation methods in terms of computing time. These gains are extremely useful for large data sets.

Reverse Random Walk Prior for Regression Discontinuity Designs*Federico Ricciardi*, Gianluca Baio*

The Regression Discontinuity Design (RDD) is a quasi-experimental design that estimates the causal effects of a treatment when its assignment is defined by a threshold value for a continuous assignment variable. The RDD assumes that subjects with measurements within a bandwidth around the threshold belong to a common population, so that the threshold can be seen as a randomizing device assigning treatment to those falling just above the threshold and withholding it from those who fall just below. Bandwidth selection represents a compelling decision for the RDD analysis, since there is a trade-off between its size and bias and precision of the estimates: if the bandwidth is small, the bias is generally low but so is precision, if the bandwidth is large the reverse is true. A number of methods to select the "optimal" bandwidth have been proposed in the literature, but their validity is questionable. We propose a methodology that, tackling the problem from an applied point of view, is able to consider the contribution of all the units in the sample, irrespectively of their distance from the threshold, but increasing their "strength" as this decreases. We use a Reverse Random Walk prior, modelling units increasingly distant from the threshold using different, although dependent, prior beliefs. We illustrate the validity of our methodology using a simulated experiment.

The Science behind the Magic: The Relation of Harry Potter "Sorting Hat Quiz" To Measures of Values and Personality - A Bayesian analysis*Lea Jakob*, Eduardo Garcia-Garzon, Hannes Jarke*

The J. K. Rowling's Harry Potter book series attracts the attention of psychology and anthropology researchers. Previous investigations have detected that individuals are able to relate personality traits with book elements (e.g., the Hogwarts School houses) and the portrayed character specific traits. This effect was found in previous research showing that individuals can—to some extent—recognize personality traits in both real-world and fictional settings. A careful examination of novels has yielded character's values to be of much greater importance for House membership than their actions (as proposed by previous research involving personality). This research aims to explore to what extent participant's membership relates to the Basic Human Values in juxtaposition to their personality traits. A Bayesian approach based on the use of Bayes Factors is used in this research.

Inference for Coupled SDE: Metropolis Algorithms via Density Tracking by Quadrature*Harish Bhat*, R. W. M. A. Madushani, Shagun Rawat*

We develop a Metropolis algorithm to perform Bayesian inference for models given by coupled stochastic differential equations. A key challenge in developing practical algorithms is the computation of the likelihood. We address this problem through the use of a fast method to track the probability density function of the stochastic differential equation. The method applies quadrature to the Chapman-Kolmogorov equation associated with a temporal discretization of the stochastic differential equation. The inference method can be adapted to scenarios in which we have multiple observations at one time, multiple time series, or observations with large and/or irregular temporal spacing. Computational tests show that the resulting Metropolis algorithm is capable of efficient inference for an electrical oscillator model.

M-quantile Bayesian regression in Small Area Estimation*Enrico Fabrizi, Nicola Salvati, Carlo Trivisano, Giovanni Riccardi**

Mixed effects models are widely used in SAE in order to estimate the effect for a particular area. However, such models depend on parametric and distributional assumptions as well as requiring specification of the random part of the model. An alternative approach to this regression estimation is the M-quantile regression. Chambers and Tzavidis (2006) applied the M-quantile regression as basis of their small area estimation method. A number of papers on M-quantile regression applied to SAE has been published since 2006 (Tzavidis et al., 2010; Fabrizi et al., 2014), concerning theoretical developments, extension to non-

linear models and various applications. But it has not been proposed any Bayesian approach. The main objective of this work is to propose a Bayesian M-quantile regression model specifying an asymmetric likelihood function based on Generalized Asymmetric Least Informative distribution (Bianchi et al., 2015). Furthermore, the Bayesian M-quantile regression will be applied to SAE in order to obtain small area estimators and their standard errors.

Bayesian autoregressive models for waiting times of recurrent events

Giorgio Paulon, Maria De Iorio, Alessandra Guglielmi*

We propose an autoregressive Bayesian semi-parametric model for the waiting times between recurrent events with covariates. Time-dependency is taken into account through an autoregressive model, but the parameters are a sample from a Dirichlet process, thus inducing clustering of the individuals in the sample. Covariates may be included in this framework. The model proposed is within the class of Dirichlet process mixtures. We build a Gibbs sampler to perform full posterior inference from this model; the code has been implemented in the Julia language. We illustrate the model through an application to recurrent hospitalizations of cancer patients.

Multiple membership model for data integration using the hierarchical beta process

Takoua Jendoubi, Korbinian Strimmer*

The parallel acquisition of high-throughput datasets from the genome, epigenome, metabolome, proteome, and transcriptome of the same biological system is now common in molecular biology. The integrated analysis of these data is promising to enhance the understanding of biological functions and pathways. Since metabolites, enzymes, or proteins can interfere in multiple biological processes, one particular challenge is to address multiple membership of investigated variables in functional modules or pathways. Here, we present an approach to data integration using the hierarchical beta process. This method has three advantages: i) by using a Bayesian nonparametric approach the number of identified clusters and functional modules is not restricted a priori but learned adaptively from the data, ii) the Beta process enables to model multiple memberships, and iii) using a hierarchical model allows to share information across data types and clusters. Using both synthetic as well as metabolomic and transcriptomic data we compare the suggested approach based on the hierarchical beta process with previously suggested related models for data integration using hierarchical Dirichlet process.

SOCIAL PROGRAM

Florence tour

Sunday, June 19

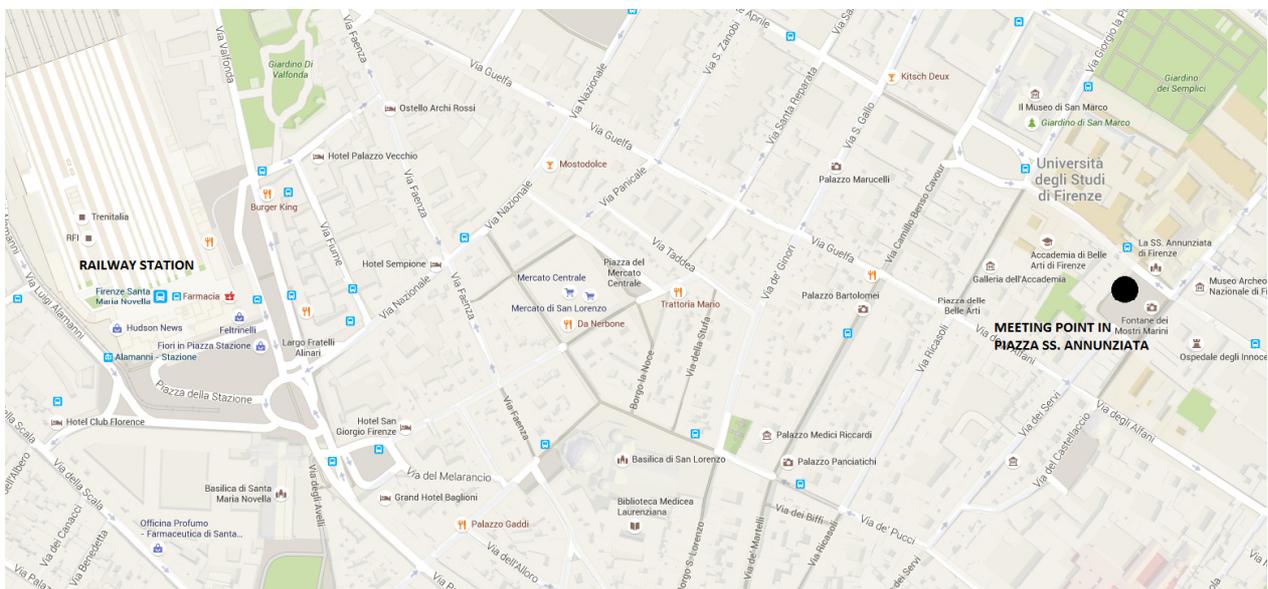


Meeting point

Piazza Ss. Annunziata
at 16:45

End of tour

Piazza Duomo
at about 20:00



Social dinner

Monday, June 20

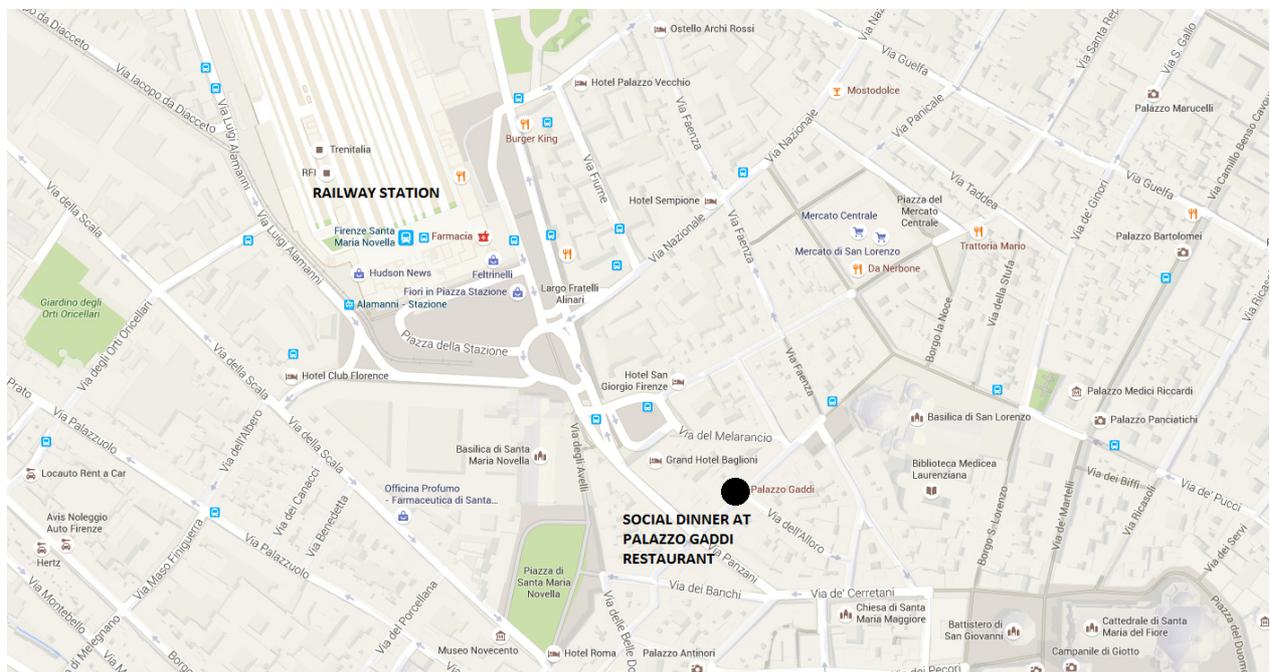


Social dinner at Palazzo Gaddi

Meeting
at 20 p.m.
in front
of the main entrance

in
Via del Giglio, 11

www.ristorantegaddifirenze.it



CONFERENCE VENUES

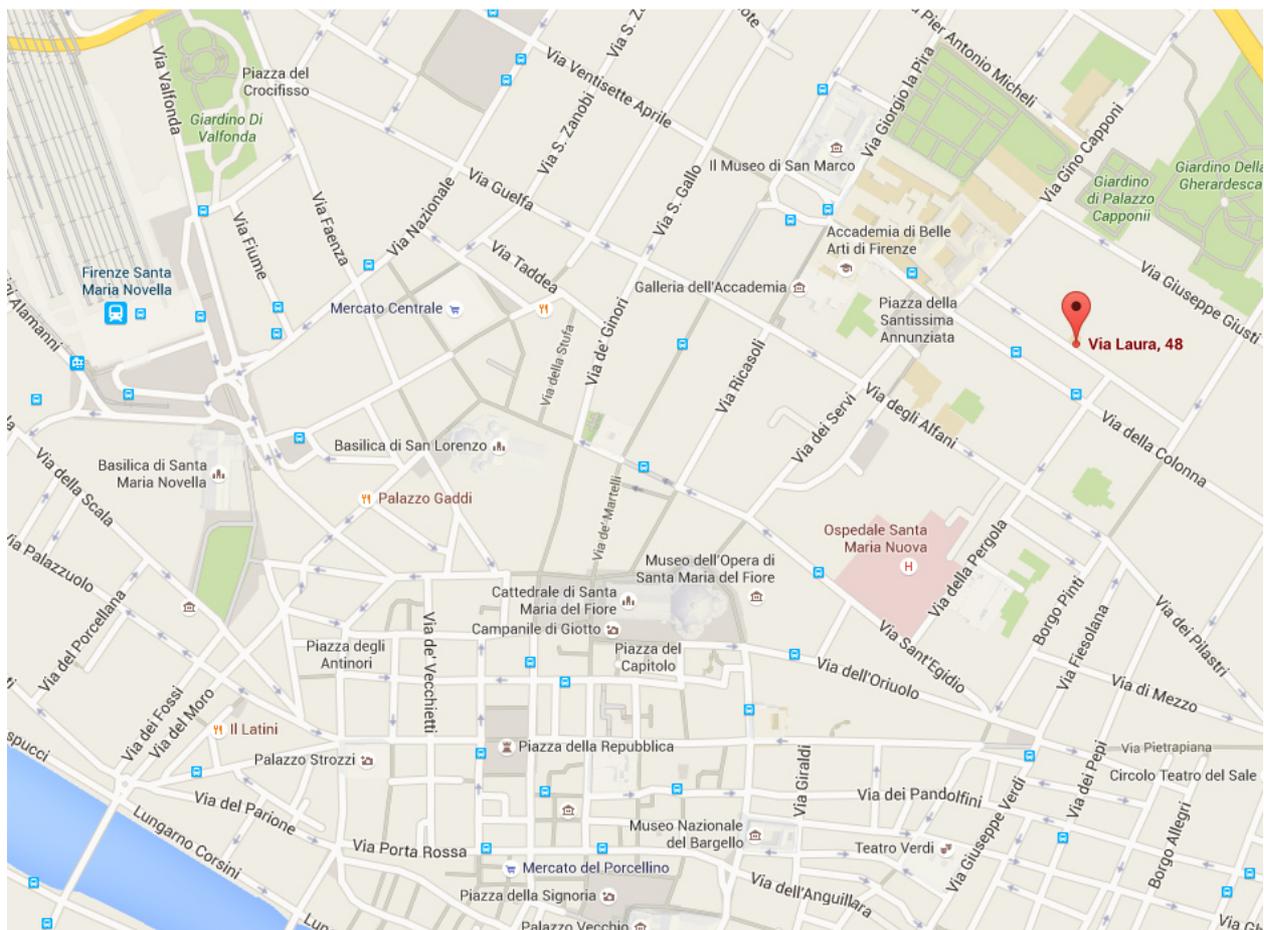
University of Florence

Monday, June 20

Via Laura 48, 50121 Florence, Italy.

First Floor

Registration: room 3; Conference: room 5

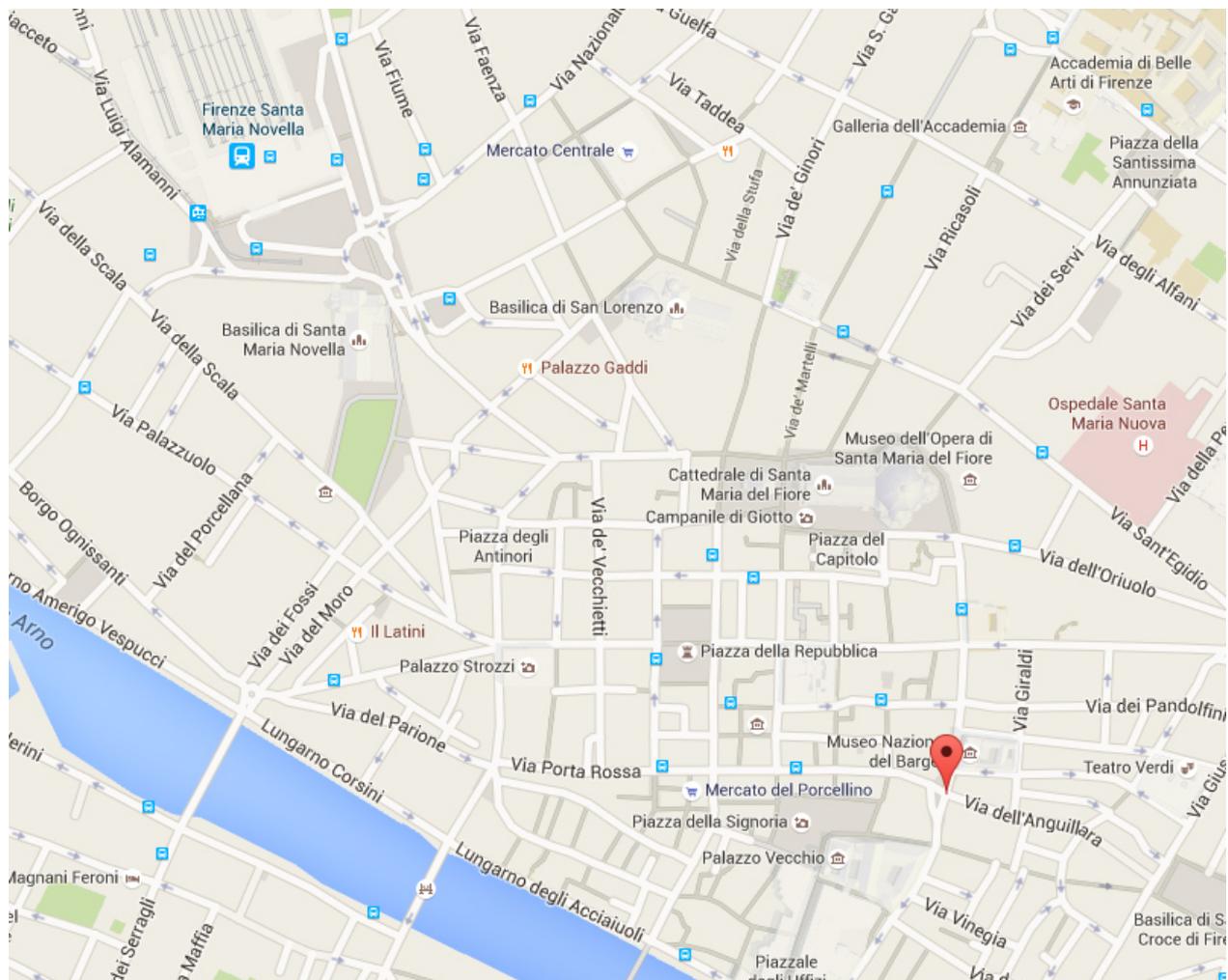


From the Railway Station of Firenze Santa Maria Novella you can take the bus n. 6 and get off at the P.za San Marco stop, in Via Cesare Battisti. Cross the road and walk beside the church on your left. Walk along under the arcade and at the end, turn left in Via Gino Capponi and then take the first on the right, that is Via Laura. Go straight on up to number 48 which is on the left.

Sala della Musica in the “ex tribunale” building

Tuesday, June 21

Piazza di San Firenze, 50122 Florence, Italy.



Location is close to the Cathedral and can be easily reached by walk.

USEFUL INFORMATION

Oral presentations

Maximum duration of each oral presentation is 15 minutes. After the presentation, there are 5 minutes for the discussant's comments and the questions from the audience. Speakers can either use their own laptop or the one provided in the room. In the second case, we suggest them to check the presentation before the beginning of the section; both power point and pdf presentation are accepted.

Poster presentations

Please ensure that your poster is no greater than standard A0 size, and is orientated in portrait mode. Poster dimensions should therefore be no larger than 84cm (33 inches) x 118cm (47 inches). Panels for the poster will be available in the room few hours before the poster session. We suggest to hang up the poster before the beginning of the presentation. Scotch tape will be available in the room.

Social Dinner

People with special dietary requirements (vegetarian, allergies, ...) can score by writing to: baysm2016@promoest.com

WiFi connections

Florence is covered by the town free wifi service. You can register for free at: <http://firezewifi.provincia.fi.it/firenze/login.php>

In addition, at the University of Florence, EDUROAM is also available.

Suggestions for lunches (not included)

We suggest the following places to have lunch close to the conference venue:

Close to University (for Monday 20)

1. Pizza al taglio Pippirilli; Borgo Pinti 61
2. Bar Cadillac; Via degli Alfani 57
3. Ristorante Tipico Trattoria Acquacotta; Via dei Pilastrini 51/R
4. Ristorante Tipico Trattoria Accadi; Borgo Pinti 56/R
5. Pizzeria le Campane; Borgo la Croce 87/R

Close to Sala della Musica (for Tuesday 21)

1. All'antico Vinaio (very good, also take away); Via De' Neri 65/R 74 76/R
2. Gelateria Vivoli; Via dell'Isola delle Stinche 7/R
3. Bar Chiaroscuro; Via del Corso, 36/R
4. Trattoria tipica Pallottino; Via Isola delle Stinche 1/R

Google maps

A map with all the places of BAYSM 2016 is available on Google maps:

<https://www.google.com/maps/d/viewer?mid=15riH3fgMpWSWxo6H36cQ02PmkDE>

An additional map with the suggested hotels is also available:

<https://www.google.com/maps/d/viewer?mid=1HTBdmNQwKhaa8UF95wfidMWWOFc>

An additional map with the suggested places for lunch is also available:

https://www.google.com/maps/d/viewer?mid=1WYvJLj62_xPFC2Lb1EcU7CXOPok

Weather in Florence

Updated information are always available at

<http://www.ilmeteo.it/meteo/Firenze>

Registration desk

The desk will be available on Monday June 20th in Via Laura from 8:30 am.

Acknowledgments

BAYSM 2016 is organized by



and co-organized by



**Università Commerciale
Luigi Bocconi**



UNIVERSITÀ
DEGLI STUDI
FIRENZE
DISIA
DIPARTIMENTO DI
STATISTICA, INFORMATICA,
APPLICAZIONI "G. PARENTI"

We thank our main sponsor:



We thank for patronizing
and supporting:



INTERNATIONAL SOCIETY FOR BAYESIAN ANALYSIS

We thank our invited plenary and keynote speakers (*Alessandra Guglielmi, Fabrizia Mealli, Peter Müller, Steven Scott, and Marina Vannucci*) and our invited session discussants (*Daniela Cocchi, Fabrizio Corradi, Emanuela Dreassi, Brunero Liseo, Antonio Pievatolo, and Fabrizio Ruggeri*) for contributing to the success of the conference.

Awards

Two talk awards and a poster award will be assigned during the conference closure (Tuesday 21, starting from 19:00).

BAYSM community

Information about BAYSM community and about the next conference editions can be found at the official website:

www.baysm.org

or on the official Facebook group:

<https://www.facebook.com/groups/326987910816120/>

Information

During the conference, in case of need, you may contact out organizing secretariat for any question:

Michela Tizzani

m.tizzani@promoest.com

+39 338 9965674

Chapter submission for the Springer Proceedings Book

Participants who are interested in publishing a chapter can submit an updated version of their contribution by replacing the previously submitted contribution of the springer website:

<https://ocs.springer.com/prom/en/home/BAYSM2016>

Manuscripts should follow the rules reported on the conference website:

<http://web.mi.imati.cnr.it/conferences/BAYSM2016/keydates.html>

The deadline for the submission is July 31, 2016.

After the submission the manuscript will be evaluated by two independent reviewers.

We ask to the authors who want to keep the already submitted manuscript for this review process to communicate this decision by mail to baysm2016@mi.imati.cnr.it

The absence of an uploaded document after the conference without such mail will be considered as a conference contribution not resubmitted for the book.

