ISBA World Meeting

July 1-7, 2024
San Giobbe
Economics Campus
Ca' Foscari University
of Venice

Tuesday 2nd, 2024

Schedule of the Day

9:00 am - 9:30 am	Opening remarks
9:30 am - 10:30 am	Foundation Lecture: Marina Vannucci
10:30 am - 11:00 am	*Coffee break*
11:00 am - 12:00 pm	Foundation Lecture: Dani Gamerman
12:00 pm- 1:30 pm	*Lunch break*
1:30 pm - 3:00 pm	Multi-track session 1
3:00 pm - 3:30 pm	*Coffee break*
3:30 pm - 4:30 pm	Foundation Lecture: Kerrie Mengersen
4:30 pm - 5:00 pm	*Unsupervised refreshment break*
5:00 pm - 6:30 pm	Multi-track session 2
7:00 pm - 9:00 pm	*Welcome reception*

9:00 am Opening remarks

- Antonio Marcomini (Deputy-Rector), Ca' Foscari University of Venice
- Roberto Casarin (Chair, Local Organizing Committee), Ca' Foscari University of Venice
- Aad van der Vaart (ISBA President), Delft Institute of Applied Mathematics
- Sinead Williamson (Chair, Scientific Committee), Apple

9:30 am Foundation Lecture: Marina Vannucci

Room	live at	Aula Magna “Guido Cazzavillan”
	streaming	Rooms 4A - 5A - 9A - 10A
	youtube	YouTube streaming

Speaker

Marina Vannucci

Noah Harding Professor of Statistics,
Rice University

Chair: Peter Müller

Bayesian Models for Complex Problems with Applications to Science

Abstract: In this talk I will present novel Bayesian methodologies that have been developed for specific applications to science. I will first motivate the talk by introducing data from different scientific studies, including neuroscience and microbiome studies. I will highlight common goals of these studies, such as variable selection, estimation of dependence structures and of latent states underlying the

observed data. I will then describe statistical methodologies specifically tailored to these scientific problems, focusing the discussion on aspects of covariate-dependent modeling, variable selection priors, graph estimation, and strategies for efficient posterior inference. Applications will include assessing seizure risk in patients with epilepsy, estimating dynamic brain connectivity states based on neuroimaging data, and integrative modeling of microbiome data.

11:00 am Foundation Lecture: Dani Gamerman

Room		
	live at	Aula Magna “Guido Cazzavillan”
	streaming at	Rooms 4A - 5A - 9A - 10A
	youtube	YouTube streaming

Speaker

Dani Gamerman
Emeritus
Universidade Federal do Rio de Janeiro

Chair: Mark Steel

Exact spatial analysis via process augmentation

Abstract: The advance of Bayesian statistics has grown considerably in the last decades leading to models with increasing complexity. This growth was accompanied by the need for approximations and we became used to them. Some of them are computational for extraction of information from the posterior distribution but other approximations are caused by the assumed inability to handle the models as they were posed.

↑ back to the schedule ↑

One important example is provided by point patterns (PP). This is one of the most commonly encountered data structure in spatial analysis, where locations of occurrences of a certain phenomenon of interest are observed. Estimation of the intensity of occurrence is the primary interest in the many usual settings. Nevertheless, the likelihood of the non-parametric intensity is not available analytically. Approximations are usually applied, inducing biases and losses in all likelihood-based inferential procedures. This inefficiency is inherited by all models that contain PP components. Further complexity is brought by the use of Gaussian processes (GP) to induce smoothness over the intensity function.

This talk will address these complications and will propose exact procedures to remedy the situation. These procedures are based on augmentation with latent processes and, more importantly, avoid model approximations.

This idea opens up a framework to handle a variety of models involving PP components without compromising their integrity. Examples include the use of PP regression with space-varying coefficients, geostatistics with sampling preferentiality, analysis of presence-only data in Ecology and nonstationary/discontinuous intensity functions. Different augmentation processes were required for the above contexts. Results from tests with synthetic data, comparison against alternatives and applications to real data are presented.

Issues associated with cost of computation with GP, parallelization and software are briefly addressed.

↑ back to the schedule ↑

1.30 pm Multi-track session 1

For invited sessions, the organizer will chair unless otherwise stated

<table>
<tr><td>Session A</td><td>Invited session: Bayesian Causal Inference for Health-Related Applications and Policy Intervention
Organizer: Marina Vannuci</td></tr>
<tr><td>Room</td><td>Magna</td></tr>
<tr><td>Francesca Dominici</td><td>Leveraging Data Science for Informed Climate and Environmental Policy</td></tr>
<tr><td colspan="2">Abstract: In this talk, I will present a comprehensive overview of data science methods, focusing on Bayesian analysis, causal inference, and machine learning, and their applications in shaping climate and environmental policy. The foundation of this research is built upon the analysis of an unprecedented data platform, comprising over 500 million observations related to the health experiences of more than 95% of the US population aged 65 years and older. This data is linked to critical factors, including air pollution exposure, climate exposures (such as heat, wildfire exposure, and tropical cyclones), and other pertinent confounders. Key highlights of the talk include the introduction of innovative Bayesian methods designed to characterize the heterogeneity of causal effects. By doing so, we can better understand how diverse populations are affected uniquely by environmental influences, enabling more targeted and effective policy interventions. An essential focus of the talk will be the differentiation of health impacts stemming from distinct sources of air pollution, such as fossil fuel power plants versus smoke. This understanding is vital in formulating precise policies to tackle the sources that contribute most to adverse health outcomes. The outcomes of this research hold great promise for informing evidence-based decision-making in climate and environmental policy.</td></tr>
<tr><td>Antonio Linero</td><td>Bayesian Nonparametric Estimation of Heterogeneous Causal Mediation Effects</td></tr>
<tr><td colspan="2">Abstract: The causal inference literature has increasingly recognized that explicitly targeting treatment effect heterogeneity can lead to improved scientific</td></tr>
</table>

understanding and policy recommendations. Towards the same ends, studying the causal pathway connecting the treatment to the outcome can be also useful. In this talk, we address these problems in the context of causal mediation analysis. We introduce a varying coefficient model based on Bayesian additive regression trees to identify and regularize heterogeneous causal mediation effects; analogously to linear structural equation models (LSEMs), these effects correspond to covariate-dependent products of coefficients. We show that, even on large datasets with few covariates, LSEMs can produce highly unstable estimates of the conditional average direct and indirect effects, while our Bayesian causal mediation forests model produces estimates that are stable. We find that our approach is conservative, with effect estimates "shrunk towards homogeneity." We examine the salient properties of our method using both data from the Medical Expenditure Panel Survey and empirically-grounded simulated data. Finally, we show how our model can be combined with posterior summarization strategies to identify interesting subgroups and interpret the model fit.

Fabrizia Mealli

Bayesian Inference for Causal Mechanisms

Abstract: In causal inference studies, interest often lies in understanding the mechanisms or channels through which a treatment affects an outcome: this may provide valuable insights for developing effective interventions and identifying potential areas for improvement. One such mechanism may arise from the inherent interconnectedness of units that may interact in clusters (e.g., schools, villages) or networks (e.g., social networks, networks of friendships). This raises the problem of interference and spillovers between observations, that is, an observational unit's outcome depends not only on its treatment but also the treatment assigned to other units. In the potential outcome approach to causal inference, principal stratification is a framework to deal with such issues. Bayesian inference is particularly suitable for developing models and methods in such settings where typically accounting for the several different sources of uncertainty is a difficult task. We will show recent advances in this area with applications to the evaluation of environmental health interventions.

↑ back to the schedule ↑

Session B	Invited session: Data Integration Organizer: David Rossell
Room	6A
Paul Rognon	**Selecting a model by integrating data: the power of non-exchangeable penalization**

Abstract: The ability to recover the true set of relevant features is fundamentally limited by sparsity: how the number of available samples compares to the number of irrelevant features and the size of effects. When working with multiple datasets, this sparsity constraint can be known to vary across sources of information. One example is when the datasets to integrate have vastly different dimensions (e.g. clinical vs genomic data). Another example is when one dataset helps tell apart relevant features from irrelevant ones in another dataset. We show how having differentiated penalization for blocks of features that are differentially sparse helps model recovery in penalized linear regression. We further propose a concrete model selection procedure based on non-exchangeable penalization that realizes those benefits and outperforms standard penalization. Finally, we describe how to tune penalties from either block sizes or an estimation of the proportions of irrelevant features.

Roberta de Vito	**Bayesian Multi-study Non Negative Matrix Factorization**

Abstract: Adopting methods to integrate multiple studies is crucial to achieving knowledge and information in epidemiological and biological data. This integration relies on two key challenges: 1- the common amount of information from all the studies and 2- the study-specific source from individual studies. The Bayesian Multi-Study Factor model (De Vito et al., 2019) achieves these two challenges and handles multiple studies. However, these methods relies on normality assumptions and continuous type data. We develop a new class of sparse Bayesian Multi-study model by adopting a non-negative matrix factorization approach that can handle counts in the data matrix. We explore different types of sparse-inducing priors to detect the latent dimension. A user-defined prior for covariate coefficient accounts for population structure and other characteristics. We assess our method's performance by different simulation settings, which clarify the benefit of our model, resulting in better accuracy and precision. We illustrate the advantages of our method through a mutational signature application.

Niccoló Anceschi

Bayesian Joint Additive Factor Regression for Multi-omics Integration

Abstract: In precision medicine, it is common to gather data from multiple modalities to characterize different aspects of a patient across biological layers. It is of interest to infer within- and across-source dependence, while combining information to improve the prediction of health outcomes. With multi-omics data, the signal-to-noise ratio can vary substantially across modalities, motivating more structured statistical tools beyond standard late and early fusion. This challenge comes with the need to preserve interpretability, allow the selection of relevant biomarkers, and obtain accurate uncertainty quantification for the predicted outcomes. While these challenges can be addressed within a Bayesian framework, state-of-the-art factor analysis formulations for multi-omics data rely on loose modeling assumptions. We propose a novel joint additive factor regression model having a structured additive design, accounting for shared and view-specific components. Moreover, we ensure identifiability via a novel dependent cumulative shrinkage process prior. We provide an efficient implementation via a partially collapsed Gibbs sampler and extend our approach to allow flexible predictor and outcome distributions.

Giovanni Parmigiani

Discussant

Session C	Invited session: Recent Advances in Bayesian Probabilistic Numerical Methods Organizer: Francois-Xavier Briol Chair: Jeremias Knoblauch
Room	3A

Chris Oates

Richardson Extrapolation meets Multi-Fidelity Modelling

Abstract: For over a century, extrapolation methods have provided a powerful tool to improve the convergence order of a numerical method. However, these tools are not well-suited to modern computer codes, where multiple continua are discretised and convergence orders are not easily analysed. To address this challenge we present a probabilistic perspective on Richardson extrapolation, a point of view that unifies classical extrapolation methods with modern multi-fidelity modelling, and handles uncertain convergence orders by allowing these to be statistically estimated. The approach is developed using Gaussian processes, leading to Gauss–Richardson Extrapolation (GRE). Conditions are established under which extrapolation using the conditional mean achieves a polynomial (or even an exponential) speed-up compared to the original numerical method. Further, the probabilistic formulation unlocks the possibility of experimental design, casting the selection of fidelities as a continuous optimisation problem which can then be (approximately) solved. A case-study involving a computational cardiac model demonstrates that practical gains in accuracy can be achieved using the GRE method.

Jon Cockayne

Learning to Solve Related Linear Systems

Abstract: The problem of solving multiple related linear systems occurs in numerous fields, such as in optimisation or sampling problems involving computer models and when solving nonlinear PDEs. This talk will explore a Gaussian-process based approach to accelerating solution of these linear systems, by constructing a model that predicts both sensible initial iterates and preconditioners for iterative linear solvers. The approach has strong ties to probabilistic linear solvers and approaches in transfer learning. We will also present some theoretical guarantees concerning how much of an improvement can be expected when using this approach, and demonstrate it numerically on test problems from numerical optimisation.

Masha Naslidnyk

Optimally-weighted Estimators of the Maximum Mean Discrepancy for Likelihood-Free Inference

Abstract: Likelihood-free inference methods typically make use of a distance between simulated and real data. A common example is the maximum mean discrepancy (MMD), which has previously been used for approximate Bayesian computation, minimum distance estimation, generalised Bayesian inference, and within the nonparametric learning framework. The MMD is commonly estimated at a root-m rate, where m is the number of simulated samples. This can lead to significant computational challenges since a large m is required to obtain an accurate estimate, which is crucial for parameter estimation. In this paper, we propose a novel estimator for the MMD with significantly improved sample complexity. The estimator is particularly well suited for computationally expensive smooth simulators with low- to mid-dimensional inputs. This claim is supported through both theoretical results and an extensive simulation study on benchmark simulators.

Session D	Invited session: Bayesian Experimental Design for Sample Size Determination Organizer: Federico Castelletti Chair: Guido Consonni
Room	2A
Kevin Wilson	**Assurance for sample size determination in diagnostic accuracy studies and cluster randomised trials**

Abstract: In clinical trials sample sizes are chosen as the smallest value to achieve a given power for the primary analysis, e.g., a hypothesis test, with a pre-specified significance level. Power is the conditional probability of rejecting the null hypothesis given a treatment effect and estimates for nuisance parameters. This can lead to under- or over-powered trials if estimates are far from their true values. We describe a Bayesian alternative to power, assurance, which provides the unconditional probability the treatment will be found to be effective in the trial. We apply assurance to diagnostic accuracy studies, choosing the sample size to provide accurate interval estimates for the test sensitivity and specificity, and cluster randomised trials, where the analysis for the treatment effect uses the posterior distribution resulting from the usual linear mixed effect model.

Stefania Gubbiotti	**Sample size determination based on Bayes risk in the presence of alternative design priors**

Abstract: The size of a clinical trial is typically selected to reach a given power for the hypothesis test on the main parameter of interest. Since this approach relies on the specification of a design value, it implies locally optimal solutions. To overcome this problem, several recent contributions advocate the use of the "probability of success" of an experiment that is based on a design prior distribution. Unfortunately, as discussed in a review paper by Kunzmann et al (2021), there is no univocal definition of the "probability of success". In this paper, thanks to a decision-theoretic framework we introduce a unifying formulation and we define sample size criteria based on the Bayes risk of a decision function, i.e. the expected value of the risk function with respect to a design prior. In the presence of uncertainty in the design scenario, several alternative priors can be considered, so that their impact on the resulting optimal sample sizes can be assessed.

Haiyan Zheng	**Sample size calculation for small population clinical trials incorporating historical data**

Abstract: Many disease conditions are so rare that the target population has only less than a hundred patients. Design and analysis of clinical trials in rare diseases are challenging, because it is clearly infeasible to enrol enough patients to achieve an adequate frequentist power. We introduce a fully Bayesian approach that leverages relevant historical information from multiple sources in a robust prior. This approach accounts for pairwise (in)commensurability between parameters that underpin the historical and new clinical trials. Specifically, we use commensurate predictive priors for borrowing of information from historical studies, and further place Gamma mixture priors on the precisions to account for potential data incommensurability. Closed-form sample size formulae are derived to ensure that the new trial has a specified chance of correctly deciding whether a new treatment is superior to or not better than the control by some clinically relevant difference. The prior mixture weights are interpretable, as these can be directly indicative of sample size saving in the new trial. We illustrate the application by revisiting the MYPAN trial, a randomised controlled trial conducted in rare and severe inflammatory blood vessel disease.

Session E	Invited session: Recent advances in Bayesian clustering for complex data Organizers: Sally Paganin and Garritt Page.
Room	10A
Alex Dombowsky	**Product Centered Dirichlet Processes for Dependent Clustering**

Abstract: While there is an immense literature on Bayesian methods for clustering, the multiview case has received little attention. This problem focuses on obtaining distinct but statistically dependent clusterings in a common set of entities for different data types. For example, clustering patients into subgroups with subgroup membership varying according to the domain of the patient variables. A challenge is how to model the across-view dependence between the partitions of patients into subgroups. The complexities of the partition space make standard methods to model dependence, such as correlation, infeasible. In this article, we propose CLustering with Independence Centering (CLIC), a clustering prior that uses a single parameter to

↑ back to the schedule ↑

explicitly model dependence between clusterings across views. CLIC is induced by the product centered Dirichlet process (PCDP), a novel hierarchical prior that bridges between independent and equivalent partitions. We show appealing theoretic properties, provide a finite approximation and prove its accuracy, present a marginal Gibbs sampler for posterior computation, and derive closed form expressions for the marginal and joint partition distributions for the CLIC model. On synthetic data and in an application to epidemiology, CLIC accurately characterizes view-specific partitions while providing inference on the dependence level

Gertraud Malsiner-Walli	**CliPS - Clustering in the parameter space**

Abstract: Each mixture model has a representation as a marked point process where the parameters of the component-distributions are the points and the component weights are their associated marks. Stephens (2000, Annals of Statistics) proposed this view to develop a Markov birth-death process for inference in the context of finite mixtures where the number of components is unknown. Frühwirth-Schnatter (2006, Chapter 1) highlighted how the label-invariance of the point-process representation helps to identify a mixture model. We investigate the use of the (marked) point process representation of a mixture model for post-processing the posterior draws of a mixture model in a clustering context. Mapping the component-distribution parameters into the parameter space provides an alternative view on the fitted model. Inspecting the results in the parameter space allows to determine a suitable number of clusters, assess the suitability of a solution for clustering, in particular when using a lower-dimensional functional to characterize how data clusters differs, and obtain an identified model. In particular we outline how clustering in the parameter space can be used for any posterior draws from a mixture model and how strategies might differ depending on if the number of data clusters are assumed to be known or to be estimated using an overfitting mixture model. This approach is easily applicable to mixture models with arbitrary component densities and can be used for finite as well as infinite mixtures.

Cecilia Balocchi	**Improving uncertainty quantification in Bayesian cluster analysis**

Abstract: The Bayesian approach to clustering is often appreciated for its ability to provide uncertainty in the partition structure. However, summarizing the posterior distribution over the clustering structure can be challenging. Wade and Ghahramani

↑ back to the schedule ↑

(2018) proposed to summarize the posterior samples using a single optimal clustering estimate, which minimizes the expected posterior Variation of Information (VI). In instances where the posterior distribution is multimodal, it can be beneficial to summarize the posterior samples using multiple clustering estimates, each corresponding to a different part of the space of partitions that receives substantial posterior mass. In this work, we propose to find such clustering estimates by approximating the posterior distribution in a VI-based Wasserstein distance sense. An interesting byproduct is that this problem can be seen as using the k-means algorithm to divide the posterior samples into different groups, each represented by one of the clustering estimates. Using both synthetic and real datasets, we show that our proposal helps to improve the understanding of uncertainty, particularly when the data clusters are not well separated, or when the employed model is misspecified.

Raffaele Argiento	**Discussant**

Session F	Contributed session: Advances in Econometrics Chair: Hedibert Lopes
Room	9A
Luis Gruber	**Dynamic Sparsity in Factor Stochastic Volatility Models**

Abstract: Appropriately selecting the number of factors in a factor model is a challenging task, and even more so if the number of factors changes over time. In this paper, we estimate a factor stochastic volatility (FSV) model through Markov chain Monte Carlo (MCMC) methods and then post-process the draws from the posterior to achieve sparsity in the factor loadings matrix. Recasting the FSV model as a homoskedastic factor model with time-varying loadings enables us to sparsify the loadings for each point in time and across MCMC draws. This enables us to back out the posterior distribution of the number of factors over time. We illustrate in simulations that our techniques accurately detect the true number of factors and apply the model to US stock market returns.

↑ back to the schedule ↑

Carson McKee

Bayesian Nonparametric Change-Point Detection for Multivariate Time Series

Abstract: This work is motivated by multivariate macroeconomic time series. Such data exhibits sharp structural breaks, or change-points, driven by events such as recessions or financial market crashes. Inferring the locations of these change-points is crucial to helping us to understand what drives macroeconomic events. In modelling such data, we have two key considerations to make. The first concerns how we model the change-point locations and the second is how we model the data conditional on these locations. In this work, we model the change-points with a Product Partition Model (PPM). PPMs assume that the data may be partitioned into contiguous segments with the beginning of each segment representing a change-point. Then, conditional on the partition, the data in each segment is independent of the data in other segments. Thus, a PPM jointly models both the number and locations of the change-points. We model the probability of a change-point occurring at a given time as a function of the previous observations, thus facilitating learning on what is driving the change-points. Additionally, we assume a Vector Autoregressive (VAR) model specification over the data in each segment, which is independent of the other segments. However, the number of parameters in a VAR scales quadratically with the dimension of the observations and they are therefore highly prone to overfitting, especially when few data points are available. We address this by placing shrinkage priors over the VAR parameters, resulting in the prior and likelihood within each segment being non-conjugate. Currently, there is a gap in computational methods for non-conjugate specifications. To that end, we develop a novel MCMC scheme which is applicable to non-conjugate change-point PPMs. We demonstrate our model on simulated and real macroeconomic datasets and show that it detects change-points on dates which correspond to well-known macroeconomic events.

Dawid Bernaciak

A loss discounting framework for model averaging and selection in time series models

Abstract: We introduce a Loss Discounting Framework for model and forecast combination which generalises and combines Bayesian model synthesis and generalized Bayes methodologies. We use a loss function to score the performance of different models and introduce a multilevel discounting scheme which allows a flexible specification of the dynamics of the model weights.

↑ back to the schedule ↑

This novel and simple model combination approach can be easily applied to large scale model averaging/selection, can handle unusual features such as sudden regime changes, and can be tailored to different forecasting problems. We compare our method to both established methodologies and state of the art methods for a number of macroeconomic forecasting examples. We find that the proposed method offers an attractive, computationally efficient alternative to the benchmark methodologies and often outperforms more complex techniques..

↑ back to the schedule ↑

3:30pm Foundation Lecture: Kerrie Mengersen

Room	live at	Aula Magna “Guido Cazzavillan”
	streaming	Rooms 4A - 5A - 9A - 10A
	youtube	YouTube streaming

Speaker

Kerrie Mengersen
Distinguished Professor,
Queensland University of Technology

Chair: Christian Robert

Being Bayesian in the 21st Century

Abstract: The Bayesian paradigm has become entrenched in Statistical Science. It is also a standard tool in many applied and social sciences. So where to from here? What is left for Bayesians to do in the 21st Century? In this presentation, I will briefly reflect on the past few decades of our discipline and focus on some of the directions that I see Bayesian statistics travelling in the next few decades. These will include new methodological and computational pathways, dealing with different data sources, expansion into other disciplines and mergers with other fields such as product development and AI.

5.00 pm Multi-track session 2

For invited sessions, the organizer will chair unless otherwise stated

Session A	Invited panel discussion: Critical Role of Bayesian Thinking in the Health Care Decision-making Organizer: Satrajit Roychoudhury
Room	10A
Panelists	**Kannan Natarajan, Natalia Muhlemann, Monika Jelizarow, Kelley Kidwell, Mark Whitlock**
Session B	Invited session: Bridging the Gap: Bayesian Statistics and Machine Learning for Enhanced Data Modeling Organizer: Clara Grazian
Room	5A
Ethan Goan	**Tools of Bayesian Analysis for Explainable Deep Learning**

Abstract: With deep learning providing unprecedented predictive capabilities for challenging perception tasks there is great interest in developing these models within a more principled probabilistic framework. Despite the interest in deploying a fully Bayesian approach to fitting these models, the increase in model size and computational complexity of modern networks quickly makes this infeasible. This talk will show how tools of Bayesian analysis can be applied to existing point-estimate networks within frameworks currently developed for explainable AI systems. We will show how adopting a probabilistic approach for explainable AI can allow for visualisation of features contributing to the strong predictive performance of existing deep-learning models whilst also highlighting features that give rise to predictive uncertainty.

Qian Jin

Generalized Partial Least Square in Deep Neural Network

Abstract: While deep learning has shown exceptional performance in many applications, the model's mathematical understanding, model designing, and model interpretation are still new areas. Combining the two cultures of deep and statistical learning can provide insight into model interpretation and high-dimensional data applications. This work focuses on combing deep learning with generalized partial least square estimation. In particular, Bilodeau et al. (2015) proposed a generalized regression with orthogonal components (GROC) model, which is more flexible than the standard partial least square (PLS), because it may involve a more complex structure of dependence and the use of the generalized additive model (GAM) instead of linear regression. We propose a deep-GROC (DGROC) model, which allows for different measures of dependence to be used and shows a high prediction accuracy. Hyperparameter selection and transfer learning in the training loop are included to boost model performance. The Bayesian neural network is used to get the Bayesian DGROC model. With the benefit of the Bayesian Neural Network, we can have a probabilistic generalized Partial Least Square model, which provides a new framework for Partial Least Square model inference. The superiority of the proposed method is demonstrated on simulations and real datasets, which show that our method achieves competitive performance compared to GROC, PLS and traditional Neural Networks.

Andrew Zammit-Mangion

Neural Bayes Estimators for Irregular Spatial Data using Graph Neural Networks

Abstract: Neural Bayes estimators are neural networks that approximate Bayes estimators. They are fast, likelihood-free, and amenable to fast bootstrap-based uncertainty quantification. Currently, neural Bayes estimators for spatial models are only available for gridded data. The estimators are also conditional on the sample locations, and need to be re-trained whenever the sample locations change; this renders them impractical in many applications. In this work, we employ graph neural networks to tackle the important problem of spatial-model-parameter estimation from arbitrary sampling locations. Our architecture leads to substantial computational benefits since training of the neural Bayes estimator now only needs to be performed once for a given spatial model, and can be used with any number or arrangement of sampling locations. We illustrate the methodology on a range of spatial models, including Gaussian processes and max-stable processes for spatial extremes, which

have an intractable likelihood function.
This is joint work with Matthew Sainsbury-Dale, Jordan Richards, and Raphaël Huser.

Session C	Invited session: Bayesian Analysis of Network Data. Organizer: Sirio Legramanti
Room	Magna
Martina Contisciani	**Latent network models to account for noisy, multiply reported social network data**

Abstract: Social network data are often constructed by incorporating reports from multiple individuals. However, it is not obvious how to reconcile discordant responses from individuals. There may be particular risks with multiply reported data if people's responses reflect normative expectations—such as an expectation of balanced, reciprocal relationships. In this talk, I will present a probabilistic model that incorporates ties reported by multiple individuals to estimate the unobserved network structure. In addition to estimating a parameter for each reporter that is related to their tendency of over- or under-reporting relationships, the model explicitly incorporates a term for 'mutuality', the tendency to report ties in both directions involving the same alter. Our model's algorithmic implementation is based on variational inference, which makes it efficient and scalable to large systems. We apply our model to data from a Nicaraguan community collected with a roster-based design and 75 Indian villages collected with a name-generator design. We observe strong evidence of 'mutuality' in both datasets, and find that this value varies by relationship type. Consequently, our model estimates networks with reciprocity values that are substantially different than those resulting from standard deterministic aggregation approaches, demonstrating the need to consider such issues when gathering, constructing, and analysing survey-based network data.

Nial Friel	**Modelling pairwise comparison data using network analysis**

Abstract: Pairwise comparison data are common in many fields including the social sciences and sports analytics. A canonical example is in a sporting context is where the data consists of the results when pairs of players compete against one another. Our objective here is to develop a statistical model for this data setting using tools from network analysis. An innovative feature of our approach is the ability to rank nodes (or players) in the network.

Federico Pavone	**Phylogenetic latent position models**

Abstract: We consider the problem of learning the underlying structure responsible for the connectivity patterns in the human brain. We analyze a population of networks representing the connections between brain regions for a set of subjects. These networks are characterized by a multiresolution organization of the nodes, responsible for the connectivity. We propose a phylogenetic latent position model, where the node latent positions are the realization of Brownian motions over a phylogenetic tree. The model reveals a tree organization of the brain regions coherent with known hemisphere and lobe partitions. Such a result uncovers new interesting possible clusterings of the brain regions at different levels of resolution.

Session D	Invited session: Bayes in the Classroom: Examples, Tools, and Recommendations Organizer: Mine Dogucu Chair: Becky Tang
Room	3A
Bertil Wegmann	**Design of content and computer examination in Bayesian courses at bachelor's and master's level**

Abstract: It is crucial to design suitable learning activities for the students in a course to facilitate the fulfillment of the course's intended learning objectives. In courses on Bayesian statistics the content and learning objectives usually differ depending on the level of education. Hence, the learning activities therefore need to be adjusted to the certain level. In this talk, I will present some ideas from my experiences about designing content and computer examination in Bayesian statistics courses at the bachelor's and master's level. Computer lab work is common to both types of courses, where students actively engage in building their own Bayesian models or solutions to specific problems using R and Stan. This learning activity is also central in preparation for the individual computer exam, to which the students are allowed to bring their own computer lab code and where the learning objectives are examined in a secured exam environment.

Mine Dogucu

Training Instructors to Advance Bayesian Thinking in STEM

Abstract: In recent decades there has been growth in use of Bayesian methods in research setting in academia and industry. There has also been growth in teaching of Bayesian methods in the classroom, however this growth has been at a much slower pace than in research. Prior educational research has shown that most of the training at the undergraduate level takes place in statistics, math, and data science programs. However, training in Bayesian methods is needed across all scientific disciplines (e.g., psychology, economy, life sciences). In this talk, I will give an overview of a training program called Bayes BATS, designed for instructors who teach in a wide range of disciplines in institutions across the United States. The goal of the program is to enable instructors to introduce Bayesian methods in their curriculum. We will share activities and resources from the program that are available open-access.

Patricia Toledo

Introducing Frequentist and Bayesian Methods in Parallel in an Undergraduate Economics Statistics Course

Abstract: Economic statistics courses typically focus on preparing undergraduate students to analyze data using exclusively frequentist methods. I adapt the structure of an existing undergraduate economic statistics course which originally emphasized the use of the frequentist approach to introduce the Bayesian approach in parallel. Since this is the first statistic course for many students, the review of concepts emphasizes intuition and applied exercises over complex formulas and analytical proofs. Students are introduced to Bayesian modelling by learning the Beta-Binomial model for proportions and the Normal-Normal model for means using the R package "bayesrules". By comparing frequentist and Bayesian approaches students would improve their understanding of concepts that are usually misunderstood in the frequentists case, such as p-value, confidence intervals, and implicit assumptions used to derive the distribution of estimators in finite samples. Furthermore, the format encourages students to develop critical thinking skills when they analyze economic data. This course format will be implemented in fall 2023. Students will complete an exit survey to provide feedback.

Session E

Invited session: Optimal Transport and Bayesian Learning
Organizer: Marta Catalano

Room	9A
Long Nguyen	**Dendrogram of mixing measures: Learning latent hierarchy and model selection for finite mixture models**

Abstract: We present a post-processing method to summarize and select mixture models via the hierarchical clustering tree (dendrogram) of overfitted latent mixing measures. Our proposed method bridges agglomerative hierarchical clustering and mixture modelling. The dendrogram's construction is derived from the theory of convergence of the latent mixing measures, and as a result one can both consistently estimate the number of mixing components and obtain the pointwise optimal convergence rate for parameter estimation from the tree, even when the model parameters are only weakly identifiable. In theory, it explicates the optimal selection of the number of clusters in hierarchical clustering. In practice, the dendrogram reveals more meaningful and robust information on the hierarchy of subpopulations compared to traditional ways of summarizing mixture models. Several simulation studies are carried out to support our theory. We also illustrate the methodology with an application to single-cell RNA sequence analysis. This work is joint with Dat Do, Linh Do, Scott McKinley and Jonathan Terhost.

Hugo Lavenant	**Quantifying the merging of opinion in Bayesian nonparametrics via optimal transport**

Abstract: We investigate the question of merging of opinions: starting from different priors and observing the same data, will there be convergence between the posterior distributions as the sample size increases? And if yes, is it possible to quantify at which rate? This question is delicate in Bayesian Nonparametrics because prior and posterior distributions are typically infinite dimensional, making already the definition of a distance between priors and posteriors a challenge. We concentrate on normalized completely random measures and propose, both a priori and a posteriori: (i) an analysis of their identifiability, (ii) the definition of an optimal transport distance based on their Lévy intensities and (iii) techniques to conduct posterior analyses to study merging of opinions obtaining both finite sample and asymptotic behavior of our distance. This helps to quantify the impact of the prior: whereas we expect small variations of the prior to lead to a small changes for the results of inference, one also would like important changes to the prior to make an impact and lead to different outcomes. This is joint work with Marta Catalano.

Ricardo Baptista	**Conditional sampling via block-triangular transport maps**

Abstract: We present an optimal transport framework for conditional sampling of probability measures. Conditional sampling is a fundamental task of solving Bayesian inverse problems and generative modeling. Optimal transport provides a flexible methodology to sample target distributions appearing in these problems by constructing a deterministic coupling that maps samples from a reference distribution (e.g., a standard Gaussian) to the desired target. To extend these tools for conditional sampling, we first develop the theoretical foundations of block triangular transport in a Banach space setting by drawing connections between monotone triangular maps and optimal transport. To learn these block triangular maps, we will then present a computational approach, called monotone generative adversarial networks (MGANs). Our algorithm uses only samples from the underlying joint probability measure and is hence likelihood-free, making it applicable to inverse problems where likelihood evaluations are inaccessible or computationally prohibitive. We will demonstrate the accuracy of MGAN for sampling the posterior distribution in Bayesian inverse problems involving ordinary and partial differential equations, and probabilistic image in-painting.

Session F	Invited session: Pitfalls in Model-based Clustering (and how to mitigate them) Organizer: Fernando A. Quintana Chair: Raffaele Argiento
Room	6A
Alessandra Guglielmi	**Attractive BNP mixtures among populations**

Abstract: We propose an "attractive" mixture model to cluster data across populations. The density within each population is modeled as a mixture of Gaussian densities with a shot-noise Cox process as the mixing measure. We assume a hierarchical framework as in the Hierarchical Dirichlet process mixtures, so that the different populations may share some clusters. However, differently from nonparametric priors in the existing literature, the sharing of clusters does not necessarily imply that the same Gaussian component is present in different

populations. In our model, "similar atoms" drive the mixing measures in different populations. The attractiveness property of the shot-noise Cox process is used to cluster together similar atoms and produces a very flexible model capable of borrowing strength. We illustrate the associated MCMC and show the characteristics of our model through simulated data.

Garritt Page	**Informed Bayesian Finite Mixture Models via Asymmetric Dirichlet Priors**

Abstract: Finite mixture models are a model-based approach to carry out probabilistic clustering. A recent focus in the model-based clustering literature is the difference between the number of components in a mixture model and the number of clusters. The number of clusters is more relevant from a practical standpoint, but to date prior distributions have only been developed for the number of mixture components. To this end, we develop a methodology that permits eliciting prior information directly on the number of clusters in an intuitive way. This is done by employing an asymmetric Dirichlet distribution as a prior on the weights of a finite mixture. Further, a penalized complexity motivated prior is employed as a prior for the Dirichlet shape parameter. We illustrate the ease in which prior information can be elicited based on our construction and the flexibility of the resulting induced prior on the number of clusters. We also demonstrate the utility of our approach using numerical experiments and real-world datasets.

Sally Paganin	**Informed Random Partition Models with Temporal Dependence**

Abstract: Model-based clustering is a powerful tool that is often used to discover hidden structure in data by grouping observational units. Recently, clustering methods have been developed that allows incorporating an initial partition informed by expert opinion. Then, using some loss metric, partitions far from the initial one are down weighted, i.e. they are assigned reduced probabilities. However, the current methods are fairly rigid in how expert opinion can be incorporated in the sense that the entire partition is either up or down weighted. It is often the case however that practitioners have clear prior information about only on a small subset of the partition. In this work, we detail an approach that is not based on partition penalties but considers individual allocation probabilities for each unit which locally weight prior information. We illustrate the substantial gains in prior specification flexibility in

simulation studies and an application to a dataset concerning spatio-temporal evolution of PM10 measurements in Germany.

Session G	Contributed session: Advances in Causal inference Chair: Sameer Deshpande
Room	4A
Dafne Zorzetto	**Causal Inference in Multivariate Outcomes: A Bayesian Factor Regression Model**

Abstract: In the context of causal inference, the investigation of causal effects for multivariate potential outcomes has not yet been extensively explored. This is due to the inherent challenge of capturing the overall causal effect of a treatment on correlated outcomes and understanding the heterogeneity of the causal effect across outcomes. However, this research question is of critical importance in relevant fields such as epigenetics to understand the causal effect of a tumor or treatment on gene expression, or environmental epidemiology to investigate the causal link between air pollution regulation and the concentration of different pollutants or between air pollution and hospitalizations for various diseases. To achieve that, we exploit the ability of Bayesian factor analysis to identify the latent treatment-specific factors that capture the causal effects within correlated multivariate outcomes. We propose a methodology that (i) introduces novel causal estimands in the general framework of multivariate outcomes, and (ii) defines a multi-treatment Bayesian factor regression model that allows us to identify and characterize the causal latent effects. The performance of the method is illustrated through simulations and real data applications in genomics and environmental epidemiology.

Carlos Cinelli	**Causally Sound Priors for Binary Experiments**

Abstract: We introduce the BREASE framework for the Bayesian analysis of randomized controlled trials with a binary treatment and a binary outcome. Approaching the problem from a causal inference perspective, we propose parameterizing the likelihood in terms of the baseline risk, efficacy, and side effects of the treatment, along with a flexible, yet intuitive and tractable jointly independent beta prior distribution on these parameters, which we show to be a generalization of the Dirichlet prior for the joint distribution of potential outcomes. Our approach has a

number of desirable characteristics when compared to current mainstream alternatives: (i) it naturally induces prior dependence between expected outcomes in the treatment and control groups; (ii) as the baseline risk, efficacy and side effects are quantities inherently familiar to clinicians, the hyperparameters of the prior are directly interpretable, thus facilitating the elicitation of prior knowledge and sensitivity analysis; and (iii) it admits analytical formulae for the marginal likelihood, Bayes factor, and other posterior quantities, as well as exact posterior sampling via simulation, in cases where traditional MCMC fails. Empirical examples demonstrate the utility of our methods for estimation, hypothesis testing, and sensitivity analysis of treatment effects.

Hedibert Freitas Lopes	**Cutoff-aware BART for Estimating Heterogeneous Treatment Effects in Regression Discontinuity Designs**

Abstract: This paper proposes a modification of the Bayesian Causal Forest algorithm (Hahn et al., 2020) - itself an extension of the BART algorithm (Chipman et al., 2010) - which uses a novel regression tree prior that incorporates the unique structure of regression discontinuity designs. Specifically, we add constraints to the tree splitting process that assure overlap within a narrow band surrounding the running variable cutoff value (where the treatment effect is identified). We show that unmodified BART and BCF models estimate RDD treatment effects poorly, while our modified model accurately recovers treatment effects at the cutoff. At the same time, our modified model retains the inherent flexibility of all BART-based models, allowing it to effectively explore heterogeneous treatment effects. Simulation studies indicate that the new approach improves upon traditional local polynomial regression on both simple and complex data generating processes in terms of estimation error, coverage, and interval length for both average and conditional average treatment effects. We illustrate the new method by analyzing data studied originally by Lindo et al. (2010) to estimate the effect of academic probation on university students' GPA; we find an average increase of 0.15 in GPA for students whose previous semester GPA lied just below the probation cutoff.

Stefano Peluso	**Bayesian learning of network structures from interventional experimental data**

Abstract: Directed acyclic graphs provide an effective framework for learning causal relationships among variables given multivariate observations. Under pure observational data, directed acyclic graphs encoding the same conditional independencies cannot be distinguished and are collected into Markov equivalence classes. In many contexts, however, observational measurements are supplemented by interventional data that improve directed acyclic graph identifiability and enhance causal effect estimation. We propose a Bayesian framework for multivariate data partially generated after stochastic interventions. To this end, we introduce an effective prior elicitation procedure leading to a closed-form expression for the directed acyclic graph marginal likelihood and guaranteeing score equivalence among directed acyclic graphs that are Markov equivalent post intervention. Under the Gaussian setting, we show, in terms of posterior ratio consistency, that the true network will be asymptotically recovered, regardless of the specific distribution of the intervened variables and of the relative asymptotic dominance between observational and interventional measurements. We validate our theoretical results via simulation and we implement a Markov chain Monte Carlo sampler for posterior inference on the space of directed acyclic graphs on both synthetic and biological protein expression data.