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

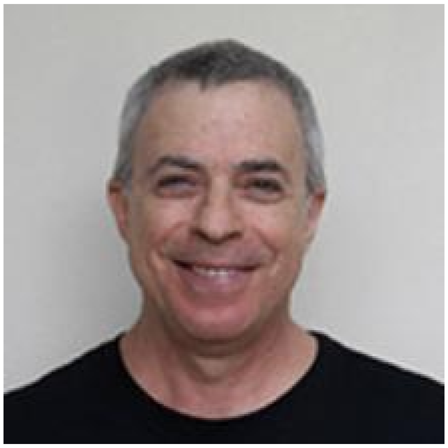
Chair: Peter Müller

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.

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.

↑ back to the schedule ↑

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 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.

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.

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.

↑ back to the schedule ↑

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.

↑ back to the schedule ↑

ISBA World Meeting

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

Wednesday 3rd, 2024

Schedule of the Day

9:00 am - 10:30 am	Multi-track session 3
10:30 am - 11:00 am	*Coffee break*
11:00 am - 12:00 pm	Foundation Lecture: Andrew Gelman
12:00 pm - 1pm	R.V. Ramamoorthi memory sharing
12:00 pm - 1:30 pm	*Lunch break*
1:30 pm - 3:00 pm	Multi-track session 4
3:00 pm - 3:30 pm	*Coffee break*
3:30 pm - 4:30 pm	Keynote Lecture: Alexandra Schmidt
5:00 pm - 6:30 pm	Bayesian Analysis Business Meeting
5:00 pm - 7:30 pm	Poster Session 1

9.00 am Multi-track session 3

For invited sessions, the organizer will chair unless otherwise stated

Session A	Bayesian Analysis Invited Session: Model Uncertainty in High-dimensional Non-Gaussian Regression Models Organizer: Mark Steel
Room	Magna
Jim Griffin	**Expressing and visualizing model uncertainty in Bayesian variable selection using Cartesian credible sets**

Abstract: Modern regression applications can involve hundreds or thousands of variables which motivates the use of variable selection methods. Bayesian variable selection defines a posterior distribution on the possible subsets of the variables (which are usually termed models) to express uncertainty about which variables are strongly linked to the response. This can be used to provide Bayesian model averaged predictions or inference, and to understand the relative importance of different variables. However, there has been little work on meaningful representations of this uncertainty beyond first order summaries. We introduce Cartesian credible sets to address this gap. The elements of these sets are formed by concatenating sub-models defined on each block of a partition of the variables. Investigating these sub-models allow us to understand whether the models in the Cartesian credible set always/never/sometimes include a particular variable or group of variables and provide a useful summary of model uncertainty. We introduce methods to find these sets that emphasize ease of understanding. The potential of the method is illustrated on regression problems with both small and large numbers of variables.

David Rossell	**Approximate Laplace Approximations and extensions**

Abstract: A main bottleneck when performing Bayesian model selection in non-Gaussian settings is computing the marginal likelihood for candidate models. Approximate Laplace Approximations (ALA) are approximate computation methods that reduce the cost to essentially that of Gaussian regression, and carry theoretical guarantees about consistent model recovery. We discuss the basic ALA formulation and some recent variations. For example, by suitably centering a Taylor expansion one can consistently recover the true model at least at the same rates as when using exact calculations, at a much cheaper cost. Another option is combining ALA with sequential Monte Carlo.

Gregor Zens	**Model uncertainty in Random Regression Single-Parameter models**

Abstract: A new class of models, named random regression single parameter (RRSP) models, is introduced. They are based on standard single parameter models (such as Poisson, Binomial, Bernoulli, Erlang etc.), but add a normal linear regression framework on a transformation of the single parameter. We allow for model uncertainty regarding the covariates included in the regression. The RRSP class typically accommodates extra dispersion in the data (an exception is discussed) and has clear advantages for deriving theoretical properties and for designing computational procedures. Posterior existence under a convenient and popular prior is formally characterised and we propose an efficient Markov chain Monte Carlo algorithm for Bayesian model averaging in RRSP models. Simulation results suggest that the framework provides accurate results that are robust to some degree of misspecification. The methodology is successfully applied to bilateral migration flow data from OECD countries and measles vaccination coverage data from villages in Ethiopia.

Merlise Clyde	**Discussant**

Session B	Invited session: Bayesian Deep Learning Organizer: Vincent Fortuin
Room	6A
Yingzhen Li	**Calibrating Transformers via Sparse Gaussian Processes**

Abstract: Transformer models have become standard building blocks for popular machine learning applications such as large-scale generative models including large language models (LLMs). Extending Transformer's success to safety-critical domains requires calibrated uncertainty estimation which remains under-explored. To address this, we propose Sparse Gaussian Process attention (SGPA), which performs Bayesian inference directly in the output space of multi-head attention blocks (MHAs) in transformer to calibrate its uncertainty. It draws connection between dot-product based attentions with sparse Gaussian processes (SGP), which are then used to approximate the posterior processes of MHA outputs. Empirically, on a suite of prediction tasks on text, images and graphs, SGPA-based Transformers achieve competitive predictive accuracy, while noticeably improving both in-distribution calibration and out-of-distribution robustness and detection. Future works include adapt SGPA to auto-regressive prediction and improve its scalability in order to apply it to LLM.

Alexander Immer	**Advances in Bayesian Neural Model Selection**

Abstract: Choosing optimal hyperparameters for deep learning can be highly expensive due to trial-and-error procedures and required expertise. Conceptually, a Bayesian approach to hyperparameter selection could help overcome such issues because it can rely on gradient-based optimization and does not require a held-out validation set. However, such an approach requires estimation and differentiation of the marginal likelihood, which is inherently intractable. In my talk, I discuss recent advances on Laplace approximations that provide efficient estimates and enable optimizing hyperparameters with stochastic gradients just like neural network weights. Further, I demonstrate successful applications of Bayesian model selection and discuss shortcomings of current algorithms.

Mark van der Wilk	**Bayesian Model Selection has an Opinion on Causal Assumptions**

Abstract: It is well known that for flexible model classes, causal assumptions (e.g. X causes Y or Y causes X) cannot be distinguished based on model fit (e.g. maximum likelihood) alone, since the best fit in both directions will be equally good. We will discuss that the introduction of Bayesian priors is enough for Bayesian Model Selection to prefer one causal direction over another, based on which causal direction gives the simplest explanation of the data. Whether this preference aligns with the true causal direction, depends (of course) on the appropriateness of the prior. However, we show that state-of-the-art results can be obtained across a wide range of causality benchmark datasets by creating a single sensible model, and applying a common variational inference technique. Finally, we will discuss how these Bayesian ideas can be scaled up using deep meta-learning. This work was done in collaboration with PhD candidate Anish Dhir (Imperial College London).

↑ back to the schedule ↑

<table>
<tr><td>Stephan Mandt</td><td>**Discussant**</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Session C</td><td>Invited session: Bayesian Industrial Statistics
Organizers: David Rios Insua and Refik Soyer
Chair: Refik Soyer</td></tr>
<tr><td>Room</td><td>3A</td></tr>
<tr><td>Daniel García-Rasines</td><td>**Personalised pricing through adversarial risk analysis**</td></tr>
<tr><td colspan="2">Abstract: Pricing stands out as one of the most critical tasks a company faces, particularly in today’s digital economy. Similar to many other business decision-making problems, pricing unfolds in a highly competitive and uncertain environment. Traditional analyses in this area have heavily relied on game theory and its variants. However, an important drawback of these approaches is their reliance on common knowledge assumptions, which are hardly tenable in competitive business domains. This paper introduces an innovative personalized pricing framework designed to assist decision-makers in undertaking pricing decisions amidst competition, considering both buyer's and competitors' preferences. Our approach (i) establishes a coherent framework for modeling competition mitigating common knowledge assumptions; (ii) proposes a principled method to forecast competitors' pricing and customer's purchasing decisions, acknowledging substantial business uncertainties; and, (iii) encourages structured thinking about the competitors' problems, thus enriching the solution process. To show these properties, in addition to presenting a general pricing template, we outline two specifications -- one from the retail domain and a more intricate one from the pension fund domain.</td></tr>
<tr><td>Jose M. Camacho</td><td>**A Cybersecurity Risk Analysis Framework for Systems with Artificial Intelligence Components**</td></tr>
<tr><td colspan="2">Abstract: The introduction of the European Union Artificial Intelligence Act, the NIST Artificial Intelligence Risk Management Framework, and related norms demands a better understanding and implementation of novel risk analysis approaches to evaluate systems with Artificial Intelligence components. This paper provides a cybersecurity risk analysis framework that can help assessing such systems. We use an illustrative example concerning automated driving systems.</td></tr>
</table>

Pablo G. Arce

A Bayesian framework for product launching

Abstract: In the world of utility-driven marketing, each company acts as an adversary to other contenders, who all have competing interests, particularly for niche products. These interests are perceived by the target buyer (or consumer), who wants to maximize the utility of the product for themselves. A major challenge for companies is that at the time of product launch, there is typically endogenous uncertainty with regards to their product, despite testing, potentially leading to market share loss. Furthermore, each company generally has little, and imperfect, information regarding the launch strategy and product quality of competing brands. However, delaying the launch of a niche product can have the downside of losing the `first-mover advantage'. From the buyer's perspective, along with the price of the product, they need to make their decision based on noisy signals they receive regarding the product's quality as promoted. This paper proposes a broad Bayesian framework to support the product launching decision by a company, in the presence of multiple competitors, and multiple buyers, who all have multiple constraints, with both sides having the objective to maximize their expected utilities. We apply the proposed framework to illustrate applications in two case studies.

Refik Soyer

Discussant

Session D

Invited session: Trees, Ensembles, and Bayesian Prediction
Organizer: Li Ma

Room 7A

Veronika Rockova

Adaptive Bayesian Predictive Inference

Abstract: Bayesian predictive inference provides a coherent description of entire predictive uncertainty through predictive distributions. We examine several widely used sparsity priors from the predictive (as opposed to estimation) inference viewpoint. Our context is estimating a predictive distribution of a high-dimensional Gaussian observation with a known variance but an unknown sparse mean under the Kullback-Leibler loss. First, we show that LASSO (Laplace) priors are incapable of achieving rate-optimal performance. This new result contributes to the literature on negative findings about Bayesian LASSO posteriors. However, deploying the Laplace prior inside the Spike-and-Slab framework (for example with the Spike-and-Slab LASSO prior), rate-minimax performance can be attained with properly tuned parameters (depending on the sparsity level sn). We highlight the discrepancy between prior calibration for the purpose of prediction and estimation. Going further, we investigate popular hierarchical priors which are known to attain adaptive rate-minimax performance for estimation. Whether or not they are rate-minimax also for predictive inference has, until now, been unclear. We answer affirmatively by showing that hierarchical Spike-and-Slab priors are adaptive and attain the minimax rate without the knowledge of sn. This is the first rate-adaptive result in the literature on predictive density estimation in sparse setups. This finding celebrates benefits of a fully Bayesian inference.

Hengrui Luo	**Shapes of the tree-induced partitions and their properties**

Abstract: In this talk, we'll start by illustrating how decision trees essentially draw multidimensional boundaries to divide the input space. Following this idea, we move into the setting of regressions for tensor inputs and outputs, showcasing the complex geometric shapes that emerge and the new challenges they bring to the tree-based models. To navigate these complexities, we introduce our innovative tensor-on-tensor tree regression approach, designed to adeptly and efficiently manage this multi-dimensional geometric partitioning. Concluding, we'll take a step back to reflect on the fundamental geometric principles underpinning tree-induced partitions and ponder future research avenues in this fascinating intersection of geometry and tree-based models.

Linxi Liu	**Posterior concentration rates for unsupervised trees and forests**

Abstract: Tree-based methods are popular nonparametric tools for capturing spatial heterogeneity and making predictions in multivariate problems. In unsupervised learning, trees and their ensembles have also been applied to a wide range of statistical inference tasks, such as multi-resolution sketching of distributional variations, localization of high-density regions, and design of efficient data compression schemes. In this talk, we will focus on the density estimation problem---a fundamental one in unsupervised learning. We consider the optional Pólya tree prior or its variations on individual trees. First we show that Bayesian density trees can achieve minimax (up to a logarithmic term) convergence over the anisotropic Besov class, which implies that tree based methods can adapt to spatially inhomogeneous features of the underlying density function, and can achieve fast convergence as the dimension increases. We will also introduce a novel Bayesian model for forests, and show that for a class of Hölder continuous functions, such type of density forests can achieve faster convergence than trees. The convergence rate is adaptive in the sense that to achieve such a rate we do not need any prior knowledge of the smoothness level of the density. The Bayesian framework naturally provides a stochastic search algorithm over either the tree space or the forest one. For both Bayesian density trees and forests, we swill provide several numerical results to illustrate their performance in the moderately high-dimensional case.

Session E	Invited session: Sample Size Determination and Assessment of Design Operating Characteristics in Bayesian Clinical Trials Organizer: Shirin Golchi Chair: Silvia Calderazzo
Room	9A
Sudipto Banerjee	**Bayesian Sample Size Determination from a Machine Learning Perspective**

Abstract: Determining the sample size to meet the inferential objectives of a study is of central importance in statistical science. There is an extensive collection of methods addressing this problem from diverse perspectives.. The Bayesian paradigm, in particular, has attracted noticeable attention and includes different perspectives for sample size determination. While traditional Bayesian methods formulate sample size determination as a decision problem that optimizes a given utility function, practical experimental settings may require a more flexible approach based upon simulating analysis and design objectives. In this talk, I will present a Bayesian framework for simulation-based sample size determination that will be geared toward some contemporary problems in machine learning. The basic idea is to eliminate the null hypothesis from the model formulation and build probabilistic learning models for the design and analysis stage. I will draw some parallels with the classical approaches, which arise as special cases, as well as with alternate Bayesian approaches embodied within this framework.

Annette Kopp-Schneider	**Frequentist operating characteristics of Bayesian clinical trials**

Abstract: Borrowing of information from historical or external data to inform inference in a current trial is an expanding field in the era of precision medicine, where trials are often performed in small patient cohorts for practical or ethical reasons. Many approaches for borrowing from external data have been proposed. Even though these methods are mainly based on Bayesian approaches by incorporating external information into the prior for the current analysis, frequentist operating characteristics of the analysis strategy are of interest. In particular, type I error and power at a prespecified point alternative are in the focus. It is well-known that borrowing from external information may lead to the alteration of type I error rate. When evaluating the frequentist operating characteristics of a hypothesis test, both type I error rate and power have to be considered simultaneously. We propose to evaluate type I error rate of the test with borrowing from external data and calibrate the test without borrowing to this type I error rate. On this basis, a fair comparison of power between the test with and without borrowing is achieved. We show that no power gains are possible in one-sided one-arm and two-arm hybrid control trials with normal endpoint, a finding that had been proven in general before. We illustrate that the Empirical Bayes power prior approach that dynamically borrows information according to the similarity of current and external data avoids the exorbitant type I error inflation occurring when borrowing with fixed power prior. We have however observed that the power in the hybrid control two-arm trial doesn't reach the power of the test calibrated to borrowing, and power losses increase when considering unconditional power.

Shirin Golchi	**Bayesian Design Operating Characteristics in Clinical Trials**

Abstract: Bayesian inference and the use of posterior or posterior predictive probabilities for decision making have become increasingly popular in clinical trials. The current approach toward Bayesian clinical trials is a hybrid Bayesian-frequentist approach where the design and decision criteria are assessed with respect to frequentist operating characteristics such as power and type I error rate conditioning on a given set of parameters. These operating characteristics are commonly obtained via simulation studies. The utility of Bayesian measures, such as "assurance", that incorporate uncertainty about model parameters in estimating the probabilities of various decisions in trials has been demonstrated recently. However, the

computational burden remains an obstacle toward wider use of such criteria. This talk is focused on methodology which utilizes large sample theory of the posterior distribution to define simple parametric models for the sampling distribution of the Bayesian test statistics, i.e., posterior tail probabilities. The approach toward the assessment of conditional and marginal operating characteristics and sample size determination can be considered as simulation-assisted rather than simulation-based. It enables formal incorporation of uncertainty about the trial assumptions via a design prior and significantly reduces the computational burden for the design of Bayesian trials in general.

Silvia Calderazzo	**Discussant**

Session F	Contributed session: Advances in Bayesian Nonparametrics Chair: Antonio Lijoi
Room	2A
Clara Grazian	**Bayesian analysis of long memory processes**

Abstract: I will investigate a Bayesian approach to estimating the parameters of long memory models. Long memory, characterized by the phenomenon of hyperbolic autocorrelation decay in time series, has garnered significant attention. This is because, in many situations, the assumption of short memory, such as the Markovianity assumption, can be deemed too restrictive. Applications for long memory models can be readily found in fields such as astronomy, finance, and environmental sciences. However, current parametric and semiparametric approaches to modeling long memory present challenges, particularly in the estimation process. In this study, I introduce various methods applied to this problem from a Bayesian perspective, along with a novel semiparametric approach for deriving the posterior distribution of the long-memory parameter. Additionally, I will establish the asymptotic properties of the model. An advantage of this approach is that it allows us to define an approximate (ABC) algorithm to expedite computations.

Maria De Iorio	**Graph Sphere: From Nodes to Supernodes in Graphical Models**

Abstract: High-dimensional data analysis typically focuses on low-dimensional structure, often to aid interpretation and computational efficiency. Graphical models provide a powerful methodology for learning the conditional independence structure in multivariate data by representing variables as nodes and dependencies as edges. Inference is often focused on individual edges in the latent graph. Nonetheless, there is increasing interest in determining more complex structures, such as communities of nodes, for multiple reasons, including more effective information retrieval and better interpretability. In this work, we propose a multilayer graphical model where we first cluster nodes and then, at the second layer, investigate the relationships among groups of nodes. Specifically, nodes are partitioned into supernodes with a data-coherent size-biased tessellation prior which combines ideas from Bayesian nonparametrics and Voronoi tessellations. This construct allows accounting also for dependence of nodes within supernodes. At the second layer, dependence structure among supernodes is modelled through a Gaussian graphical model, where the focus of inference is on superedges. We provide theoretical justification for our modelling choices. We design tailored Markov chain Monte Carlo schemes, which also enable parallel computations. We demonstrate the effectiveness of our approach for large-scale structure learning in simulations and a transcriptomics application.

Elizabeth Bersson	**Covariance Meta Regression**

Abstract: This article considers the task of covariance estimation for high dimensional data consisting of a large amount of features relative to the number of samples. Standard approaches to such covariance modeling include imposing structural assumptions based on auxiliary information regarding the features or utilizing unsupervised dimension reduction with latent factors. In this work, we present a prior distribution for a covariance matrix that incorporates dimension reduction via latent factors and flexibly allows for shrinkage towards a structure learned from feature-level explanatory covariates, or, meta covariates. As with classical regression analysis, the proposed prior can flexibly utilize covariates of mixed types such as categorical and continuous. If no meta covariates are available, we show the proposed prior more accurately estimates non-diagonal population covariance matrices than standard alternatives. One natural application of this work is in modeling chemical exposures where broad chemical classes may be labeled for convenience rather than scientific differentiation. In this application, a researcher may wish to allow for robustness to class label in order to more accurately represent an

across-chemical covariance matrix instead of assuming conditional independence. We demonstrate the utility of this prior in jointly modeling chemical exposures in the NHANES data set.

Xenia Miscouridou

Cox-Hawkes: doubly stochastic spatiotemporal Poisson processes

Abstract: Hawkes processes are point process models that have been used to capture self-excitatory behavior in social interactions, neural activity, earthquakes and viral epidemics. They can model the occurrence of the times and locations of events. Here we develop a new class of spatiotemporal Hawkes processes that can capture both triggering and clustering behavior and we provide an efficient method for performing inference. We use a log-Gaussian Cox process (LGCP) as prior for the background rate of the Hawkes process which gives arbitrary flexibility to capture a wide range of underlying background effects (for infectious diseases these are called endemic effects). The Hawkes process and LGCP are computationally expensive due to the former having a likelihood with quadratic complexity in the number of observations and the latter involving inversion of the precision matrix which is cubic in observations. Here we propose a novel approach to perform MCMC sampling for our Hawkes process with LGCP background, using pre-trained Gaussian Process generators which provide direct and cheap access to samples during inference. We show the efficacy and flexibility of our approach in experiments on simulated data and use our methods to uncover the trends in a dataset of reported crimes in the US.

Session G	Contributed session: Robust and Privacy-Preserving Bayesian Inference Chair: Harrison Quick
Room	10A
Ayush Bharti	**Learning Robust Statistics for Simulation-based Inference under Model Misspecification**

Abstract: Simulation-based inference (SBI) methods such as approximate Bayesian computation (ABC), synthetic likelihood, and neural posterior estimation (NPE), rely on simulating statistics to infer parameters of intractable likelihood models. However, such methods are known to yield untrustworthy and misleading inference outcomes

↑ back to the schedule ↑

under model misspecification, thus hindering their widespread applicability. In this work, we propose the first general approach to handle model misspecification that works across different classes of SBI methods. Leveraging the fact that the choice of statistics determines the degree of misspecification in SBI, we introduce a regularized loss function that penalises those statistics that increase the mismatch between the data and the model. Taking NPE and ABC as use cases, we demonstrate the superior performance of our method on high-dimensional time-series models that are artificially misspecified. We also apply our method to real data from the field of radio propagation where the model is known to be misspecified. We show empirically that the method yields robust inference in misspecified scenarios, whilst still being reasonably accurate when the model is well-specified.

Jack Jewson	**Differentially Private Statistical Inference through beta-Divergence One Posterior Sampling**

Abstract: Differential privacy guarantees allow the results of a statistical analysis involving sensitive data to be released without compromising the privacy of any individual taking part. Achieving such guarantees generally requires the injection of noise, either directly into parameter estimates or into the estimation process. Instead of artificially introducing perturbations, sampling from Bayesian posterior distributions has been shown to be a special case of the exponential mechanism, producing consistent, and efficient private estimates without altering the data generative process. The application of current approaches has, however, been limited by their strong bounding assumptions which do not hold for basic models, such as simple linear regressors. To ameliorate this, we propose betaD-Bayes, a posterior sampling scheme from a generalised posterior targeting the minimisation of the beta-divergence between the model and the data generating process. This provides private estimation that is generally applicable without requiring changes to the underlying model and consistently learns the data generating parameter. We show that betaD-Bayes produces more precise inference estimation for the same privacy guarantees, and further facilitates differentially private estimation via posterior sampling for complex classifiers and continuous regression models such as neural networks for the first time.

Yanran Li

Impacts of census differential privacy for small-area disease mapping to monitor health inequities

Abstract: The U.S. Census Bureau will implement a modernized privacy-preserving disclosure avoidance system (DAS), which includes application of differential privacy, on publicly released 2020 census data. There are concerns that the DAS may bias small-area and demographically stratified population counts, which play a critical role in public health research, serving as denominators in estimation of disease/mortality rates. Using three DAS demonstration products, we quantify errors attributable to reliance on DAS-protected denominators in standard small-area disease mapping models for characterizing health inequities. We conduct simulation studies and real data analyses of inequities in premature mortality at the census tract level in Massachusetts and Georgia. Results show that overall patterns of inequity by racialized group and economic deprivation level are not compromised by the DAS. While early versions of DAS induce errors in mortality rate estimation that are larger for Black than non-Hispanic white populations in Massachusetts, this issue is ameliorated in newer DAS versions.

Kelly Li

Large-scale Bayesian sparse logistic regression via transfer learning on federated observational health data

Abstract: Large-scale sparse logistic regression is commonly used in clinical patient-level prediction models and has wide applications in medical decision-making, risk assessment, and more. However, smaller databases tend to suffer from a lack of available information, leading to overfitting and poor predictive performance. To address these issues, OHDSI (Observational Health Data Sciences and Informatics) is an open-science collaborative focusing on standardizing and combining electronic health records (EHRs) and administrative claims from databases into an international federated network of observational data. Unfortunately, distributed data sites are unable to share patient-level data. We propose a Bayesian approach for sparse logistic regression. The objective is to transfer posterior information about regression coefficients from analyses in larger databases (order 10^5 ~ 10^6 patients by 10^5 generic medical features) to facilitate model fitting in smaller struggling databases. The model imposes an informed prior on each regression coefficient in the informed model that is a discrete mixture of the Bayesian bridge shrinkage prior, and a normal distribution centered and scaled around the larger model's coefficients' posterior means and variances. The mixture indicators are

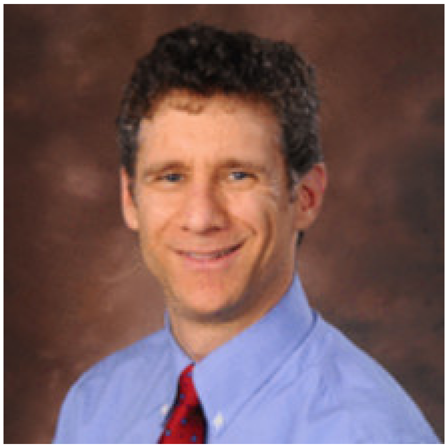
sampled via a beta-binomial hierarchical model.

MCMC is conducted using Nishimura and Suchard (2022)’s “conjugate gradient (CG) accelerated” sampler, which overcomes the computational bottleneck from repeated sampling of the high-dimensional Gaussian from the Polya-Gamma augmented logistic regression model. The desired distribution is generated by solving a linear system with a fast-converging CG algorithm, rather than directly computed. We compare the performance of three PS models with a small sample size (order 10^3 by 10^3): an informed model utilizing the prior distribution previously mentioned, an L1-regularized model, and an "vanilla" model applying a Bayesian bridge prior to all regression coefficients. We assess their effectiveness based on predictive performance, measured by the area under the curve (AUC), and sparsity/interpretability of the resulting solutions.

↑ back to the schedule ↑

11:00 am Foundation Lecture: Andrew Gelman

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



Speaker

Andrew Gelman
Professor of Statistics and Political Science,
Columbia University

Chair: Steve MacEachern

Holes in Bayesian statistics

Abstract: Every philosophy has holes, and it is the responsibility of proponents of a philosophy to point out these problems. Here are a few holes in Bayesian data analysis: (1) the usual rules of conditional probability fail in the quantum realm, (2) flat or weak priors lead to terrible inferences about things we care about, (3) subjective priors are incoherent, (4) Bayesian decision picks the wrong model, (5) Bayes factors fail in the presence of flat or weak priors, (6) for Cantorian reasons we need to check our models, but this destroys the coherence of Bayesian inference. Some of the problems of Bayesian statistics arise from people trying to do things they should not be trying to do, but other holes are not so easily patched. In particular, it may be a good idea to avoid flat, weak, or conventional priors, but such advice, if followed, would go against the vast majority of Bayesian practice and requires us to confront the fundamental incoherence of Bayesian inference. This does not mean that we think Bayesian inference is a bad idea, but it does mean that there is a tension between Bayesian logic and Bayesian workflow which we believe can only be resolved by considering Bayesian logic as a tool, a way of revealing inevitable misfits and incoherences in our model assumptions, rather than as an end in itself. This work is joint with Yuling Yao.

12.00 pm R.V. Ramamoorthi memory sharing

An unstructured session for attendees to share memories about R.V. Ramamoorthi.

Room Magna

1.30 pm Multi-track session 4

For invited sessions, the organizer will chair unless otherwise stated

Session A	Invited session: Novel and Flexible Bayesian Approaches for Causal Inference in Complex Settings Organizers: Mike Daniels and Maria Josefsson Chair: Arman Oganisian
Room	8A
Ashkan Ertefaie	**Nonparametric Bayesian Q-learning for optimization of dynamic treatment regimes in the presence of partial compliance**

Abstract: Dynamic treatment regimes (DTR) are sequential, individualized treatment rules, which are of increasing interest in clinical studies. Existing methods for estimation of DTR's are limited to intention-to-treat analyses--which estimate the effect of randomization to a particular treatment regime without considering the compliance behavior of patients. In this article, we propose a novel Bayesian nonparametric Q-learning approach to construct optimal DTR's that adjusts for partial compliance. We consider the popular potential compliance framework, where some potential compliances are latent and need to be imputed. The key challenge is learning the joint distribution of the potential compliances, which we do using a

↑ back to the schedule ↑

Dirichlet process mixture model. Our approach provides two sets of decision rules: (1) principal treatment regimes, which are functions of the potential compliances; and (2) marginal treatment regimes, where the potential compliances are marginalized.

Herbert Susmann	**Targeted Bayesian Learning for Causal Inference: The Best of Both Worlds?**

Abstract: Targeted Maximum Likelihood Estimation (TMLE) is a framework for constructing asymptotically efficient estimators of statistical parameters in non-parametric models. A popular use case for TMLE is to develop non-parametric estimators of parameters relevant in causal inference, such as the Average Treatment Effect (ATE). For parameters like the ATE, TMLE has the well-known "double-robust" property: it is consistent if either of two nuisance parameters (the outcome regression and propensity score model) are estimated consistently. While TMLE was developed in the framework of frequentist semi-parametric efficiency theory, it also admits a Bayesian analogue. Bayesian TMLE inherits favorable properties of the frequentist estimator, such as double robustness, as well as benefits of the Bayesian paradigm, such as the ability to incorporate complex prior information on the parameter of interest. We will outline the construction of Bayesian TMLE estimators and review their theoretical properties. We will then explore how Bayesian TMLE can be combined with Bayesian hierarchical priors to estimate group treatment effects. The described estimators are accompanied with implementations in the Stan probabilistic programming language to illustrate how they can be used in practice.

Session B	Invited session: Advances in Optimization-Based Bayesian Inference Organizer: Michael Stanley Smith
Room	6A
Linda Tan Siew Li	**Variational inference based on a subclass of closed skew normals**

Abstract: Gaussian distributions are widely used in Bayesian variational inference to approximate intractable posterior densities, but the ability to accommodate skewness can improve approximation accuracy significantly, especially when data or prior information is scarce. We study the properties of a subclass of closed skew normals constructed using affine transformation of independent standardized univariate skew normals as the variational density, and illustrate how this subclass provides increased flexibility and accuracy in approximating the joint posterior density in a variety of applications by overcoming limitations in existing skew normal variational approximations. The evidence lower bound is optimized using stochastic gradient ascent, where analytic natural gradient updates are derived. We also demonstrate how problems in maximum likelihood estimation of skew normal parameters occur similarly in stochastic variational inference and can be resolved using the centered parametrization.

Michael Stanley Smith	**Cutting Feedback in Misspecified Copula Models**

Abstract: In copula models the marginal distributions and copula function are specified separately. We treat these as two modules in a modular Bayesian inference framework, and propose conducting modified Bayesian inference by ``cutting feedback''. Cutting feedback limits the influence of potentially misspecified modules in posterior inference. We consider two types of cuts. The first limits the influence of a misspecified copula on inference for the marginals, which is a Bayesian analogue of the popular Inference for Margins (IFM) estimator. The second limits the influence of misspecified marginals on inference for the copula parameters by using an approximate likelihood to define the cut model. We establish that if only one of the modules is misspecified, then the appropriate cut posterior gives accurate uncertainty quantification asymptotically for the parameters in the other module. Computation of the cut posteriors is difficult, and new variational inference methods to do so are proposed. The efficacy of the new methodology is demonstrated using a multivariate time series application from macroeconomic forecasting. In the latter, cutting feedback from misspecified marginals to a 1096 dimensional copula improves posterior inference and predictive accuracy greatly, compared to conventional Bayesian inference.

David John Nott	**Cutting feedback and modularized analyses in generalized Bayesian inference**

Abstract: This talk considers Bayesian inference under misspecification for complex statistical models comprised of simpler submodels, referred to as modules, that are coupled together. Such multi-modular models often arise when combining information from different data sources, where there is a module for each data source. When some of the modules are misspecified, the challenges of Bayesian inference under misspecification can sometimes be addressed by using "cutting feedback" methods, which modify conventional Bayesian inference by limiting the influence of unreliable modules. Here we investigate cutting feedback methods in the context of generalized posterior distributions, which are built from arbitrary loss functions, and present novel findings on their behaviour. We make three main contributions. First, we describe how cutting feedback methods can be defined in the generalized Bayes setting, and discuss the appropriate scaling of the loss functions for different modules to each other and the prior. Second, we derive a novel result about the large sample behaviour of the posterior for a given module's parameters conditional on the parameters of other modules. This formally justifies the use of conditional Laplace approximations, which provide better approximations of conditional posterior distributions compared to conditional distributions from a Laplace approximation of the joint posterior. Our final contribution leverages the large sample approximations of our second contribution to provide convenient diagnostics for understanding the sensitivity of inference to the coupling of the modules, and to implement a new semi-modular posterior approach for conducting robust Bayesian modular inference. The usefulness of the methodology is illustrated in several benchmark examples from the literature on cut model inference. This is joint work with David Frazier.

David Frazier	**Discussant**
Session C	Invited session: Bayesian Statistical Methods to Operationalize Structural Racism: From Causal Mediation to Spatial Uncertainty and What Lies In Between Organizer: Loni Philip Tabb
Room	9A
Jarvis Chen	**Incorporating uncertainty into the shared component model for decomposing racialized disparities in cancer mortality**

Abstract: The shared component model can provide insight into the spatial patterning of health disparities by decomposing spatially-varying risk surfaces for two population groups into a global component (a health disparity between groups that is constant over space), a shared component (spatial variation in risk that is common to both groups but may be scaled differently), and group-specific components (spatial variation in risk that is unique to each population group). When using population surveillance data, however, estimates of small-area rates may be sensitive to uncertainty in the population denominator, particularly for racially minoritized population groups and for data coming from intercensal years. We present an extension to the shared component model that incorporates multiple sources of population data in a Bayesian hierarchical modeling framework and propagates population denominator uncertainty for inference on the global, shared, and group-specific components of health disparity. We illustrate the application of the model to understanding county-level patterns in cancer mortality using data from the US National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program.

Andrew B. Lawson	**Bayesian Mediation of Deprivation and Segregation Indices in Ovarian Cancer Survival**

Abstract: Deprivation indices (DIs) from area-level socio-economic indicators, and segregation indices (SIs) from the area-level divisions in ethnicity and socio-economic status, attempt to allow for the acknowledged relationship between socio-economic status and health outcomes. It has been observed that African American women experiences significantly poorer survival than White American women in ovarian cancer. To investigate the key factors that contribute to the poor survival experience in African -American women, we evaluated the effects of DI and SI on ovarian cancer survival among African-American women in a Bayesian hierarchical joint mediation framework. Deprivation and segregation effects considered in this study were: Urbanization dimension of Kolak social determinants of health (URB), the Yost index, and the ICE race/income measures. We analyzed a cohort (n=559) of African-American ovarian cancer patients from the African American Cancer Epidemiology Study (AACES). We investigated the direct effect of different DIs and SIs on overall survival and the indirect effects of DIs/SIs through histology, stage, residual disease and delayed diagnosis using the Bayesian structural equation models with Gibbs variable selection. We also examined multiple mediation effects for these mediators with shared random effects.

Emily Peterson

A Bayesian hierarchical small area population model accounting for data source specific methodologies from American Community Survey, Population Estimates Program, and Decennial Census Data

Abstract: Accurate small area race-group specific population counts are necessary for robust health monitoring and reduction of data and health inequities across marginalized populations. In the United States, small area race-specific population counts are published by the United States Census Bureau (USCB) in the form of the Decennial census counts, intercensal population projections (PEP), and American Community Survey (ACS) estimates. Although there are significant relationships between these three data sources, there are important contrasts in data collection, data availability, and accuracy, such that each set of reported race-specific population counts may be subject to different sources and magnitudes of error yet differences in their quality and accuracy are often not considered within spatial epidemiological studies. Consequently, in public health studies, small area race-group specific disease/mortality rates may differ depending on which data source is used for denominator data. To accurately estimate annual small area race-specific population counts, and their associated uncertainties, we present a Bayesian population (BPop) model, which fuses information from all three USCB sources, accounting for data source specific methodologies and race-group specific errors. We produce comprehensive small area race-stratified estimates of the true population, and associated uncertainties, given the observed trends in all three USCB population estimates. The main features of our framework are: 1) a single model integrating multiple data sources, 2) accounting for data source specific data generating mechanisms, and specifically accounting for data source specific errors across race groups, and 3) prediction of population counts for years without USCB reported data. We focus our study on the Black and White only populations for 159 counties of Georgia, and produce estimates for years 2006-2023. We compare BPop population estimates to decennial census counts, PEP annual counts, and ACS 5-year estimates. Our Bayesian population model can be extended to other applications at smaller spatial granularity and for demographic sub-populations defined further by race, age, and sex, and/or for other geographical regions.

Loni Philip Tabb

Evaluating racialized economic segregation in the presence of spatial autocorrelation and statistical uncertainty

Abstract: Research on residential segregation, a prominent domain of structural racism in the United States (US), has been active since the 1950s and originated in a desire to quantify the level of racial/ethnic segregation in neighborhoods. The Index of Concentration at the Extremes (ICE), an operationalization of racialized economic segregation that simultaneously captures racial, and income polarization, has been a popular topic in public health research since it’s development in 2015, with a particular focus on social epidemiology. However, the construction of the ICE metric typically ignores the spatial autocorrelation that may be present in residential segregation; additionally, it is frequently presented without indicating its degree of statistical and spatial uncertainty. To address these issues, we propose reformulating the ICE metric using a Bayesian methodological framework. Assuming a variety of residential segregation scenarios, we conduct simulation studies to evaluate the performance of our novel approach for the reformulation of the ICE measure in comparison to other approaches often used when operationalizing residential segregation. Additionally, the application presented is based on racialized economic segregation in the state of Georgia (located in the southeastern portion of the US), and the proposed modeling approach aims to determine whether racialized economic segregation has changed over two non–overlapping time points.

Session D	Invited session: Bayesian Methods in Ecology Organizer: Alessandro Zito
Room	10A

Bokgyeong Kang

Analyzing whale calls through novel multivariate Hawkes processes

Abstract: Sound is the primary mode of communication among many marine species. Studying recording of these sounds helps to understand the function of the acoustic signals. With increasing anthropogenic noise in the ocean, understanding its impact on the acoustic behavior of marine mammals is needed. One important type of vocalization is an "up-call", thought to serve as a contact call between individuals. Motivated by a dataset recorded by a network of hydrophones in Cape Cod Bay, Massachusetts, utilizing automatically detected up-calls in recordings, we aim to study the communication process of the endangered North Atlantic right whale. We propose novel spatiotemporal excitement modeling consisting of background and excitement processes. The background process describes the intensity of contact calls, providing inference for the impact of diurnal patterns and noise on acoustic behavior of the whales. The excitement process accounts for the potential self- and mutual-excitement. Call incidence is found to be clustered in space and time; a call seems to excite more calls nearer to it in time and space. We find evidence that whales make more calls during twilight hours, respond to other whales nearby, and are likely to remain quiet in the presence of increased ambient noise.

Michele Peruzzi

Bayesian multi-species N-mixture models for large scale spatial data in community ecology

Abstract: Community ecologists seek to model the local abundance of multiple animal species while taking into account that observed counts only represent a portion of the underlying population size. Analogously, modeling spatial correlations in species' latent abundances is important when attempting to explain how species compete for scarce resources. We develop a Bayesian multi-species N-mixture model with spatial latent effects to address both issues. On one hand, our model accounts for imperfect detection by modeling local abundance via a Poisson log-linear model. Conditional on the local abundance, the observed counts have a binomial distribution. On the other hand, we let a directed acyclic graph restrict spatial dependence in order to speed up computations, and use recently developed gradient-based Markov-chain Monte Carlo methods to sample a posteriori in the multivariate non-Gaussian data scenarios in which we are interested.

Narmadha Meenu Mohankumar

Incorporating machine learning within a Bayesian framework to model spatial dependence in occupancy data

Abstract: Spatial models for occupancy data are used to estimate and map the true presence of a species, which may depend on biotic and abiotic factors as well as spatial autocorrelation. Traditionally researchers have accounted for spatial autocorrelation in occupancy data by using a correlated normally distributed site-level random effect, which might be incapable of modeling nontraditional spatial dependence such as discontinuities and abrupt transitions. Machine learning approaches have the potential to model nontraditional spatial dependence, but these

approaches do not account for observer errors such as false absences. By combining the flexibility of Bayesian hierarchal modeling and machine learning approaches, we present a general framework to model occupancy data that accounts for both traditional and nontraditional spatial dependence as well as false absences. We demonstrate our framework using six synthetic occupancy data sets and then apply our approach to understand the spatial dynamics of Thomson’s gazelle (Eudorcas thomsonii) in Tanzania and sugar gliders (Petaurus breviceps) in Tasmania. Our results demonstrate how to model both traditional and nontraditional spatial dependence in occupancy data, which enables a broader class of spatial occupancy models that can be used to improve predictive accuracy and model adequacy.

Session E	Invited session: Well-Calibrated Bayesian Data Science in the Presence of Model Uncertainty: Contemporary Perspectives and Future Directions Organizer: David Draper Chair: Ed George
Room	Magna
Stephen G. Walker	**Model Choice using Martingale Posteriors**

Abstract: Statistical uncertainty arises from what has not been seen and so a model for the unseen sample, which if observed would make all correct choices known, is constructed given what has been observed. A sequence of one step ahead predictive densities would be appropriate and the martingale posterior ensures the future sequence are either conditionally identically distributed or are asymptotically exchangeable. Both of which guarantee the existence of a posterior distribution on the parameters. When the model is unclear, the extension of the posterior to include a number of models can be handled in a straightforward way by using the one step ahead predictive to be based on the best model at that point.

David Blei	**The Posterior Predictive Null**

Abstract: Bayesian model criticism is an important part of the practice of Bayesian statistics. Traditionally, model criticism methods have been based on the predictive

check, an adaptation of goodness-of-fit testing to Bayesian modeling. In modern practice, however, researchers iteratively build and develop many models. While classical predictive checks can help assess each one, they cannot help the researcher understand how they relate to each other. To address this problem, we present the posterior predictive null (PPN). Suppose model B is more complex than model A, but each passes its predictive check. The idea behind a PPN is to check whether data from model B's predictive distribution would pass the check for model A. If it does then we need not select the more complex model. Broadly, we can apply a collection of PPNs to a collection of models, to understand which models are equivalent and which provide different perspectives on the data. We will demonstrate PPNs on several examples.

Yuling Yao	**Calibration for simulation-based inference**

Abstract: Simulation-based inference has been a powerful tool for likelihood-free computation---but the sampling quality is often hard to examine for high-dimensional parameter space; and even if can we run a test to successfully detect miscalibration, we need additional calibration tools to adjust for the sampling error. To this end, I present two methods. First, the discriminative calibration method turns the sampling-check problem into a classification problem, and use a classifier to estimate various divergence between the true posterior and the approximation. It typically has higher power than the traditional rank-based tests. Second, we present simulation-based stacking, a tool to combine multiple non-mixed posterior inferences and return a better fit in terms of different score rules, such as the log score, the rank-calibration, and quantile coverages.

Session F	Contributed session: Simulation-Based Inference Chair: Sam Power
Room	2A
Lorenzo Rimella	**Simulation Based Composite Likelihood**

Abstract: Inference for high-dimensional hidden Markov models is challenging due to

the exponential-in-dimension computational cost of the forward algorithm. To address this issue, we introduce an innovative composite likelihood approach called ``Simulation Based Composite Likelihood'' (SimBa-CL). With SimBa-CL, we approximate the likelihood by the product of its marginals, which we estimate using Monte Carlo sampling. In a similar vein to approximate Bayesian computation (ABC), SimBa-CL requires multiple simulations from the model, but, in contrast to ABC, it provides a likelihood approximation that guides the optimization of the parameters. Leveraging automatic differentiation libraries, it is simple to calculate gradients and Hessians to not only speed-up optimization, but also to build approximate confidence sets. We conclude with an extensive experimental section, where we empirically validate our theoretical results, conduct a comparative analysis with SMC, and apply SimBa-CL to real-world Aphtovirus data.

Umberto Picchini	**Fast, accurate and lightweight sequential simulation-based inference using Gaussian locally linear mappings**

Abstract: Algorithms for simulation-based inference, also known as likelihood-free inference (LFI), come to the rescue when exact inference is hindered by an intractable likelihood. Recent LFI methods have made use of neural networks (NN) to provide approximate, yet expressive constructs for the unavailable likelihood function and the posterior distribution. However, NN-based inference algorithms do not generally achieve an optimal trade-off between accuracy and computational demand, where training time is a significant overhead. Moreover they produce considerable energy expenditure. We propose an alternative that provides approximations to both the likelihood and the posterior distribution in closed form, using structured mixtures of probability distributions. Our approach improves posterior inference compared to state-of-the-art NN-based LFI methods for Bayesian inference, such as sequential neural likelihood (SNL) and SNPE-C, while exhibiting a much smaller computational footprint, by reducing the running time by a factor of 10 compared to SNPE-C as well as considerably reducing the memory requirements and energy consumption. We illustrate our results on several benchmark models.

Wentao Li	**Optimal combination of composite likelihoods using approximate Bayesian computation with application to state-space models**

Abstract: Composite likelihood provides approximate inference when the full

likelihood is intractable. When evaluating the likelihood for some subsets of the data is straightforward, then a composite likelihood can be used as a direct replacement to the likelihood in both Bayesian and frequentist inference. It has been successfully applied in complex models such as max-stable models, state-space models, etc. However, its wider application is limited by two issues. First, the choice of weights to combine these marginal likelihoods can have a significant impact on the information efficiency and is currently an open question. Second, Bayesian inference with composite likelihoods requires calibration to adjust the curvature and is difficult for data without repeated measurements. This work shows that approximate Bayesian computation can properly address these two issues by using marginal score functions as summary statistics. It is shown that the ABC posterior distribution is calibrated and gives optimal Godambe information among a broad class of estimators based on estimating functions that include any combination of marginal scores. Furthermore, for state-space models where the marginal scores do not have closed forms, we propose a novel Monte Carlo approach to estimating the marginal scores and give conditions for the additional Monte Carlo noise to be negligible. This approach provides an off-line static parameter estimator for state-space models that aims at optimal Godambe information. It is shown that the computational cost is O(n) where n is the data length, much cheaper than the O(n^2) cost of standard approaches, e.g. particle MCMC, in both asymptotic results and numerical examples.

Session G	Contributed session: Recent Innovations in Econometrics Chair: Roberto Casarin
Room	3A
Boyuan Zhang	**Incorporating Prior Knowledge of Latent Group Structure in Panel Data Models**

Abstract: The assumption of group heterogeneity has become popular in panel data models. We develop a constrained Bayesian grouped estimator that exploits researchers' prior beliefs on groups in a form of pairwise constraints, indicating whether a pair of units is likely to belong to a same group or different groups. We propose a prior to incorporate the pairwise constraints with varying degrees of confidence. The whole framework is built on the nonparametric Bayesian method, which implicitly specifies a distribution over the group partitions, and so the posterior analysis takes the uncertainty of the latent group structure into account. Monte Carlo experiments reveal that adding prior knowledge yields more accurate estimates of coefficient and scores predictive gains over alternative estimators. We apply our method to two empirical applications. In a first application to forecasting U.S. CPI inflation, we illustrate that prior knowledge of groups improves density forecasts when the data is not entirely informative. A second application revisits the relationship between a country,Äôs income and its democratic transition; we identify heterogeneous income effects on democracy with five distinct groups over ninety countries.

Kaoru Irie

Robust Bayesian Modeling of Counts with Zero Inflation and Outliers

Abstract: Count data with zero inflation and large outliers are ubiquitous in many scientific applications. However, posterior analysis under a standard statistical model, such as Poisson or negative binomial distribution, is sensitive to such contamination. This study introduces a novel framework for Bayesian modeling of counts that is robust to both zero inflation and large outliers. In doing so, we introduce rescaled beta distribution and adopt it to absorb undesirable effects from zero and outlying counts. The proposed approach has two appealing features: the efficiency of the posterior computation via a custom Gibbs sampling algorithm and a theoretically guaranteed posterior robustness, where extreme outliers are automatically removed from the posterior distribution. We demonstrate the usefulness of the proposed method by applying it to trend filtering and spatial modeling using predictive Gaussian processes. (Joint work with Yasuyuki Hamura and Shonosuke Sugasawa.)

Karla Vianey Palacios Ramirez

Semiparametric Bayesian modeling of nonstationary joint extremes

Abstract: We propose a novel Bayesian model for inferring about the intensity of observations in the joint tail over time, and for assessing if two stochastic processes are asymptotically dependent. To model the intensity of observations exceeding a high threshold, we develop a Bayesian nonparametric approach that defines a prior on the space of what we define as EDI (Extremal Dependence Intensity) functions. In addition, a parametric prior is set on the coefficient of tail dependence. An extensive battery of experiments on simulated data showcases that the proposed methods are able to recover the true targets in a variety of scenarios. An application of the

↑ back to the schedule ↑

proposed methodology to a set of big tech stocks, known as FAANG, sheds light on some interesting features on the dynamics of their combined losses over time.

Makoto Takahashi	**Realized Stochastic Volatility Model with Skew-t Distributions for Improved Volatility and Quantile Forecasting**

Abstract: Forecasting volatility and quantiles of financial returns is essential for accurately measuring financial tail risks, such as value-at-risk and expected shortfall. The critical elements in these forecasts involve understanding the distribution of financial returns and accurately estimating volatility. This paper introduces an advancement to the traditional stochastic volatility model, termed the realized stochastic volatility model, which integrates realized volatility as a precise estimator of volatility. To capture the well-known characteristics of return distribution, namely skewness and heavy tails, we incorporate three types of skew-t distributions. Among these, two distributions include the skew-normal feature, offering enhanced flexibility in modeling the return distribution. We employ a Bayesian estimation approach using the Markov chain Monte Carlo method and apply it to major stock indices. Our empirical analysis, utilizing data from US and Japanese stock indices, indicates that the inclusion of both skewness and heavy tails in daily returns significantly improves the accuracy of volatility and quantile forecasts.

3:30 pm Keynote Lecture: Alexandra Schmidt

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

Speaker

Alexandra Schmidt
Professor of Biostatistics,
McGill University

Chair: Amy Herring

Moving beyond normality in spatial and spatiotemporal processes

Abstract: Gaussian processes (GPs) are routinely used in the modelling of spatial and spatio-temporal processes. This is because they are fully specified by their mean and covariance functions, and prediction to unobserved locations easily follows from the properties of the partition of multivariate normal distributions. However, observations associated with environmental processes rarely follow a normal distribution. Although data are commonly transformed to attain normality, transformation can affect the quantification of uncertainty of predictions. I will discuss some spatial and spatiotemporal models for processes on their original scale. In particular, I will propose an extension of the popular multivariate dynamic linear model to accommodate heavy tails both for univariate and multivariate spatio-temporal processes. GPs require the evaluation of the determinant and inverse of covariance matrices, which becomes expensive as the number of observed sites grows large. As the proposed class of models involve more than one GP, inference can become computationally challenging when using Markov chain Monte Carlo

methods to obtain samples from the resultant posterior distribution. To address this challenge, I will discuss how to approximate the proposed models when the number of observations across space is large. Examples using the proposed extensions include the analysis of temperature measurements across the Basque country, and the joint modelling of PM10 and ozone across the UK.

5.00 pm Bayesian Analysis Business Meeting

Invite-only

Room Meeting Room 1

↑ back to the schedule ↑

ISBA World Meeting

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

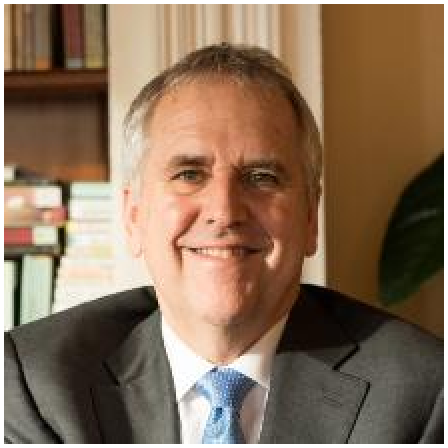
Thursday 4th, 2024

Schedule of the Day

9:00 am - 10:30 am	De Finetti Lecture: Michael Jordan
10:30 am - 11:00 am	*Coffee break*
11:00 am - 12:00 pm	Keynote Lecture: Antonietta Mira
12:00 pm - 1:30 pm	*Lunch break*
1:30 pm - 4:00 pm	Poster Session 2
4:00 pm - 4:30 pm	*Coffee break*
4:30 pm - 6:00 pm	Multi-track session 5
6:00 pm - 7:00 pm	j-ISBA meeting

9:00 am De Finetti Lecture: Michael Jordan

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



Speaker

Michael Jordan
Pehong Chen Distinguished Professor,
UC Berkeley

Discussants
David Dunson, Duke University
Chris Holmes, University of Oxford

Chair: Matthias Katzfuß

Uncertainty, Asymmetry of Information, and Statistical Contract Theory

Abstract: Contract theory is the study of economic incentives when parties transact in the presence of private information, such that prior distributions may differ and not be common knowledge. We augment classical contract theory to incorporate a role for learning from data, where the overall goal of the adaptive mechanism is to obtain desired statistical behavior. We consider applications of this framework to problems in federated learning, the delegation of data collection, and principal-agent regulatory mechanisms.

11:00 am Keynote Lecture: Antonietta Mira

Room	live at	Aula Magna "Guido Cazzavillan"
	streaming	Rooms 4A - 5A - 9A - 10A
	youtube	[YouTube streaming]

Speaker

Antonietta Mira
Professor of Statistics,
Università della Svizzera italiana

Chair: Francois Caron

Beyond the noise: intrinsic dimension estimation and exploitation

Abstract: Real-world datasets tend to show a high degree of (possibly) non-linear correlations and constraints between their features. This means that, despite a very large embedding dimensionality, data typically lie on a manifold characterized by a much lower intrinsic dimension (ID) which, in the presence of noise, may depend on the scale at which the data is analysed. This fact raises interesting questions: How many variables, or combinations there of, are necessary to describe a real-world data set without significant information loss? What is the appropriate scale at which one should analyze and visualize the data? These two issues, which are often considered unrelated, are actually strongly entangled, and can be addressed within a unified framework. We introduce an approach in which the optimal number of variables and the optimal scale are determined self-consistently, recognizing and bypassing the scale at which the data are affected by noise. To this aim we estimate the data ID in an adaptive way, and exploit it as a summary statistics in Approximate Bayesian Computation for inference in network type data. Sometimes, within the same dataset,

it is possible to identify more than one ID meaning that different subsets of the data points lie onto manifolds with different IDs. Identifying these manifold provides a clustering of the data, and in many real world applications a simple topological feature, like the ID, allows to uncover a rich data structure and improves our insight into subsequent statistical analysis. Examples of these applications range from gene expression to protein folding, pandemic evolution, FMRI, all the way to finance, sport data and the analysis of the representations of neural networks.

4.30 pm Multi-track session 5

For invited sessions, the organizer will chair unless otherwise stated

Session A	Invited session: Bayesian Learning in Highly Structured Latent Variable Models Organizers: Anindya Bhadra and Sylvia Fruehwirth-Schnatter
Room	7A
Nadja Klein	**A Framework for Consistent Effect Estimation and Selection in Generalized Sparse Partially Linear Additive Models**

Abstract: Accurately selecting and estimating smooth functional effects in additive models with potentially many functions is a challenging task, especially if the components are decomposed into linear and nonlinear effects. We provide a rigorous definition of the true linear and nonlinear effects of an additive component using projections and introduce a new construction of the Demmler-Reinsch basis for penalized splines. We prove that our representation allows to consistently estimate the true effects as opposed to the commonly employed mixed model representations. Equipping the reparameterized regression coefficients with normal beta prime spike and slab priors allows us to automatically determine whether a continuous covariate has a linear, a nonlinear or no effect at all. We provide new theoretical results for the prior and a compelling explanation for its superior Markov chain Monte Carlo mixing performance compared to the spike-and-slab group lasso prior. Finally, we illustrate the developed methodology along effect selection on the hazard rate of a time-to-event response in the additive Cox regression model with spatial effects in simulations and on leukemia survival data.

Sylvia Fruehwirth-Schnatter	**Exchangeable shrinkage process priors for learning latent dimensions with an application to mixture-of-factor-analysers**

Abstract: This talk addresses the challenging task of learning the dimension of latent variables in highly structured latent variable models using exchangeable shrinkage process (ESP) priors. An ESP prior assigns spike-and-slab distributions on a finite sequence of model parameters, where the spike probability arises from a hierarchical prior. It is shown that such an ESP prior has a representation as (generalized) cumulative shrinkage process (CUSP) prior where an increasing probability is assigned to the spike, allowing in this way to learn the number of active model parameters in the model needed to explain the data. Establishing this connection between ESP and CUSP priors is fruitful both from a methodological as well as a computational viewpoint. It allows a deeper understanding regarding the choice of hyperparameters and hyperpriors and leads to considerable improvement in MCMC estimation compared to common data augmentation techniques for CUSP priors. For illustration, an application to mixture-of-factor-analyser models which are commonly applied to data that exhibit both low-dimensional structures and unobserved heterogeneity is discussed. For this specific model class, the entire model-specification including the number of clusters, the number of factors and the factor loading matrix within each cluster are unknown and can be learned from the data with the help of ESP priors.

Anindya Bhadra	**Likelihood Based Inference in Fully and Partially Observed Exponential Family Graphical Models with Intractable Normalizing Constants**

Abstract: Probabilistic graphical models that encode an underlying Markov random field are fundamental building blocks of generative modeling to learn latent representations in modern multivariate data set with complex dependency structures. Among these, the exponential family graphical models are especially popular given their fairly well-understood statistical properties and computational scalability to high-dimensional data based on pseudo-likelihood methods. These models have been successfully applied in many fields such as statistical physics (Ising) and genetics (count graphical models). Another strand of models allows some nodes to be latent, so as to the marginal distribution of the observable nodes can depart from exponential family to capture more complex dependence. These approaches form the basis of generative models in AI, such as the Boltzmann machines and their restricted versions. A fundamental barrier to likelihood-based (i.e., both maximum likelihood and fully Bayesian) inference in both fully and partially observed cases is the intractability of the likelihood. The usual workaround is via adopting pseudolikelihood based approaches, following the pioneering work of Besag (1974). The goal of this paper is to demonstrate that full likelihood based analysis of these models is feasible in a computationally efficient manner. The chief innovation lies in using a technique of Geyer (1991) to estimate the intractable normalizing constant, as well as its gradient, for count graphical models. Extensive numerical results, and comparisons with pseudo-likelihood based approaches demonstrate the applicability of the proposed method.

Session B	Invited panel discussion: Which Future for ISBA Conferences? Organizers: Gertraud Malsiner-Walli and Guido Consonni
Room	Magna
Christian Robert	**Through the looking glass: let a thousand meetings bloom, let a hundred schools of thought contend!**

Abstract: The pandemic and the fabulous ISBA 2021 demonstrated that hybrid meetings were more than the sum of their parts, by allowing a wider range of the community to take part into World meetings, without having to cross half (or more) of the World. Based on the experience of mirror conferences I helped organise, I would like to make a plea for the generalisation of such mirror meetings to all ISBA backed conferences, so that a wider diversity of attendants could participate and mix local with global input to the community. My main argument is that creating mirrors to the main meeting are only contributing positively, by allowing people with diverse reasons for not travelling to be or to feel a component of the Bayesian community. I will discuss several ways of creating such multilevel conferences, based on previous

experiments.

Ron Wasserstein | **Balancing the Multiple Purposes of Societal Conferences**

Abstract: Societal conferences, particularly the large annual meetings of professional societies, serve multiple purposes. Attendees can network, share and get feedback on their research, take professional development coursework, and more. Societies promote their activities and develop membership through these conferences. And, for many organizations, the annual meeting provides an essential revenue source. But in-person societal annual conferences exclude people who cannot afford or are otherwise unable to travel to them. Those excluded are often those who would most benefit from the meeting –students and early career individuals and those from geographically distant locations. The conferences also come at a cost to the environment. As we search for solutions to these issues, the multiple purposes of the conferences must be acknowledged and prioritized.

Kate Lee | **Barriers to participation in academic conferences: a personal view**

Abstract: Academic conferences are incredibly valuable opportunities for researchers to exchange ideas and findings. On the other hand, large international in-person conferences can sometimes unintentionally exclude certain groups. Some well-discussed factors include geographical isolation, financial hurdles, limited accessibility, language barriers, lack of diversity and inclusion initiatives, among others. Despite efforts by institutions like ISBA and world meeting organizers to foster inclusivity, numerous barriers persist. In this talk, I will offer my personal view on these obstacles, their impact on research career opportunities, and discuss strategies that can be put in place to cultivate a more inclusive scholarly community.

Session C | Invited session: Bayesian Methods for Astrophysics Discovery based on Big Data

↑ back to the schedule ↑

	Organizer: Hyungsuk Tak Chair: David van Dyk
Room	2A
Yang Chen	**Multi-instrument Cross-Calibration in Astronomy with Multiplicative Shrinkage**

Abstract: Calibration data are often obtained by observing several well-understood objects simultaneously with multiple instruments, such as satellites, for measuring astronomical sources. Analyzing such data and obtaining proper concordance among the instruments is challenging when the physical source models are not well understood, when there are uncertainties in "known" physical quantities, or when data quality varies in ways that cannot be fully quantified. Furthermore, the number of model parameters increases with both the number of instruments and the number of sources. Thus, the concordance of the instruments requires careful modeling of the mean signals, the intrinsic source differences, and measurement errors. We propose a log-Normal model and a more general log-t model that respects the multiplicative nature of the mean signals via a half-variance adjustment yet permits imperfections in the mean modeling to be absorbed by residual variances. We present analytical solutions in the form of power shrinkage in special cases and develop reliable Markov chain Monte Carlo algorithms for general cases, both of which are available in the Python module CalConcordance. We apply our method to several datasets, including a combination of observations of active galactic nuclei (AGN) and spectral line emission from the supernova remnant E0102, obtained with a variety of X-ray telescopes such as Chandra, XMM- Newton, Suzaku, and Swift. We demonstrate that our method provides helpful and practical guidance for astrophysicists when adjusting for instrument disagreements.

Sujit Ghosh	**A Nonparametric and Probabilistic Framework for Characterizing Planetary Samples in Higher Dimensions**

Abstract: Fundamental to our understanding of planetary bulk compositions is the relationship between their masses and radii (M-R), two properties that are often not simultaneously known for most exoplanets. However, while many previous studies have modeled the two-dimensional relationship between planetary mass and radii, this approach largely ignores the dependencies on other properties that may have influenced the formation and evolution of the planets. This work provides a non-trivial

extension of the M-R relation by including additional variables while also incorporating asymmetric measurement uncertainties and upper limits in the data. The proposed probabilistic framework is based on nonparametric estimation of copula and the marginals for arbitrary fixed dimension making it suitable for various applications for planetary relationships. We showcase the potential of this multidimensional approach to three science cases: (i) a four-dimensional joint fit to planetary mass, radius, insolation, and stellar mass, hinting of changes in planetary bulk density across insolation and stellar mass; (ii) a three-dimensional fit to the California Kepler Survey sample showing how the planet radius valley evolves across different stellar masses; and (iii) a two-dimensional fit to a sample of Class-II protoplanetary disks in Lupus while incorporating the upper limits in dust mass measurements.

Kaisey Mandel	**GausSN: Bayesian Time-Delay Estimation for Strongly Lensed Supernovae**

Abstract: We present GausSN, a Bayesian semi-parametric Gaussian Process (GP) model for time-delay estimation from the brightness time series of the resolved multiple images of strongly gravitationally lensed supernovae (glSNe). The gravitational field of an intervening galaxy can split the light from a distant exploding star, or supernova, into multiple images of the same event. The time delays between the multiple images of the explosion can be used to estimate the expansion rate of the Universe, the Hubble constant. GausSN models the underlying light curve non-parametrically using a GP. Without assuming a template light curve shape for each supernova type, GausSN fits for the time delays of all images using data in any number of wavelength filters simultaneously. We also introduce a novel time-varying magnification model to capture the effects of microlensing by stars along the line-of-sight, alongside time-delay estimation. In this analysis, we model the time-varying relative magnification as a sigmoid function, as well as a constant for comparison to existing time-delay estimation approaches. We demonstrate that GausSN provides robust time-delay estimates for simulations of glSNe from the Nancy Grace Roman Space Telescope and the Vera C. Rubin Observatory's Legacy Survey of Space and Time (Rubin-LSST). We find that up to 43.6% of time-delay estimates from Roman and 52.9% from Rubin-LSST have fractional errors of less than 5%. We then apply GausSN to Supernova Refsdal and find the time delay for the fifth image is consistent with the original analysis, regardless of microlensing treatment.

↑ back to the schedule ↑

Therefore, GausSN maintains the level of precision and accuracy achieved by existing time-delay extraction methods with fewer assumptions about the underlying shape of the light curve than template-based approaches, while incorporating microlensing into the statistical error budget. GausSN is scalable for time-delay cosmography analyses for current projections of glSNe discovery rates from Rubin-LSST and Roman. https://arxiv.org/abs/2311.17997

David van Dyk	**Discussant**

Session D	Invited session: Monte Carlo Algorithms using Modern Hardware Organizer: Charles Margossian
Room	6A
Matt Hoffman	**Running Many-Chain Hamiltonian MCMC on Cheap GPUs**

Abstract: Traditional Markov chain Monte Carlo (MCMC) workflows run a small number of parallel chains for many steps. This workflow made sense in the days when parallel computation was limited to the number of cores on one's CPU, since it lets one amortize the cost of burn-in/warmup (i.e., discarding early samples to eliminate bias) over as many useful samples as possible. But modern commodity GPUs (<$500 retail, or a few dollars per hour to rent in the cloud), in conjunction with modern software frameworks like TensorFlow Probability, make it possible to run 50–100 chains in parallel without paying much of a price in wallclock time. For many Bayesian inference problems, this means that we can get reasonably low-variance estimates of posterior expectations using as little as one sample per chain, dramatically reducing the time we spend waiting for results. In this talk, I will present some of our work on new Hamiltonian Monte Carlo algorithms that can take full advantage of GPUs.

Louis Grenioux	**Sampling from multimodal distributions with stochastic localization**

Abstract: Building upon score-based learning, new interest in stochastic localization techniques has recently emerged. In these models, one seeks to noise a sample from the data distribution through a stochastic process, called observation process, and progressively learns a denoiser associated to this dynamics. In this talk, we consider a general stochastic localization framework and introduce an explicit class of observation processes, associated with flexible denoising schedules. I will provide a complete methodology, Stochastic Localization via Iterative Posterior Sampling (SLIPS), to obtain approximate samples of these dynamics, and as a by-product, samples from the target distribution.

Nicolas Chopin	**Connection between Tempering SMC and Entropic Mirror Descent**

Abstract: In this talk, I will make connections between tempering SMC (Sequential Monte Carlo) and entropic mirror descent to sample from a target probability distribution whose unnormalized density is known. My co-authors and myself have established that tempering SMC corresponds to entropic mirror descent applied to the reverse Kullback-Leibler (KL) divergence and obtain convergence rates for the tempering iterates. Our result motivates the tempering iterates from an optimization point of view, showing that tempering can be seen as a descent scheme of the KL divergence with respect to the Fisher-Rao geometry, in contrast to Langevin dynamics that perform descent of the KL with respect to the Wasserstein-2 geometry. We use the connection between tempering and mirror descent iterates to justify common practices in SMC and derive adaptive tempering rules that improve over other alternative benchmarks in the literature. (joint work with Francesca Crucinio and Anna Korba).

Session E	Invited session: Bayesian Methods for Biomedical Imaging Data Organizers: Michele Guindani and John Kornak
Room	9A
Jeffrey S Morris	**Spatiotemporal Distributional Regression of Climate Data**

Abstract: In many areas of science, technological advances have led to devices that produce an enormous number of measurements per experimental units, including wearable devices, biomedical image data, and geospatial monitoring data. Frequently, researchers deal with these data by extracting summary statistics from these data (e.g. mean or variance) and then modeling those, but this approach can miss key insights when the summaries do not capture all of the relevant information in the raw data. In this talk, we will discuss Bayesian frameworks for modeling the entire distribution of the measurements observed for each experimental unit as responses, relating them to covariates and accounting for potential spatial and/or temporal structure among the experimental units. We apply our approach to model climate data, modeling the distribution of daily temperatures within a given site and year across all sites and years using a spatiotemporal distributional regression model, able to characterize the climate change in Iceland over the past 7 decades. The general distributional regression framework has great applicability to large data streams obtained from many commonly used applications, and provides a scalable method for rigorous regression analyses summarizing how subject-specific distributions vary with covariates.

Laura D'Angelo

Uncovering the heterogeneity of the neuronal activity in calcium imaging studies via Bayesian mixture models

Abstract: The modern technique of calcium imaging is revolutionizing the understanding of the nervous system, thanks to its ability to image the activity of individual neurons over time in freely moving animals. This technology has led to remarkable insights into how neurons process and encode information, both individually and collectively. In this talk, we discuss how Bayesian mixture models can help uncover the complex functioning of neurons in different experimental studies. First, we discuss a nested mixture model to analyze how the activity of an individual cell is affected by visual stimuli that vary over time. Then, we move to a multivariate model to identify groups of co-activating neurons, where the information on the anatomical proximity is introduced to inform the clustering procedure. At the basis of both models is a simple but effective state-space formulation that describes the calcium dynamic and relates it to the latent firing events. Through these examples, we show how the Bayesian nonparametric framework is ideal for flexibly modeling the unobserved neuronal activity, identifying patterns of activity, borrowing information across experimental conditions, and including additional knowledge.

Veera Baladandayuthapani

Bayesian Statistical Models for Spatial Imaging in Cancer

Abstract: Recent breakthrough in cancer imaging techniques such as spatial multiplex immunofluorescence imaging and digital pathology imaging facilitate detailed assessment, characterization, and visualization of various aspects of spatial

heterogeneity within and across tumors. The resulting high-resolution data raise many statistical and modeling challenges such as complex spatial relationships, high-degree of spatial heterogeneity within and between samples and non-conformable spaces for population analyses. In this talk, I will present Bayesian spatial models to address some key scientific questions for such data that include probabilistic quantification of spatial interactions through point-processes, spatially structured regression models to integrate imaging and genomics data, and spatial clustering models for non-conformable spaces. The methods will be exemplified using several case studies in cancer.

Jaroslaw Harezlak	**Discussant**
Session F	Contributed session: Advances in Bayesian Modeling Chair: Tamara Broderick
Room	4A
Marcio Alves Diniz	**Bayesian and Ridge regression models: a comparison by conformal prediction**

Abstract: Ridge regression (RR) is a method of estimation of the coefficients of multiple linear regression models that were introduced mainly to deal with some limitations of the ordinary least squares (OLS) estimators, namely, when the covariates are highly correlated and/or the number of covariates is proportionally high when compared to the sample size. Although biased, RR estimators have smaller mean squared errors, but the estimated regression function may have low predictive accuracy, an important topic in machine learning problems. Hence, the RR estimators have also been evaluated according their predictive performance, usually considering the squared error as loss function. In this work we compare the predictive performance of RR and linear Bayesian regression (LBR) under the usual assumptions of Gaussian distribution of the independent error terms and conjugate prior (normal-inverse gamma) for the parameter vector. In order to not restrict the comparison to point predictions, we use LBR and RR point prediction estimates as inputs to conformal prediction, an approach to obtain prediction intervals that contain the target or explained variable with a preassigned probability. The comparison is made by means of simulation studies under different specifications of the generating data and two real data sets. The first one includes variables corresponding to real estate single-family residential apartments in Tehran, Iran, and the second brings salaries and performance measures of 874 ice hockey players for the 2016/2017 season of the North American league, National Hockey League (NHL).

Alejandra Avalos-Pacheco

Rotational Invariant Sparse Factor Models with the l1-ball prior

Abstract: Integrative factor models have proven to be crucial for identifying reproducible biological pathways shared by different cancer studies that traditional factor analysis approaches may miss due to systematic biases. Existing integrative factor models, while valuable, often neglect the impact of covariates and confounders, introducing bias into the signal and/or lack a study-specific factor structure totally independent of the common latent structure. Moreover, these models require post-processing steps for loadings, such as varimax rotation, and restrictions on the loading matrix for identifiability, essential for interpretation. To address these challenges, we present a novel class of integrative factor models: "Rotational Invariant Sparse Factor Models" (RISFM). RISFM displays several advantages:

1. Providing sparse low-dimensional common and study-specific factors while adjusting for confounding effects using Bayesian methods.
2. Addressing identifiability issues crucial for interpretation through the l1-ball prior.
3. Ensuring computational efficiency for practical applicability.

We validate the proposed RISFM approach through extensive simulations and its application to hepatocellular carcinoma human-mice gene expression cancer data. The results showcase the utility of RISFM to determine the overall human-mice genomic similarities, to identify the most appropriate mouse model for studying different human patient sub-populations, and to obtain the co-regulation mechanisms that are idiosyncratic to only humans or mice. The main goal is to improve the current understanding of mouse mutation models and contribute to develop new ones for precision medicine.

Catherine Forbes	**Assessment of local influence for Bayesian moment condition models**

Abstract: Moment condition models are common in Economics and related disciplines where theory implies a set of moment constraints, yet a full generative probability model is not specified. Case influence diagnostics provide a way to assess the sensitivity of the posterior distribution to perturbations of the data – something that is especially important in this context because the empirical moment conditions can be greatly affected by the presence of extreme observations. Measures of divergence between the full-data and case-deleted moment condition posterior are considered, as are measure of local influence from infinitesimal perturbations to the moment-based likelihood function. These more traditional methods are compared against a new technique that embeds data perturbations directly inside the moment conditions, resulting in posterior information regarding case-influence.

Sylvia Richardson	**Bayesian graphical horseshoe for joint network inference: a strategy for characterising differences in biological function between clinical groups**

Abstract: Network models are useful tools for modelling complex associations. In statistical omics, such models are increasingly popular for identifying and assessing functional relationships and pathways. If a Gaussian graphical model is assumed, conditional independence is determined by the non-zero entries of the inverse covariance (precision) matrix of the data. The Bayesian graphical horseshoe estimator provides a robust and flexible framework for precision matrix inference, as it introduces local, edge-specific parameters which prevent over-shrinkage of non-zero off-diagonal elements. However, its applicability is currently limited in statistical omics settings, which often involve high-dimensional data from multiple conditions that might share common structures. In this talk, we will discuss a novel joint graphical horseshoe estimator, which borrows information across multiple related networks to improve estimation, and its scalable implementation via an expectation conditional maximisation (ECM) algorithm. Using simulated data, we show that the approach successfully shares information between networks while capturing their differences, ensuring better interpretability of the resulting edges. Building on this, we propose a strategy to highlight pathway activations which differ between networks of related clinical conditions. The strategy is applied to proteomic data from a phase II clinical trial investigating the effect of neoadjuvant therapy in breast carcinomas, leading to

the identification of potential mechanisms for disease progression and treatment response. arXiv:2206.11820.

Session G	Contributed session: Bayes in Practice: from Social Networks to Epidemiological Surveillance Chair: Daniele Durante
Room	5A
Brisilda Ndreka	**Bayesian Estimation of Contagion Effect: An Application of Friendship Networks and Alcohol Behavior**

Abstract: When examining behavioral patterns within specific social groups, social networks primarily focus on the contagion effect phenomenon. Nonetheless, the impact of peer effects characterized by the tendency to imitate the behaviors of friends and the selection process, where individuals tend to affiliate with others sharing similar traits, significantly contribute to shaping social behavior and frequently interconnecting. This article presents a Bayesian approach that utilizes latent space estimation methods to detect and examine contagion effects, considering the impact of social selection. The research encompasses a methodological explanation followed by a sequence of simulation trials designed to explore operational functionalities and possible real-world applications. To illustrate the potential correlation between changes in alcohol use and the influence of social networks, the study concludes by presenting an example of adolescent drinking behavior.

Jennifer Hill	**Bayesian Causal software: Is it used? Does it work?**

Abstract: Bayesian researchers, along with other statisticians and empirical scientists, have succeeded in recent years in providing a wide range of software to enable other researchers to capitalize on their methodological advances. However few, if any, researchers have rigorously evaluated whether the software they provide actually positively contributes to the research endeavor. This novel study provides experimental evidence about the effectiveness of a new tool for Bayesian causal inference, thinkCausal, and evaluates its effectiveness with respect to accurate estimation of treatment effects and ease of use relative to freely available alternatives.

Saptarshi Roy	**Bayesian Causal Structure Learning using Directed Cyclic Graph from Multivariate Functional Data**

Abstract: Exploring causal relationships from multivariate functional data has recently gained considerable attention. This article introduces a functional structural equation model designed for learning causal structures, particularly when the underlying graph incorporating multivariate functions may contain cycles. To enhance interpretability, our model incorporates a low-dimensional causal embedded space, preserving all relevant causal information from the multivariate functional data within this reduced subspace. We demonstrate the causal identifiability of the proposed model under standard assumptions commonly found in the causal discovery literature. For model inference, we establish a comprehensive Bayesian framework with appropriate prior specifications and incorporate uncertainty quantification through posterior summaries. Through extensive simulation studies, we showcase the superior performance of our method in terms of causal graph estimation compared to existing methods. Additionally, we apply the proposed method to analyze a brain EEG dataset, further illustrating its effectiveness.

Leonardo Bastos	**Advances in delay correction for epidemiological surveillance of infectious diseases**

Abstract: Real-time early warning systems for monitoring infectious diseases are needed to provide fast public health answers to avoid outbreaks and epidemics or to reduce severe outcomes. One problem for real-time analysis is the reporting delay, which consists of the time between the case being identified and the case becoming known in the surveillance system. There are some approaches to learning about the delay and correcting it. In this work, we extend the chain-ladder model by constructing a fast Bayesian random effects model for spatiotemporal disease counts, adding age structure motivated by COVID-19, and proposing a multivariate approach inspired by co-occurrence of two mosquito-borne diseases (dengue and chikungunya) in Brazil. We present some results associated with two Brazilian early warning systems monitoring respiratory infectious disease hospitalizations and notifications of dengue and chikungunya fever. Our methods are implemented in both early-warning systems and have been used by the Brazilian Ministry of Health to monitor these diseases and provide public health actions to mitigate epidemics.

6.00 pm j-ISBA meeting

Room Magna

The j-ISBA meeting welcomes all early-career Bayesian researchers, whether or not they are members of j-ISBA (the junior section of ISBA). This one-hour session is designed to facilitate connections among junior fellows, provide opportunities to interact with the j-ISBA board, and offer insights into j-ISBA's activities. Additionally, a glass of Prosecco will be served to enhance the convivial atmosphere. Join us for an engaging and enjoyable hour.

↑ back to the schedule ↑

ISBA World Meeting

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

Friday 5th, 2024

Schedule of the Day

8:30 am - 10:00 am	Savage Award Session: Applied Methodology
8:30 am - 10:00 am	Multi-track session 6
10:00 am - 10:30 am	*Coffee break*
10:30 am - 12:00 pm	Savage Award Session: Theory & Methods
10:30 am - 12:00 pm	Multi-track session 7
12:00 pm - 1:30 pm	*Lunch break*
1:30 pm - 2:30 pm	Keynote Lecture: Omiros Papaspiliopoulos
2:30 pm - 3:00 pm	*Coffee break*
3:00 pm - 4:30 pm	Multi-track session 8
5:00 pm - 7:30 pm	Poster session 3

8.30 am Savage Award Session: Applied Methodology

Room 6A
Chair Radu Craiu

Raquel Barata

Flexible dynamic quantile linear models with application to the characterization of atmospheric rivers

Abstract: The primary component of many atmospheric river detection schemes is the thresholding of the integrated water vapor transport (IVT) magnitude at a single quantile over time. Utilizing a recently developed family of parametric distributions for quantile regression, we develop a flexible dynamic quantile linear model (exDQLM) which enables versatile, structured, and informative estimation of a dynamic quantile. In addition to a Markov chain Monte Carlo (MCMC) algorithm, we develop an efficient importance sampling variational Bayes (ISVB) algorithm for fast approximate Bayesian inference which produces comparable results to the MCMC in a fraction of the computation time. Further, we develop a transfer function extension to our exDQLM as a method for quantifying non-linear relationships between a quantile of a climatological response and an input. The utility of our transfer function exDQLM is demonstrated in capturing both the immediate and lagged effects of El Niño Southern Oscillation Longitude Index on the estimation of the 0.85 quantile IVT.

Wei Jin

Novel Bayesian Methods for Precision Medicine in HIV

Abstract: Numerous adverse effects (e.g., depression) have been reported for combination antiretroviral therapy (cART) despite its remarkable success in viral suppression in people with HIV. To improve long-term health outcomes for people with HIV, there is an urgent need to design personalized optimal cART with the lowest risk of comorbidity in the emerging field of precision medicine for HIV. Large-scale HIV studies offer researchers unprecedented opportunities to optimize personalized cART in a data-driven manner. However, the large number of possible drug combinations for cART makes the estimation of cART effects a high-dimensional combinatorial problem, imposing challenges in both statistical inference and decision-making. We develop a two-step Bayesian decision framework for optimizing sequential cART assignments. In the first step, we propose a dynamic model for individuals' longitudinal observations using a multivariate Gaussian process. In the second step, we build a probabilistic generative model for cART assignments and design an uncertainty-penalized policy optimization using the uncertainty quantification from the first step. Applying the proposed method to a dataset from the Women's Interagency HIV Study, we demonstrate its clinical utility in assisting physicians to make effective treatment decisions, serving the purpose of both viral suppression and comorbidity risk reduction.

Matteo Pedone	**Covariate-dependent Bayesian Models for Heterogeneous Populations**

Abstract: In this thesis, we propose two novel Bayesian models for the analysis of health and genomic data, for which traditional methods are often found to be inefficient or unsuitable. Our approaches are motivated by the emerging field of precision medicine, whose ultimate goal is to select the optimal treatment accounting for patient and disease's variability. The main distinctive mark of statistical methodology in the precision medicine paradigm is to leverage patients' heterogeneity to obtain subject-specific inference.

First, motivated by a microbiota study on patients affected by colorectal cancer, we propose a model designed to analyze data that exhibit a hierarchical structure induced by measurements from multiple tissues of the same patient. Our goal is to capture patients’ heterogeneity and similarities in terms of effects altering microbiota composition.

Building upon the Dirichlet-multinomial model, we propose a flexible regression model, where coefficients are allowed to be smooth functions of the covariates. This results in a subject-specific model where varying coefficients include two-way linear and non-linear interactions as special cases. This allows us to recover associations and interactions patterns that may be specific for each individual rather than estimated at population level.

In the second contribution, we develop a predictive model for the selection of the personalized optimal treatment in oncology, when a predictive signature and a set of prognostic biomarkers are available. Predictive covariates are used to drive a clustering process that results in homogeneous groups of patients. This step is integrated into a prognostic model to predict response to competing treatments for

↑ back to the schedule ↑

new untreated patients. Finally, a utility-based approach allows us to select the treatment that ensures the larger predicted utility for new patients, based on their genetic profiles. We employed a Bayesian nonparametric model for random partition to build our integrative approach. In particular, we explored the use of the Normalized Generalized Gamma process as cohesion function in a product partition model with covariates. In contrast with existing methods, we jointly estimate model-based clustering and treatment assignment from the data, and hence treatment selection fully accounts for patients' variability.

Annie Sauer Booth

Deep Gaussian Process Surrogates for Computer Experiments

Abstract: This talk provides an overview of Bayesian deep Gaussian processes (DGPs) as surrogate models for computer experiments. Computer experiments are invaluable tools for replacing and/or supplementing direct experimentation, particularly in settings where physical experimentation is restricted by ethical, time, financial, or practicality constraints. Such simulations are necessarily complex and require statistical "surrogate" models, trained on a limited budget of simulator evaluations, which can provide predictions and uncertainty quantification at untried input configurations. Gaussian process (GP) surrogates are the canonical choice, but they are limited by stationarity constraints. DGPs upgrade ordinary GPs through functional composition, in which intermediate GP layers warp the original inputs, providing flexibility to model non-stationary dynamics. In large data settings, we integrate Vecchia approximation for faster computation. In small data settings, we utilize strategic active learning/sequential designs with a variety of objectives including variance reduction, Bayesian optimization, and reliability analysis. We showcase implementation in the "deepgp" package for R on CRAN.

8.30 am Multi-track session 6

For invited sessions, the organizer will chair unless otherwise stated

<table>
<tr><td>Session A</td><td>Invited session: Advances in Latent Variable Modeling
Organizer: Abel Rodriguez
Chair: Alex Jiang</td></tr>
<tr><td>Room</td><td>7A</td></tr>
<tr><td>Elena Erosheva</td><td>Modeling Preferences: A Bayesian Mixture of Finite Mixtures for Rankings and Ratings</td></tr>
<tr><td colspan="2">Abstract: Rankings and ratings are commonly used to express preferences but provide distinct and complementary information. Rankings give ordinal and scale-free comparisons but lack granularity; ratings provide cardinal and granular assessments but may be highly subjective or inconsistent. Collecting and analyzing rankings and ratings jointly has not been performed until recently due to a lack of principled methods. In this work, we propose a flexible, joint statistical model for rankings and ratings under heterogeneous preferences: the Bradley-Terry-Luce-Binomial (BTL-Binomial). We employ a Bayesian mixture of finite mixtures (MFM) approach to estimate heterogeneous preferences, understand the inherent uncertainty, and make accurate decisions based on ranking and ratings jointly. We demonstrate the efficiency and practicality of the BTL-Binomial MFM approach on real and simulated datasets of ranking and rating preferences in peer review and survey data contexts.</td></tr>
<tr><td>Rayleigh Lei</td><td>When the extremes meet: Dynamic spatial voting models for estimating the ideology of U.S. Supreme Court Justices</td></tr>
<tr><td colspan="2">Abstract: Latent factor models are widely used in the social and behavioral science as scaling tools to map discrete multivariate outcomes into low dimensional, continuous scales. In political science, dynamic versions of classical factor models have been widely used to study the evolution of justice's preferences in multi-judge courts. In this paper, we discuss two new classes of dynamic factor model (one that relies on a latent circular space and one that is based on an unfolding of the latent space) that can accommodate voting behaviors in which justices commonly understood to be on opposite ends of the ideological spectrum vote together on a substantial number of otherwise closely-divided opinions. We apply these models to data on</td></tr>
</table>

non-unanimous decisions made the U.S. Supreme Court between 1937 and 2021, and show that both of the new model indicate that there are at least two periods (1949-1952 and 1967-1970) when voting patterns can be better described by our new models.

Donatello Telesca	**Modeling functional brain imaging data with mixed membership models**

Abstract: We characterize functional brain imaging data, obtained via electroencephalograms, as functional data. Due to the highly observed levels of signal heterogeneity, functional regression analysis tends to only explain a small amount of the observed variance. We discuss a novel analytical framework combining regression and mixed membership modeling, to obtain deeper insights into sources of data heterogeneity, and single out important functional features which change with covariate information. Throughout the work, we pay particular attention to issues of robustness, conditional consistency in estimation, as well as adjustments for phase variability. Our proposed methodological framework is based on flexible representations of Gaussian Processes. Time permitting, we will discuss generalizations to more general stochastic processes and metrics.

Session B	Invited session: Innovative Approaches to Model Selection and Clustering Organizer: Antonio Lijoi
Room	4A
Federica Zoe Ricci	**Bayesian temporal biclustering with applications to multi-subject neuroscience studies**

Abstract: We consider the problem of analyzing multivariate time series collected on multiple subjects, with the goal of identifying groups of subjects exhibiting similar trends in their recorded measurements over time as well as time-varying groups of associated measurements. To this end, we propose a Bayesian model for temporal biclustering featuring nested partitions, where a time-invariant partition of subjects induces a time-varying partition of measurements. Our approach allows for data-driven determination of the number of subject and measurement clusters as well as estimation of the number and location of changepoints in measurement partitions. To efficiently perform model fitting and posterior estimation with Markov Chain Monte Carlo, we derive a blocked update of measurements' cluster-assignment sequences. We illustrate the performance of our model in two applications to functional magnetic resonance imaging data and to an electroencephalogram dataset. Results indicate that the proposed model can combine information from potentially many subjects to discover a set of interpretable, dynamic patterns. Experiments on simulated data compare the estimation performance of the proposed model against ground-truth values and other statistical methods, showing that it performs well at identifying ground-truth subject and measurement clusters even when no subject or time dependence is present.

Bryon Aragam

Statistical aspects of nonparametric latent variable models

Abstract: One of the key paradigm shifts in statistical machine learning over the past decade has been the transition from handcrafted features to automated, data-driven representation learning, typically via deep neural networks. As these methods are being used in high-stakes settings such as medicine, health care, law, and finance, where accountability and transparency are not just desirable but often legally required, it has become necessary to place representation learning on a rigorous scientific footing. The statistical foundations of nonparametric latent variable models are revisited and discussed how even basic statistical properties such as identifiability and consistency are surprisingly subtle. New results are also discussed, characterizing the optimal sample complexity for learning simple nonparametric mixtures, which turn out to have a nonstandard super-polynomial bound. With time permitting, applications will end to deep generative models widely used in practice.

Francesco Gaffi

Nonparametric priors for graph matching

Abstract: The graph matching problem consists of establishing correspondences between nodes in two graphs. Specifically, connections are observed in two networks, which share the exact same set of nodes, but the nodes' labeling is unknown, so the object of inference is the bijection that identifies the nodes of the two networks. While graph matching has countless applications, especially in computer vision and bioinformatics, exact solutions are often elusive due to complexity. Even classical estimators that aim to minimize the edge mismatches, allowing for noisy observations, are very difficult to compute. We present a novel modeling idea for graph matching that takes inspiration from the extensive Bayesian nonparametric methodologies developed for random partition models. In particular, we combine a correlated Erdős–Rényi likelihood with a nonparametric prior on the space of random permutations, obtained via an upgrade of the well-known Chinese restaurant process (CRP). Posterior inference is carried out via MCMC algorithms. We believe this to be the first model-based Bayesian nonparametric approach to the graph matching problem. The possible generalizations of the proposed model are several: the CRP prior for random permutations can be replaced by more complex constructions, inducing structural information in the model. In this spirit, we furthermore propose a model extension based on conditional partial exchangeability, in which the cycle structure of the random permutation is partially exchangeable conditionally on the block structure originating the connections. This enforces, a priori, a higher probability of matching with nodes with same connectivity behaviour, without excluding the possibility of matching across-blocks. Joint work with Nathaniel Josephs and Lizhen Lin.

Session C	Invited session: Variational Inference: When, Why and Why Not? Organizer: David Blei
Room	Magna
Charles Margossian	**Variational Inference for Uncertainty Quantification: an Analysis of Trade-offs**

Abstract: Given an intractable distribution p, the problem of variational inference (VI) is to find the best approximation from some more tractable family Q. Commonly, one chooses Q to be a family of factorized distributions (i.e., the mean-field assumption), even though p itself does not factorize. I'll show that this mismatch leads to an impossibility theorem: if p does not factorize, then any factorized approximation q can correctly estimate at most one of the following three measures of uncertainty: (i) the marginal variances, (ii) the marginal precisions, or (iii) the generalized variance (which can be related to the entropy). In practice, the best variational approximation in Q is found by minimizing some divergence $D(q, p)$ between distributions. In this talk, I'll explore how the choice of divergence determines which measure of uncertainty, if any, is correctly estimated by VI.

Tamara Broderick	**Black Box Variational Inference with a Deterministic Objective: Faster, More Accurate, and Even More Black Box**

Abstract: Automatic differentiation variational inference (ADVI) offers fast and easy-to-use posterior approximation in multiple modern probabilistic programming languages. However, its stochastic optimizer lacks clear convergence criteria and requires tuning parameters. Moreover, ADVI inherits the poor posterior uncertainty estimates of mean-field variational Bayes (MFVB). We introduce "deterministic ADVI" (DADVI) to address these issues. DADVI replaces the intractable MFVB objective with a fixed Monte Carlo approximation, a technique known in the stochastic optimization literature as the "sample average approximation" (SAA). By optimizing an approximate but deterministic objective, DADVI can use off-the-shelf second-order optimization, and, unlike standard mean-field ADVI, is amenable to more accurate posterior covariances via linear response (LR). In contrast to existing worst-case theory, we show that, on certain classes of common statistical problems, DADVI and the SAA can perform well with relatively few samples even in very high dimensions, though we also show that such favorable results cannot extend to variational approximations that are too expressive relative to mean-field ADVI. We show on a variety of real-world problems that DADVI reliably finds good solutions with default settings (unlike ADVI) and, together with LR covariances, is typically faster and more accurate than standard ADVI.

Lu Zhang	**Pathfinder and the Geometry of Bayesian Inference**

Abstract: Our recent research has introduced Pathfinder: a parallel quasi-Newton algorithm that can efficiently reaches regions of high probability. Comparative simulation studies reveal Pathfinder's superiority to automatic differentiation variational inference (ADVI) and its competitive performance against dynamic Hamiltonian Monte Carlo (HMC), assessed by the 1-Wasserstein distance. However, Pathfinder's effectiveness is notably influenced by the posterior's geometry. The talk will present case studies illustrating the causes of Pathfinder's sensitivity to the pathological geometry of the posterior and discuss new developments aiming at overcoming these obstacles.

Session D	Invited session: Bayesian Approaches to Statistical and Algorithmic Fairness

↑ back to the schedule ↑

Organizer: Brieuc Lehmann

Room	9A
Ioanna Manolopoulou	**Combining observational data with non-representative randomised data: a case for capturing unfair de-confounding**

Abstract: Building statistical models using non-randomly sampled data is a well-known challenge in statistics, and is especially challenging when any part of the statistical model is not fully identifiable. In causal inference, and in particular in the estimation of heterogeneous treatment effects, this arises when observational data are used which may be affected by unobserved confounding. One approach to correct for such confounding is to combine observational data with randomised experiments. However, when these randomised experiments are not representative of the whole population, the effect of de-confounding will be poor for subsets of the population that fall outside the range of these experiments. Depending on the structure of the model and the nature of the prior distributions used within a Bayesian model, this will be addressed by borrowing information from other parts of the space. In this work, we highlight the importance of building models that can account for uncertainty due to unobserved confounding in regions where no de-confounding is possible. To this end, we embed a combination of randomised and observational data into Bayesian Causal Forests (BCF), and make use of adaptive modular inference to harness as much reliable information from the observational data as possible, without leading to over-confidence in regions of poor identifiability. We implement our methods on a set of simulated and real data examples.

Christos Dimitrakakis	**Fairness is in the eye of the beholder**

Abstract: Even with a fixed definition of fairness, it is hard to agree about whether a given policy satisfies it. This is because a policy only satisfies group fairness constraints with respect to a specific underlying distribution. Once this distribution changes, the policy stops being fair. The obvious solution is to adopt a Bayesian viewpoint, which will allow us to minimise fairness violation in expectation, with respect to some subjective belief. This problem does not go away when we are dealing with individual fairness notions such as meritocracy, especially when we utilise an uncertain predictive model to justify our actions on individuals. However, I argue that the Bayesian approach is especially suitable in this case, because simply

marginalising out model parameters may lead the policy to ignore parameter and outcome correlations across similar individuals.

Silvia Chiappa

Graph-based statistical causality for fairness

Abstract: In this talk, I will discuss the usefulness of graph-based statistical causality for addressing fairness issues in different learning settings. I will cover the following topics: how graph-based statistical causality can be used to consider complex scenarios in the classical classification and regression settings, how graph-based statistical causality can be used to enable the learning of fair policies from offline data in the decision making setting, and how graph-based statistical causality can be used to identify pitfalls of data balancing techniques for mitigating undesired dependencies in the fine-tuning or partial retraining setting.

Kristian Lum

Discussant

Session E

Contributed session: Advances in Scalable Inference
Chair: Alejandra Avalos Pacheco

Room

5A

Matteo Gianella

Consensus Monte Carlo for large spatial datasets

Abstract: A vast literature on spatial and spatio-temporal Bayesian models is available both for georeferenced and areal data. Modelling those types of data is usually accommodated via a covariance function (for georeferenced data) or a neighbouring structure (for areal data). A common issue those models share is that they scale badly with respect to the number of datapoints or areal units. Scalable Bayesian models, especially for georeferenced data, have recently become very popular. Such models try to reduce the number of data points used to sample each element of the spatial process that governs the dependence between data. Methods for the analysis of high-dimensional areal data, on the other hand, are way less investigated. We here propose a Consensus Monte Carlo algorithm for the analysis and clustering of areal datasets. The spatial dependency among areas is modelled via a spatial product partition model, thus inducing clustering among observations which controls and balances local and global spatial dependence. Our novel contribution is the extension to the spatial setting itself, which has not been explored so far in the available literature on Consensus Monte Carlo algorithms. The algorithm we propose is valid under any model assumed for the data within clusters. Joint work with Fernando A. Quintana and Alessandra Guglielmi.

Meini Tang

Fast approximate posterior inference for modeling disease dynamics via state-space models

Abstract: The outbreak of COVID-19 significantly increased the demand for flexible models to better understand and predict disease dynamics. Various models have been used to describe the evolution of cases over time, including self-exciting branching processes like the Hawkes process, Poisson autoregressions, and distributed lag models, also referred to as transfer function models. We present a general class of Bayesian dynamic transfer function models that encompass the aforementioned models as specific cases. We notice the inherent computational challenges to obtain filtering, smoothing and forecasting in this class of models due to their non-Gaussian and/or non-linear nature at one or more levels of their hierarchical structure. To address this issue, we provide methods to facilitate fast and accurate online inference and forecasting. These methods range from linear conjugate Bayes approximations, sequential Monte Carlo and variational Bayes, to hybrid methods that combine some of these approaches. We illustrate our proposed models and related methods for exact and approximate inference in several simulation settings and in the context of the analysis of daily Covid-19 cases at various locations.

Ryan Giordino

The Bayesian Infinitesimal Jackknife for Variance

Abstract: The frequentist variability of Bayesian posterior expectations can provide meaningful measures of uncertainty even when models are misspecified. Classical methods to asymptotically approximate the frequentist covariance of Bayesian estimators such as the Laplace approximation and the nonparametric bootstrap can be practically inconvenient, since the Laplace approximation may require an intractable integral to compute the marginal log posterior, and the bootstrap requires computing the posterior for many different bootstrap datasets. We develop and explore the infinitesimal jackknife (IJ), an alternative method for computing asymptotic frequentist covariance of smooth functionals of exchangeable data, which is based on the "influence function" of robust statistics. We show that the influence

↑ back to the schedule ↑

function for posterior expectations has the form of a simple posterior covariance, and that the IJ covariance estimate is, in turn, easily computed from a single set of posterior samples. Under conditions similar to those required for a Bayesian central limit theorem to apply, we prove that the corresponding IJ covariance estimate is asymptotically equivalent to the Laplace approximation and the bootstrap. In the presence of nuisance parameters that may not obey a central limit theorem, we argue using a von Mises expansion that the IJ covariance is inconsistent, but can remain a good approximation to the limiting frequentist variance. We demonstrate the accuracy and computational benefits of the IJ covariance estimates with simulated and real-world experiments.

Katja Ickstadt	**Scalable Bayesian p-generalized probit and logistic regression**

Abstract: The logit and probit link functions are arguably the two most common choices for binary regression models. Many studies have extended the choice of link functions to avoid possible misspecification and to improve the model fit to the data. We introduce the p-generalized Gaussian distribution (p-GGD) to binary regression in a Bayesian framework. The p-GGD has received considerable attention due to its flexibility in modeling the tails, while generalizing, for instance, over the standard normal distribution where p = 2 or the Laplace distribution where p = 1. Here, we estimate the posterior distributions using Markov Chain Monte Carlo (MCMC) sampling for the model parameters and the link function parameter p. We use simulated and real-world data to verify the effect of different parameters p on the estimation results, and how logistic regression and probit regression can be incorporated into a broader framework. To make our approach scalable in the case of large data, we also incorporate coresets to reduce the data before running the complex and time-consuming MCMC analysis. This allows us to perform very efficient calculations while retaining the original posterior parameter distributions up to little distortions both, in practice, and with theoretical guarantees.

Session F	Contributed session: Advances in Objective Bayes Chair: Luis Carvalho

<table>
<tr><td>Room</td><td>10A</td></tr>
<tr><td>Alessandro Zito</td><td>**Compressive Bayesian non-negative matrix factorization for mutational signature analysis**</td></tr>
<tr><td colspan="2">Abstract: Non-negative matrix factorization (NMF) is widely used in many applications for dimensionality reduction. Inferring an appropriate number of factors for NMF is a challenging problem, and several approaches based on information criteria or sparsity-inducing priors have been proposed. However, inference in these models is often complicated and computationally challenging. In this paper, we introduce a novel methodology for overfitted Bayesian NMF models using "compressive hyperpriors" that force unneeded factors down to negligible values while only imposing mild shrinkage on needed factors. The method is based on using simple semi-conjugate priors to facilitate inference, while setting the strength of the hyperprior in a data-dependent way to achieve this compressive property. We apply our method to mutational signatures analysis in cancer genomics, where we find that it outperforms state-of-the-art alternatives. In particular, we illustrate how our compressive hyperprior enables the use of biologically informed priors on the signatures, yielding significantly improved accuracy. We provide theoretical results establishing the compressive property, and we demonstrate the method in simulations and on real data from a breast cancer application</td></tr>
<tr><td>Dimitris Fouskakis</td><td>**Shrinkage Priors Via Imaginary Data**</td></tr>
<tr><td colspan="2">Abstract: In this work, focus is given in the Bayesian variable selection problem for high-dimensional linear regression problems. The use of shrinkage priors, when the number n of available observations is less than the number p of explanatory variables, is a well-established method, which shares great theoretical and empirical properties. By using imaginary data and shrinkage priors as baseline priors, under the Power-Expected-Posterior (PEP) prior methodology, objective shrinkage priors are being created. In addition, we explore the idea of augmenting the imaginary design matrix in order to make it with orthogonal columns and thus to produce independent PEP-shrinkage priors, based on default baseline priors. Under this setup, properly chosen hyperpriors are placed on the power parameters of the PEP methodology, in order to produced mixtures of independent priors suitable for the variable selection problem when n << p. This second approach provides us with algorithmically flexibility and less time-consuming procedures. We check the theoretical properties of our</td></tr>
</table>

proposed methods and we explore their behavior via simulated studies.

Daniel Roy

On Priors That Assign Infinitesimal Mass And The Resolution Of Longstanding, Open Problems In Statistical Decision Theory

Abstract: We give an exact characterization of admissibility in statistical decision problems in terms of Bayes optimality in a so-called nonstandard extension of the original decision problem, as introduced by Duanmu and Roy (Ann. Statist. 49(4): 2053-2078, 2021). Unlike the consideration of improper priors or other generalized notions of Bayes optimality, the nonstandard extension is distinguished, in part, by having priors that can assign "infinitesimal" mass in a sense that can be made rigorous using results from nonstandard analysis. With these additional priors, we find that, informally speaking, a decision procedure d0 is admissible in the original statistical decision problem if and only if, in the nonstandard extension of the problem, the nonstandard extension of d0 is Bayes optimal among the extensions of standard decision procedures with respect to a nonstandard prior that assigns at least infinitesimal mass to every set in a so-called determining family. We use the above theorem to give further characterizations of admissibility, one related to Blyth's method, one to a condition due to Stein which characterizes admissibility under some regularity assumptions; and finally, a characterization using finitely additive priors in decision problems meeting certain regularity requirements. Our results imply that Blyth's method is a sound and complete method for establishing admissibility. The family of nonstandard priors are also shown to always admit a least favorable prior, yielding a new minimax theorem, equating the standard upper values (minimax risk) to a nonstandard lower value (a nonstandard Bayes risk).

Saptati Datta

Bayes Factor Functions

Abstract: We base Bayes factors on test statistics and describe a generalization of the class of distributions used to model the distribution of the non-centrality of these test statistics under the alternative hypothesis (Bayes Factors Based on Test Statistics, Johnson (2005)). We illustrate how generalizing the class of probability distributions to model non-centrality parameters of test statistics can enhance the support for the true alternative hypothesis when using Bayes factors based on test statistics. This generalization is necessary when aggregating evidence across replicated experimental designs. The choice of hyperparameters differs for evidence collected from multiple studies compared to a single study. We also discuss the interpretability of Bayes factors indexed by prior densities corresponding to standardized effect sizes and the simplification in model specification that is achieved by modeling test statistics rather than the full data distribution. We provide mathematical justification and practical examples to illustrate this approach to the definition of Bayes factor functions.

10.30 am Savage Award Session: Theory and Methods

Room	6A
Chair	Matteo Ruggiero

Mario Beraha	**Bayesian Learning of Random Probability Measures**

Abstract: The study of random probability measures is a lively research topic that has attracted interest from different fields in recent years. We consider random probability measures in the context of Bayesian nonparametrics, where the law of a random probability measure is used as prior distribution, and in the context of distributional data analysis, where, in the simplest setting, the goal is to perform inference given an independent and identically distributed sample from the law of a random probability measure.

In this talk, I will provide an overview of recent advancements in Bayesian nonparametrics, with a particular focus on a novel framework for constructing informative prior distributions for model-based clustering tasks. Specifically, I will discuss the use of repulsive and attractive point process priors, the related distributional theory, and computational algorithms.

Jeremias Knoblauch	**Post-Bayesian Machine Learning**

Abstract: In this talk, I provide my perspective on the efforts to develop inference procedures with Bayesian characteristics that go beyond Bayes' Rule as an epistemological principle. I will explain why these efforts are needed, as well as the forms which they take. Focusing on some of my own contributions to the field, I will trace out the community's most important milestones, as well as some of the challenges that lie ahead.

Jeffrey Negrea

Approximations and scaling limits of Markov chains with applications to MCMC and approximate inference.

Abstract: Markov chains form the basis for efficient exact and approximate inference methods. One may wish to use approximate inference methods when exact ones are too computationally expensive. However theoretical quantification of the approximation error and justification for tuning of approximation parameters are needed to best implement such methods.The contributions of my dissertation include: (i) a perturbation analysis of reversible geometrically ergodic Markov chains, which characterizes the stability of the stationary distribution and rate of convergence under changes in the transition dynamics; (ii) a joint characterization of the optimal proposal scaling and shaping for the random-walk Metropolis algorithm; and (iii) a complete characterization of the statistical asymptotics of stochastic gradient algorithms as methods for approximate inference, with recommendations on how to tune them for accuracy and efficiency. Results are obtained via two key theoretical methods: (1) an analysis of the perturbation sensitivity of Markov chains by viewing their transition kernels as linear operators, and (2) through scaling limits of Markov chains and comparisons to continuous-time processes. Through these results, my dissertation offers insight into the viability of approximate inference methods based on approximations and scaling limits of Markov chains.

Lionel Riou-Durand

On sampling from a log-concave density using kinetic Langevin diffusions

Abstract: Langevin diffusion processes and their discretizations are often used for sampling from a target density. The most convenient framework for assessing the quality of such a sampling scheme corresponds to smooth and strongly log-concave densities defined on Rp. The present work focuses on this framework and studies the behavior of the Monte Carlo algorithm based on discretizations of the kinetic Langevin diffusion. We first prove the geometric mixing property of the kinetic Langevin diffusion with a mixing rate that is optimal in terms of its dependence on the condition number. We then use this result for obtaining improved guarantees of

sampling using the kinetic Langevin Monte Carlo method, when the quality of sampling is measured by the Wasserstein distance. We also consider the situation where the Hessian of the log-density of the target distribution is Lipschitz-continuous. In this case, we introduce a new discretization of the kinetic Langevin diffusion and prove that this leads to a substantial improvement of the upper bound on the sampling error measured in Wasserstein distance.

10.30 am Multi-track session 7

Session A	Invited session: Bayesian Predictivism and Decision Analysis under Model Uncertainty: New Theory, Methodology and Macroeconomic Applications Organizer: Mike West Chair: Antonietta Mira
Room	10A
Emily Tallman	**Bayesian predictive decision synthesis (BPDS): Foundations, theory and applied scope**

Abstract: Decision-guided perspectives on model uncertainty expand traditional statistical thinking about managing, comparing, and combining inferences from sets of models. Bayesian predictive decision synthesis (BPDS) advances conceptual and theoretical foundations at the intersection of model uncertainty and decision theory. BPDS methodology exploits the integration of decision-analytic outcomes as well as predictive performance into the evaluation, comparison and potential combination of candidate models. The theory underlying BPDS involves relaxed entropic tilting, which also reinterprets and extends the Bayesian foundation of related developments of so-called generalized and/or robust Bayes. With an overview of the ideas and framework of BPDS, this talk highlights questions of BPDS model specification and technical aspects of analysis in specific settings, with motivating examples drawn from areas including financial forecasting for portfolio decisions and design/control decision settings.

↑ back to the schedule ↑

Tony Chernis

Decision synthesis in monetary policy

Abstract: Monetary policymakers are tasked with simple, but hard to achieve, objectives. Common objectives are minimizing deviations from an inflation target or stabilizing output growth using the interest rate as the policy lever. The macroeconomy is a sophisticated dynamic system and there are significant uncertainties which complicate modelling. In response, decision makers consider multiple models which provide different predictions and policy recommendations which are then synthesized into a policy decision. In this paper, we use Bayesian predictive decision synthesis (BPDS) to create optimal monetary policy recommendations using multiple VAR models. We show how BPDS weights the models differently from BMA by considering decision-making criteria instead of solely statistical criteria.

Niko Hauzenberger

Predictive density combination using a tree-based synthesis function

Abstract: Bayesian predictive synthesis (BPS) provides a method for combining multiple predictive distributions based on agent/expert opinion analysis theory and encompasses a range of existing density forecast pooling methods. The key ingredient in BPS is a "synthesis" function. This is typically specified parametrically as a dynamic linear regression. In this paper, we develop a nonparametric treatment of the synthesis function using regression trees. We show the advantages of our tree-based approach in two macroeconomic forecasting applications. The first uses density forecasts for GDP growth from the euro area's Survey of Professional Forecasters. The second combines density forecasts of US inflation produced by many regression models involving different predictors. Both applications demonstrate the benefits – in terms of improved forecast accuracy and interpretability – of modeling the synthesis function nonparametrically.

Session B	Invited session: Advanced Bayesian Models for Climate and Energy Market Analysis Organizer: Francesco Ravazzolo.
Room	3A
Francesca Panero	**Forecasting food insecurity with Gaussian processes on auxiliary data**

Abstract: Estimating the share of population that is food insecure is one of the major tasks that governments and international organisations need to address in order to direct their humanitarian aid efficiently. Typically, the estimation is conducted through surveys that are collected via internet, phone or in person. For certain countries, the data collection procedure can be expensive, potentially dangerous and subject to irregular time patterns. To address this, we will forecast food insecurity in different regions of selected countries using auxiliary information which is readily available, such as economical, weather and conflict data. Gaussian processes are a strong candidate to perform this estimation, thanks to their flexibility and interpretability in the description of spatio-temporal data, the integration of expert opinions into the choice of kernels and priors, the natural uncertainty quantification and the possibility of introducing hierarchies at spatial level. We will present the model framework and present our forecasting results on some real datasets, comparing with other approaches in literature. This work is joint with the United Nations World Food Programme Hunger Monitoring Unit.

Luigi Gifuni	**Do high frequency text data help forecast crude oil prices? MF-VAR vs. MIDAS**

Abstract: This paper investigates the predictability of monthly real oil prices when daily and weekly text data are combined with oil market fundamentals. Text data are retrieved from 6,447,630 full oil-related articles featured in The Financial Times, Thomson Reuters and The Independent from January 1982 to June 2021. I show that models containing high-frequency financial and commodity variables do not yield significant improvements on the no-change forecast. In contrast, when text data are used along with commodity variables and oil market fundamentals, the preferred models reduce the MSPE by 18%. However, despite this marginal improvement, gains are low. Indeed, the corresponding models with variables observed at homogeneous frequency, generate similar out-of-sample forecasts in terms of accuracy. I thus conclude that variables sampled at different frequencies do not significantly improve the predictability of monthly real oil prices. This is true for point and density forecasts.

Luca Rossini	**Is the price cap for Gas useful? Evidence from European countries**

Abstract: Since Russia's invasion of Ukraine, many countries have pledged to end or restrict their oil and gas imports to curtail Moscow's revenues and hinder its war effort. Thus, the European ministers agreed to trigger a cap on the gas price. To detect the importance of the price cap for gas, we provide a mixture representation for the gas price to detect the presence of outliers made by a truncated normal distribution and a uniform one. We focus our analysis on Germany and Italy, which are major Russian gas importers by exploiting the response of the different commodities to a gas shock through a Bayesian vector autoregressive (VAR) model. As a result, including a lower gas price cap smooths the impact of a gas shock on electricity prices, while not considering a price cap will increase exponentially this impact. Regarding the other commodities, gas shocks matter in the short and long run when a price cap is not considered.

Session C	Invited session: New Approaches for Bayesian Analysis of Complex Data Incorporating Fourier Space Analysis, Causality and Network Models Organizer: Sharmistha Guha
Room	9A
John Kornak	**An MCMC Approach to Bayesian Image Analysis in Fourier Space**

Abstract: Bayesian methods are commonly applied to solve image analysis problems such as noise-reduction, feature enhancement and object detection. A primary limitation of these approaches is the computational complexity due to the interdependence of neighboring pixels which limits the efficiency of performing full posterior sampling through Markov chain Monte Carlo (MCMC). To alleviate this problem, we develop a new posterior sampling method that is based on modeling the prior and likelihood in the space of the Fourier transform of the image. One

advantage of Fourier-based methods is that a large set of spatially correlated processes in image space can be represented via independent processes over Fourier space. A recent approach known as Bayesian Image Analysis in Fourier Space (or BIFS), has introduced the concept of parameter functions to describe prior expectations about distributional parameters over Fourier space. To date BIFS has relied on Maximum a Posteriori (MAP) estimation for generating posterior estimates; providing just a single point estimate. The work presented here extends BIFS to a posterior sampling approach for BIFS that can explore a range of posterior estimators beyond the MAP estimate. In addition, credible intervals can be provided to measure the uncertainty and variability of the posterior estimates. Computational efficiency of MCMC for BIFS is much improved over that for conventional Bayesian image analysis, and mixing concerns that commonly have to be dealt with in high dimensional Markov chain Monte Carlo sampling problems are avoided.

Yanxun Xu	**Enhancing Decision Making with Causal Inference and Unmeasured Confounders in a Bayesian Framework**

Abstract: Effectively navigating decision-making demands a comprehensive understanding of causal relationships, especially with unmeasured confounders in the environment. Traditional causal inference methods often relies on auxiliary data sources to identify true causal effects, such as instrumental variables or proxies. Unfortunately, such data might be difficult or impractical to acquire in observational studies, leading to potential inaccuracies and incomplete inference. To address this limitation, we propose a novel approach that integrates Bayesian joint modeling with causal inference for effective decision-making under the presence of unmeasured confounding. By taking advantage of proper model design and assumptions, the proposed framework can identify true causal effects without the reliance on additional data sources, thereby leading to more informed and effective decisions in complex real-world observational scenarios.

Sharmistha Guha	**Supervised Modeling of Heterogeneous Networks: Investigating Functional Connectivity Across Various Cognitive Control Tasks**

Abstract: We present a neuroimaging-driven study examining the relationship between functional connectivity across cognitive control domains and cognitive phenotypes, aiming to identify specific brain regions significantly associated with these phenotypes. We propose a generalized linear modeling framework incorporating multiple network responses and predictors, allowing for diverse interconnections between edges. Leveraging hierarchical Bayesian modeling, our approach estimates regression coefficients and identifies predictor-linked nodes with precise uncertainty quantification. Empirical investigations, including simulation studies and functional connectivity data analysis, demonstrate our framework's superior performance compared to competitors.

Session D	Invited session: Bayesian Statistical Demography Organizer: Zehang Richard Li
Room	Magna
Adrian Raftery	**Bayesian Forecasting of International Migration**

Abstract: Population projections have traditionally been deterministic, but since 2015 the UN has based its official projections for all countries on a Bayesian method developed by our group at the University of Washington. This does not yet account for uncertainty about international migration. Minimally, to do so requires probabilistic projections of net international migration. Azose and Raftery (2015) introduced such a model, but it does not account for the changing age composition of the population. We propose a way of doing so. We show that migration rates can be age-standardized in a remarkably simple way, using a quantity we call the migration age structure index (MASI). We then use a Bayesian hierarchical model to generate joint probabilistic forecasts of total and age- and sex-specific net migration rates from 2025 to 2100, accounting for age structure. We find that accounting for population age structure in historic and forecast net migration rates leads to narrower prediction intervals by the end of the century for most countries. Also, accounting for population age structure leads to less out-migration among countries with rapidly aging populations that are forecast to contract most rapidly by the end of the century. This leads to less drastic population declines than are forecast without accounting for population age structure. This is joint work with Nathan Welch and Hana Sevcikova.

↑ back to the schedule ↑

Andrea Riebler

Education-specific differences in time trends of age-stratified backpain data of US adults

Abstract: Back pain has been a significant health problem worldwide. Adults with lower educational attainment are more likely to be affected by back pain, and it has been suspected that the educational disparities widened for recent birth cohorts. In this study, we analyze back pain data obtained from the 1997 to 2018 National Health Interview Survey (NHIS) for US adults aged 30 to 69 years. We use multivariate age-period-cohort models to investigate the effect of four different education levels on back pain across the three time axes: age, period and birth cohort. Smoothing priors are used to stabilise estimates of time trends, and joint hyperpriors are chosen based on expert knowledge. The complex survey design is accounted for by adapting ideas of Fay-Herriot area modelling. We find an increasing effect across age groups indicating that the effect that people with lower levels of education are more likely to experience backpain is increasing with age. This effect is strongest for people with less than a high school degree. Similar results are found across cohorts indicating that the gap between eduction levels is getting wider for more recent generations. This is joint work with Anna Zajacova, Snehalata Huzurbazar, Markus Trætli and Ingeborg Hem Sørmoen.

Maciej Danko

Modeling migration flows in Europe and the role of freedom of movement in shaping migration patterns

Abstract: While migration has become a key driving force of population change in Europe, the quality and availability of migration data remains highly heterogeneous across countries. In this study we develop a Bayesian hierarchical model to analyze international migration flows in Europe from 2002 to 2018. The approach addresses data quality issues, harmonizes migration definitions, and merges administrative and EU Labor Force Survey data. Then we examine the impact of freedom of movement and labor market access in destination countries. As an illustrative example, we discuss the case of Poland, representing the A-8 countries that joined the EU in 2004. More specifically, we investigate how the free movement of workers and labor market access influence migration patterns. Our findings suggest that the largest component of emigration flows from Poland shifted from Germany to the UK and other countries following Poland's 2004 EU accession. However, in 2011, the pattern reversed when Germany opened its labor market to A-8 country workers, resulting in a significant increase in immigration.

↑ back to the schedule ↑

Session E	Contributed session: Advances in Sampling-Based Inference Chair: Michael Minyi Zhang
Room	7A
Alexandros Beskos	**Manifold Markov Chain Monte Carlo Methods for Bayesian Inference in Diffusion Models**

Abstract: Bayesian inference for nonlinear diffusions, observed at discrete times, is a challenging task that has prompted the development of a number of algorithms, mainly within the computational statistics community. We propose a new direction, and accompanying methodology - borrowing ideas from statistical physics and computational chemistry - for inferring the posterior distribution of latent diffusion paths and model parameters, given observations of the process. Joint configurations of the underlying process noise and of parameters, mapping onto diffusion paths consistent with observations, form an implicitly defined manifold. Then, by making use of a constrained Hamiltonian Monte Carlo algorithm on the embedded manifold, we are able to perform computationally efficient inference for a class of discretely observed diffusion models. Critically, in contrast with other approaches proposed in the literature, our methodology is highly automated, requiring minimal user intervention and applying alike in a range of settings, including: elliptic or hypo-elliptic systems; observations with or without noise; linear or non-linear observation operators. Exploiting Markovianity, we propose a variant of the method with complexity that scales linearly in the resolution of path discretisation and the number of observation times. Python code reproducing the results is available at http://doi.org/10.5281/zenodo.5796148.

Fernando Antonio Zepeda Herrera	**D-ALPS: A Non-Reversible extension of the Annealed Leap-Point Sampler**

Abstract: It is well known that when a target distribution is multimodal, traditional MCMC algorithms such as RWM, MALA or HMC get often trapped in local modal regions, resulting in poor space exploration and, consequently, unreliable Monte Carlo estimates. There is a growing literature of alternative methods proposed to address this challenge, with one of its main streams being Parallel Tempering (Geyer, 1991), abbreviated as PT. Many variants arising from this meta-algorithm have been proposed like the Adaptive version of Miasojedow et al (2013), the Optimized Paths proposal of Syed et al (2021), or the use of Variational References put forward by Surjanovic et al (2022), to name just a few. Another stream of research has made use of optimization techniques to find modes and leverage this information to better perform inference on multimodal targets (Tjelmeland and Hegstad, 2001; Zhou, 2011; Pompe et al 2020). The Annealed Leap Point Sampler (ALPS) of Tawn et al (2021) is a novel algorithm that bridges these areas of research and incorporates modal information into a Reversible PT setting. Here, we present D-ALPS, a Non-Reversible counterpart that makes use of the deterministically alternating replica exchange strategy used originally by Okabe et al (2001). As proved by Syed et al (2022), Non-Reversible PT is an improvement over Reversible PT. We verify this empirically by contrasting the performance of ALPS with that of D-ALPS. We further extend D-ALPS so that, besides tackling multimodal targets, it is able to perform transdimensional sampling. There is a clear analogy between well-separated modal regions of a common space and the disjoint nature of transdimensional state spaces. Moreover, some multimodal distributions arise in the context of model selection or model averaging, which are very frequently transdimensional tasks. One would therefore like to have an algorithm capable of addressing both situations.

Paul Fearnhead	**Stochastic gradient PDMP samplers**

Abstract: Recently it has been shown that piecewise deterministic Markov processes (PDMPs) can be used as an alternative to MCMC. The idea is to simulate a PDMP that has been designed to have the posterior distribution as its stationary distribution, with there being simple rules for specifying the dynamics of the PDMP to enable this. Furthermore, the PDMP samplers are non-reversible, and thus can mix better than reversible MCMC in high-dimensions, and they can be implemented with sub-sampling ideas to reduce the per-iteration cost. Unfortunately, whilst the dynamics of these PDMPs are easy to define, simulating a continuous-time realisation is challenging in general. To overcome this, we will show we can approximately simulate the dynamics of a PDMP with subsampling. The resulting algorithm is easy to implement, and is computationally efficient as it involves just accessing one or two data points per iteration. The resulting algorithm can be viewed as an alternative to the popular stochastic-gradient Langevin dynamics (SGLD) algorithm, but with a PDMP replacing the Langevin dynamics. The resulting stochastic gradient PDMP algorithm has a number of advantages over SGLD: the underlying dynamics are non-reversible; the algorithm is more stable; and we can leverage its continuous trajectories to more naturally incorporate model selection. This is joint work with Sebastiano Grazzi, Chris Nemeth and Gareth Roberts

Jun Yang	**Stereographic Barker's MCMC Proposal: Efficiency and Robustness at Your Disposal**

Abstract: We introduce a new family of robust gradient-based MCMC samplers under the framework of Stereographic MCMC (Yang et al., 2022) which maps the original high dimensional problem in Euclidean space onto a sphere. Compared with the existing Stereographic Projection Sampler (SPS) which is of a random-walk Metropolis type algorithm, our new family of samplers is gradient-based using the Barker proposal (Livingstone and Zanella, 2022), which improves SPS in high dimensions and is robust to tuning. Meanwhile, the proposed algorithms enjoy all the good properties of SPS, such as uniform ergodicity for a large class of heavy and light-tailed distributions and ``blessings of dimensionality''.

Session F	Contributed session: Clinical Research Chair: Juhee Lee
Room	2A
Massimiliano Russo	**A cautious decision-theoretic use of auxiliary outcomes in clinical trials**

Abstract: Most clinical trials collect data on multiple outcomes, such as overall survival (OS), progression-free survival (PFS), and response to treatment (RT). In most cases, however, study designs only use primary outcome data for interim and final decision-making. In several disease settings, clinically relevant outcomes, for example OS, become available years after patient enrollment. Moreover, the effects of experimental treatments on OS might be less pronounced compared to auxiliary outcomes such as RT. We develop a Bayesian decision-theoretic framework that leverages the joint analysis of primary and auxiliary outcomes for interim and final decision-making. The framework allows investigators to control standard frequentist operating characteristics, such as the type I error rate. The approach can be used with auxiliary outcomes from emerging technologies such as circulating tumor assays.

False positive rates and other frequentist operating characteristics are rigorously controlled without any assumption about the concordance between primary and auxiliary outcomes. We discuss algorithms to implement this decision-theoretic approach and show that incorporating auxiliary information into interim and final decision-making can lead to relevant efficiency gains according to established and interpretable metrics. We illustrate the approach through subgroup analyses and multi-stage clinical trial designs.

Victor Medina-Olivares	**Dynamic Copulas for Joint Models of Longitudinal and Survival Data**

Abstract: Joint modeling of longitudinal and survival data is an active area of research. Joint models simultaneously analyze repeated measures and time-to-event data and are instrumental, e.g., in medical research, for studying disease progression and the time until an event, such as death or disease recurrence. They have also been recently introduced in other domains like social sciences, economics, and engineering. Joint models offer several benefits over benchmark models, including a principled way to include endogenous time-varying covariates and a more accurate and robust forecasting framework. A typical joint model comprises two sub-models: one for the longitudinal process and another for the survival process. Given a shared latent structure, these sub-models are assumed to be conditionally independent, allowing for a correlation between them. A prevalent method to link both processes is through normally distributed shared random effects. Such models have been well-explored using Bayesian estimation, and scalability to large datasets can be handled computationally conveniently using integrated nested Laplace approximations (INLA). However, normally distributed shared random effects are often too restrictive to model more complex and potentially more realistic scenarios. To overcome these limitations, we introduce a flexible copula representation to associate the longitudinal and survival processes. This framework extends previous works by allowing the copula parameters to change over time, creating a dynamic link between both processes that can account for dependency on the event's life cycle stage. Despite the enhanced modeling capabilities, these flexible representations result in highly parameterized models, making the estimation process through MCMC schemes computationally expensive or impractical for large datasets. To address this, we present a factorized Gaussian variational approximation that allows efficient handling of the high-dimensional problem through a low-dimensional factor covariance structure. This method enables computationally efficient estimation of model parameters while also providing a measure of uncertainty. The benefits of this approach are demonstrated through a simulation study and an application to a large dataset.

Ray Bai

A Bayesian Semiparametric Approach to Informative Cluster Size with Application to Kidney Transplantation

Abstract: In biomedical studies, it is common to observe clustered outcomes. In these settings, informative cluster size (ICS) is often present. ICS occurs when the cluster sizes influence the outcomes of interest. If nonignorable cluster size is not properly accounted for, then it can lead to biases and incorrect inferences. Unfortunately, there is very little work on Bayesian approaches to ICS, and Bayesian multi-level models almost never correct for ICS in inference. In this talk, we propose a novel semiparametric Bayesian approach for ICS. We introduce a joint model for the primary regression and the cluster size that properly accounts for ICS. Using Dirichlet process mixture models, our joint model allows for flexible modeling of both the primary outcome of interest and the count distribution for the cluster size. We validate our method through simulation studies and an application to kidney transplantation where patient volume is correlated with survival time.

Youngsoo Baek

Bayesian semi-supervised learning for predictive modeling of electronic health records (EHR) clinical outcomes

Abstract: Electronic health records (EHR) have emerged as an important data resource for conducting observational studies in healthcare and medicine. In recent years, predictive modeling has become a central focus motivated by precision medicine and improving clinician decision making for individual patients. A common problem in developing predictive models in EHR data is that predictor information is easy to obtain at scale, but the outcome is expensive or time consuming to collect. In this setting, semi-supervised (SS) learning algorithms can be used to improve efficiency of prediction models by leveraging large amounts of unlabeled data to extract features from the predictors that are useful for predicting within the small subset of labeled data. Though numerous SS learning methods have been proposed in recent years, these methods often implicitly assume that the missingness mechanism is missing completely at random, which is often inappropriate when modeling clinical outcomes in the EHR. In this work, we introduce a fully Bayesian SS

↑ back to the schedule ↑

learning algorithm using a selection model specification that allows for data to be both missing at random and missing not at random. We evaluate the method thoroughly in simulation studies and in a real-world EHR example predicting psychological distress in glaucoma patients.

↑ back to the schedule ↑

1:30 pm Keynote Lecture: Omiros Papaspiliopoulos

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

Speaker

Omiros Papaspiliopoulos
Professor of Decision Sciences,
Bocconi University

Chair: Judith Rousseau

From random effects to random graphs: large scale inference for mixed models

Abstract: Generalized bi-linear mixed models are the workhorse of applied Statistics and they are used for varied tasks such as small area estimation, item response theory, recommendation and analysis of networks. In modern applications, from political science to electronic marketing, it is common that both the size of the data and the number of random effects are large. As a result, there is an increasing need amongst scientists to have access to inferential and computational frameworks that can cope with large-scale problems and provide reasonable uncertainty quantification. The posterior dependence in these models is sparse but in an unstructured way that relates to random graphs. As a result, popular inferential approaches and implementations with reasonable uncertainty quantification (such as lmer or INLA), even though highly optimized they eventually have costs that scale polynomially with the size of data and model. The talk will provide an overview of recent methodologies that have provably linear computational complexity and provide provably good uncertainty quantification. The methods include collapsed Gibbs samplers, partially factorized variational inference and approximate conjugent

gradient Markov chain Monte Carlo. The desired properties are obtained leveraging spectral random graph theory, and a convergence-accuracy duality for variational inference that is of broader interest.

3:00 pm Multi-track session 8

For invited sessions, the organizer will chair unless otherwise stated

Session A	Invited session: Recent Highlights from Biometrika Organizer: Paul Fearnhead
Room	6A
Chris Nemeth	**Semi-exact control functionals from Sard’s method**

Abstract: A novel control variate technique is proposed for the post-processing of Markov chain Monte Carlo output, based on both Stein’s method and an approach to numerical integration due to Sard. The resulting estimators of posterior expected quantities of interest are proven to be polynomially exact in the Gaussian context, while empirical results suggest that the estimators approximate a Gaussian cubature method near the Bernstein–von Mises limit. The main theoretical result establishes a bias-correction property in settings where the Markov chain does not leave the posterior invariant. Empirical results across a selection of Bayesian inference tasks are presented. This is joint work with Leaht South, Toni Karvonen, Mark Girolami and Chris Oates.

Tommaso Rigon	**A generalized Bayesian framework for probabilistic clustering**

Abstract: Loss-based clustering methods, such as k-means clustering and its variants, are standard tools for finding groups in data. However, the lack of quantification of uncertainty in the estimated clusters is a disadvantage. Model-based clustering based on mixture models provides an alternative approach, but such methods face computational problems and are highly sensitive to the choice of kernel. In this article we propose a generalized Bayes framework that bridges between these paradigms through the use of Gibbs posteriors. In conducting Bayesian updating, the loglikelihood is replaced by a loss function for clustering, leading to a rich family of clustering methods. The Gibbs posterior represents a coherent updating of Bayesian beliefs without needing to specify a likelihood for the data, and can be used for characterizing uncertainty in clustering. We consider losses based on Bregman divergence and pairwise similarities, and develop efficient deterministic algorithms for point estimation along with sampling algorithms for uncertainty quantification. Several existing clustering algorithms, including k-means, can be interpreted as generalized Bayes estimators in our framework, and thus we provide a method of uncertainty quantification for these approaches, allowing, for example, calculation of the probability that a data point is well clustered. This is joint work with Amy Herring and David Dunson

Giacomo Zanella	**Optimal design of the Barker proposal and other locally balanced Metropolis–Hastings algorithms**

Abstract: We study the class of first-order locally balanced Metropolis–Hastings algorithms introduced in Livingstone & Zanella (2022). To choose a specific algorithm within the class, the user must select a balancing function and a noise distribution for the proposal increment. Popular choices within the class are the Metropolis-adjusted Langevin algorithm and the recently introduced Barker proposal. We first establish a general limiting optimal acceptance rate of 57% and scaling of the step-size that is the cube-root of dimension, as dimension tends to infinity, among all members of the class under mild smoothness assumptions and when the target distribution for the algorithm is of product form. In particular, we obtain an explicit expression for the asymptotic efficiency of an arbitrary algorithm in the class, as measured by expected squared jumping distance. We then consider how to optimize this expression under various constraints. We derive an optimal choice of noise distribution for the Barker proposal, an optimal choice of balancing function under a Gaussian noise distribution, and an optimal choice of first-order locally balanced algorithm among the entire class, which turns out to depend on the specific target distribution. Numerical simulations confirm our theoretical findings, and in particular, show that a bimodal choice of noise distribution in the Barker proposal gives rise to a practical algorithm that is consistently more efficient than the original Gaussian version. This is joint work with Jure Vogrinc and Sam Livingstone.

↑ back to the schedule ↑

Session B	Invited session: Bayesian Sampling as an Optimisation Problem Organizer: Chris Nemeth
Room	2A
Sinho Chewi	**Variational inference via Wasserstein gradient flows**

Abstract: Variational inference (VI), which seeks to approximate the Bayesian posterior by a more tractable distribution within a variational family, has been widely advocated as a scalable alternative to MCMC. However, obtaining non-asymptotic convergence guarantees has been a longstanding challenge. In this talk, I will argue that viewing this problem as optimization over the Wasserstein space of probability measures equipped with the optimal transport metric leads to the design of principled algorithms which exhibit strong practical performance and are backed by rigorous theory. In particular, we address Gaussian VI, as well as (non-parametric) mean-field VI.

Anna Koraba	**Sampling through optimization of discrepancies**

Abstract: Sampling from a target measure when only partial information is available (e.g. unnormalized density as in Bayesian inference, or true samples as in generative modeling) is a fundamental problem in computational statistics and machine learning. The sampling problem can be formulated as an optimization over the space of probability distributions of a well-chosen discrepancy (e.g. a divergence or distance). In this talk, we'll discuss several properties of sampling algorithms for some choices of discrepancies (well-known ones, or novel proxies), both regarding their optimization and quantization aspects.

Louis Sharrock	**Learning-Rate-Free Methods for Scalable Bayesian Inference**

Abstract: In recent years, particle-based variational inference (ParVI) algorithms such as Stein variational gradient descent (SVGD) have gained popularity as scalable alternative methods for Bayesian inference. Unfortunately, such methods invariably depend on hyperparameters such as the learning rate, which must be carefully tuned by practitioners in order to ensure convergence to the target measure at a suitable rate. In this talk, we introduce several new particle-based sampling algorithms which are entirely learning-rate free, leveraging ideas from convex optimisation and the viewpoint of sampling as an optimisation problem over the space of probability measures. We discuss how to adapt these algorithms for sampling in constrained domains. We then illustrate their performance on a range of numerical examples, demonstrating comparable performance to other ParVI algorithms with no need to tune a learning rate.

Session C	Invited session: Innovations in Spatial and Spatiotemporal Statistics and its Applications Organizers: Lu Zhang, Aritra Halder, and Didong Li
Room	7A
Toryn Schafer	**Bayesian models for complex animal movement data**

Abstract: The evolution of ecological technology has ushered in an era of abundant and diverse wildlife movement data, necessitating robust modeling approaches. This presentation introduces Bayesian methodologies tailored to tackle the scientific challenges posed by these datasets. I highlight multiscale data collected from remote wildlife devices including high-dimensional accelerometers and video collars and GPS based locations. Emphasis is placed on addressing nonlinear interactions among organisms and their environment, accounting for non-stationarity, managing multiple data sources, and handling instances of missing-not-at-random. The proposed Bayesian models offer a systematic framework to extract insights into animal behavior, decision-making processes, and ecological dynamics. By integrating prior scientific knowledge and quantifying uncertainty, these models contribute to a more rigorous understanding of the complexities inherent in wildlife movement data analysis.

Pulong Ma	**Residual Treed Gaussian Processes**

Abstract: Gaussian processes (GP) enjoy wide popularity in spatial statistics, uncertainty quantification, and machine learning. With the advance of measurement technologies and increasing computing power, large numbers of measurements and large-scale numerical simulations make traditional GP models and computational strategies inadequate in dealing with spatially heterogeneous and big data, especially in multi-dimensional domains. In recent years, several multi-scale or tree-based extensions of the GP have been introduced to model spatial nonstationarity and/or achieve scalable computation. In this talk, I introduce a new Bayesian tree-based GP inference framework called residual treed GP (ResTGP). ResTGP combines key features of the treed GP (Gramacy and Lee 2008) and the multi-resolution GP (Fox and Dunson 2012), thereby enjoying the computational efficiency of the formal and the flexibility of the latter. Our main idea is to decompose a Gaussian process as well as the data at a cascade of resolutions across locations through iteratively computing predictive and residual processes, thereby characterizing the underlying covariance structure and achieving divide-and-conquer on the data points simultaneously. We also introduce a new computational strategy for Bayesian inference for ResTGP that does not rely on Metropolis-Hastings based stochastic tree search algorithms but is based on recursive message passing. This is joint work with Li Ma at Duke University.

Harrison Quick	**Reliable event rates for disease mapping**

Abstract: When analyzing spatially referenced event data, the criteria for declaring rates as "reliable" is still a matter of dispute. What these varying criteria have in common, however, is that they are rarely satisfied for crude estimates in small area analysis settings, prompting the use of spatial models to improve reliability. While reasonable, recent work has quantified the extent to which popular models from the spatial statistics literature can overwhelm the information contained in the data, leading to oversmoothing. Here, we begin by providing a definition for a "reliable" estimate for event rates that can be used for crude and model-based estimates and allows for discrete and continuous statements of reliability. We then construct a spatial Bayesian framework that allows users to infuse prior information into their models to improve reliability while also guarding against oversmoothing. We apply our approach to county-level birth data from Pennsylvania, highlighting the effect of oversmoothing in spatial models and how our approach can allow users to better focus their attention to areas where sufficient data exists to drive inferential decisions. We then conclude with a brief discussion of how this definition of reliability can be used in the design of small area studies.

Rajarshi Guhaniyogi	**Discussant**

Session D	Invited session: Bayesian Nonlinear Inverse Problems Organizer: Botond Szabo

↑ back to the schedule ↑

Room	9A
Aad van der Vaart	**Linear methods for nonlinear inverse problems**

Abstract: We consider the recovery of an unknown function f from a noisy observation u_f of the solution to a partial differential equation that has f as a parameter or boundary function. The challenging, but realistic, case is that the forward map f -> u_f is nonlinear, making this into a nonlinear inverse problem. We follow a standard (nonparametric) Bayesian approach, thus regularising the solution of the inverse problem through a prior. However, we reformulate the problem as a combination of an embedded Bayesian linear problem and an analytic nonlinear problem, thus making it possible to obtain the posterior distribution using known and computationally efficient approaches for linear inverse problems in combination with numerical methods to map back to the original nonlinear problem. We consider several examples, including the Schrödinger and Darcy equations. After reviewing results for linear problems, we present contraction rates for the posterior distribution and coverage of credible sets for the nonlinear problems. We also discuss distributed posteriors to further alleviate the computational burden. [Joint work with Geerten Koers (TU Delft) and Botond Szabó (Bocconi, Milano).]

Fanny Seizilles	**The Bayesian approach to inverse Robin problems**

Abstract: We investigate the Bayesian approach to certain elliptic boundary value problems of determining a Robin coefficient on a hidden part of the boundary from Cauchy data on the observable part. Such a nonlinear inverse problem arises naturally in the initialisation of large-scale ice sheet models. In this talk we will specifically focus on the computational routine to estimate posterior densities for the Robin coefficient, as illustrated in https://arxiv.org/abs/2311.17542.

The Bayesian approach is motivated for a prototypical Robin inverse problem by showing that the posterior mean converges in probability to the data-generating ground truth as the number of observations increases. Related to the stability theory for inverse Robin problems, a logarithmic convergence rate for Sobolev-regular Robin coefficients is established, whereas for analytic coefficients an algebraic rate can be attained. Our numerical results on synthetic data illustrate the convergence property in two observation settings. (Joint work with Aksel Kaastrup Rasmussen, Ieva Kazlauskaite and Mark Girolami).

↑ back to the schedule ↑

Kolyan Ray	**Bayesian nonparametric inference in a McKean-Vlasov model**

Abstract: We study nonparametric estimation of the interaction term in a McKean-Vlasov model where noisy observations are drawn from the nonlinear parabolic PDE arising in the mean-field limit as the number particles grows to infinity. In this model, the long-time invariant state can be uninformative about the interaction potential. We therefore show that under certain regularity conditions on the initial state, the short-time behaviour of this system contains sufficient information to consistently recover the interaction potential using Gaussian process priors. This involves establishing a stability-type estimate for this PDE to solve the resulting inverse problem. This is joint work with Richard Nickl and Greg Pavliotis.

Session E	Invited session: Bayesian Decision Making and Learning: Theoretical Foundations and Approximate Inference Organizer: Julyan Arbel
Room	Magna
Florence Forbes	**Bayesian experimental design - a sequential Monte Carlo folk-song sung to a stochastic optimization new tune**

Abstract: We propose a new procedure for Bayesian experimental design, that performs sequential design optimization by simultaneously providing accurate estimates of successive posterior distributions for parameter inference. The sequential design process is carried out via a contrastive estimation principle, using stochastic optimization and Sequential Monte Carlo (SMC) samplers to maximise the Expected Information Gain (EIG). As larger information gains are obtained for larger distances between successive posterior distributions, this EIG objective may worsen classical SMC performance. To handle this issue, tempering is proposed to have both a large information gain and an accurate SMC sampling, that we show is crucial for performance. This novel combination of stochastic optimization and tempered SMC allows to jointly handle design optimization and parameter inference. We provide a proof that the obtained optimal design estimators benefit from some consistency property. Numerical experiments confirm the potential of the approach, which outperforms other recent existing procedures.

Thomas Möllenhoff

Variational Learning for Large Deep Networks

Abstract: We present extensive evidence against the common belief that variational Bayesian learning is ineffective for large neural networks. First, we show that a popular deep learning method called sharpness-aware minimization (SAM) solves an optimal convex relaxation of the variational Bayesian objective. Then, we demonstrate that a principled optimization of the variational objective with an Improved Variational Online Newton method (IVON) can consistently match or outperforms Adam for training large networks such as GPT-2 and ResNets from scratch. IVON's computational costs are nearly identical to Adam but its predictive uncertainty is better. We show several new use cases of variational learning where we improve fine-tuning and model merging in Large Language Models, accurately predict generalization error, and faithfully estimate sensitivity to data. We find overwhelming evidence in support of effectiveness of variational methods for deep learning.

Hien Nguyen

Bayesian Likelihood Free Inference using Mixtures of Experts

Abstract: This work presents an extension of Bayesian Synthetic Likelihood (BSL) methods to encompass non-Gaussian approximations of the likelihood function. Within this framework, we introduce a class of neural network models known as Mixtures of Experts (MoEs). These serve as surrogate likelihoods, exhibiting beneficial properties in terms of approximation theory. Furthermore, MoEs can be estimated using approaches based on the Expectation-Maximization algorithm, such as the Gaussian Locally Linear Mapping model estimators that we employ in this work. We provide theoretical results regarding our procedure’s capability to estimate and approximate a broad class of likelihood functions. Through a series of simulations, we show that our approach can be a viable alternative to existing BSL variants, demonstrating potential improvements in both the accuracy of posterior approximation and computational efficiency.

<table>
<tr><td>Judith Rousseau</td><td>Discussant</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Session F</td><td>Contributed session: Advances in Computational Statistics.
Chair: Dario Plumbo</td></tr>
<tr><td>Room</td><td>4A</td></tr>
<tr><td>Sameer Deshpande</td><td>Oblique BART</td></tr>
<tr><td colspan="2">Abstract: Regression trees built using axis-aligned decision rules have enjoyed considerable practical and theoretical success. However, recent work has shown that oblique trees built using linear combinations of features enjoy strong theoretical support and often display predictive performance superior to that of axis-aligned trees and similar to that of neural networks. To explore the extent to which such phenomena carry over to Bayesian Additive Regression Trees, we developed and implemented an oblique version of BART based on trees that recursively partition the unit square with random hyperplanes. We introduce several proposal distributions for random hyperplanes that involve rapidly (i) sampling points from linearly constrained polytopes and (ii) randomly drawing the normal direction of each hyperplane. We compare the mixing rates of oblique BART implemented with these different proposals and also perform extensive experiments to compare oblique BART's predictive power to that of the basic axis-aligned BART.</td></tr>
<tr><td>Mikolaj Kasprzak</td><td>How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences</td></tr>
<tr><td colspan="2">Abstract: The Laplace approximation is a popular method for providing posterior mean and variance estimates. But can we trust these estimates for practical use? One might consider using rate-of-convergence bounds for the Bayesian Central Limit Theorem (BCLT) to provide quality guarantees for the Laplace approximation. But the bounds in existing versions of the BCLT either: require knowing the true data-generating parameter, are asymptotic in the number of samples, do not control the Bayesian posterior mean, or apply only to narrow classes of models. Our work provides the first closed-form, finite-sample quality bounds for the Laplace approximation that simultaneously (1) do not require knowing the true parameter, (2) control posterior means and variances, and (3) apply generally to models that satisfy</td></tr>
</table>

the conditions of the asymptotic BCLT. In fact, our bounds work even in the presence of misspecification. We compute exact constants in our bounds for a variety of standard models, including logistic regression, and numerically demonstrate their utility. We provide a framework for analysis of more complex models. This is joint work with Ryan Giordano (UC Berkeley) and Tamara Broderick (MIT). A preprint is available here: arXiv:2209.14992.

Helga Wagner	**Flexible Bayesian treatment effects for panel outcomes**

Abstract: Identification and estimation of treatment effects is challenging due to unobserved confounding if treatment selection is endogeneous. Bayesian approaches to treatment effects estimation rely on a joint model for selection and the potential outcomes and thus allows inference on effects of observed confounders on selection as well as heterogeneity of treatment effects. For the analysis of treatment effects on longitudinally observed outcomes dependence of treatment selection and the potential outcomes as well as the longitudinal dependence in the potential outcomes has to be taken into account. Motivated by our empirical analysis on the effect of the length of maternity leave on mothers' incomes after their return to work, we extend the recently proposed bifactor treatment effects model to accommodate zero earnings, which are observed for a relevant proportion of mothers. Additionally, to capture treatment effects as well as covariate effects appropriately we model them as time-varying parameters with priors that allow shrinkage to time-constant or even zero covariate effects to avoid overspecification.

Andrea Pandolfi	**Sparse linear algebra methods for high-dimensional GLMMs**

Abstract: Generalized linear mixed models (GLMMs) are a widely used tool in statistical analysis. Both for Bayesian and frequentist inference, the main bottleneck of many computational approaches (e.g. Laplace approximations, Gibbs Sampling, Variational inferences, etc) lies in the inversion of high dimensional precision matrices. Such matrices are typically very sparse; however, the sparsity pattern resembles a K-partite graph, which does not lend itself well to default sparse linear algebra techniques. Notably, we prove that, for typical GLMMs, the Cholesky factor is dense even when the original precision is sparse. We thus turn to approximate linear algebra techniques, in particular to conjugate gradient and incomplete Cholesky methods. Through a detailed analysis of the spectrum of GLMMs precision matrices, we can borrow results from random graph theory to show that these methods typically

achieve a fixed approximation error with a total cost that scales linearly with the number of observations. Numerical illustrations with both real and simulated data confirm the theoretical findings. Joint work with Papaspiliopoulos O. and Zanella G.

Session G	Contributed session: Bayes in Industry Chair: Refik Soyer
Room	3A
Giulio Grossi	**SMaC: Spatial Matrix Completion method**

Abstract: Synthetic control methods are commonly used in panel data settings to evaluate the effect of an intervention. In many of these cases, the treated and control time series correspond to spatial areas such as regions or neighborhoods. We work in a setting where a treatment is applied at a given location and its effect can emanate across space. Then, an area of a certain size around the intervention point is considered to be the treated area. Synthetic control methods can be used to evaluate the effect that the treatment had in the treated area, but it is often unclear how far the treatment's effect propagates. Therefore, researchers might consider treated areas of different sizes and apply synthetic control methods separately for each one of them. However, this approach ignores the spatial structure of the data, and can lead to efficiency loss in spatial settings. We propose to deal with these issues by developing a Bayesian spatial matrix completion framework that allows us to predict the missing potential outcomes in the different areas around the intervention point while accounting for the spatial structure of the data. Specifically, the missing time series in the absence of treatment for the treated areas of all sizes are imputed using a weighted average of control time series, where the weights are assumed to vary smoothly over space according to a Gaussian process. Our motivating application is the construction of the first line of the Florentine tramway, which could have had an effect on the prevalence of businesses in the neighborhood of the construction site, and at various distances from the tramway stops.

Rosangela Helena Loschi	**Dynamic multivariate Gamma-Gamma general path model: An alternative approach to time-variant degradation rates**

Abstract: We introduce a general path Gamma-Gamma model for degradation

↑ back to the schedule ↑

measures, related to different inspection times functions, obtaining flexible forms of degradation paths. One important contribution of the proposed model is the way the degradation rate is modeled. It is composed of two random components: one random effect quantifying the specific features of each device and a dynamic effect, common to all devices, measuring the impact of the environment on the degradation. The model is identifiable under mild constraints. Besides producing gains regarding the interpretability of the parameters, this decomposition generates a parsimonious model, reducing computational time. The relation between degradation and failure time is obtained, allowing a computational approximation for the failure time distribution. The model performance is evaluated through simulation, helping to guide the prior specifications to model identification. The proposed model is applied to analyze fatigue crack growth data. We compare the proposed model with the traditional linear Weibull model and with a dynamic linear Normal model. Results show that the proposed methodology is competitive in predicting failure times and estimating the remaining useful life.

Kevin Li

ProSpar-GP: scalable Gaussian process modeling with massive non-stationary datasets

Abstract: Gaussian processes (GPs) are a popular class of Bayesian nonparametric models, but its training can be computationally burdensome for massive training datasets. While there has been notable work on scaling up these models for big data, existing methods typically rely on a stationary GP assumption for approximation, and can thus perform poorly when the underlying response surface is non-stationary, i.e., it has some regions of rapid change and other regions with little change. Such non-stationarity is, however, ubiquitous in real-world problems, including our motivating application for surrogate modeling of computer experiments. We thus propose a new Product of Sparse GP (ProSpar-GP) method for scalable GP modeling with massive non-stationary data. The ProSpar-GP makes use of a carefully-constructed product- of-experts formulation of sparse GP experts, where different experts are placed within local regions of non-stationarity. These GP experts are fit via a novel variational inference approach, which capitalizes on mini-batching and GPU acceleration for efficient optimization of inducing points and length-scale parameters for each expert. We further show that the ProSpar-GP is Kolmogorov-consistent, in that its generative distribution defines a valid stochastic process over the prediction space; such a property provides essential stability for variational inference, particularly in the presence of non-stationarity. We then demonstrate the improved performance of the ProSpar-GP over the state-of-the-art, in a suite of numerical experiments and an application for surrogate modeling of a satellite drag simulator.

Shengbin Ye

Iterative BART - Operator-Induced Structural Variable Selection for Identifying Materials Genes

Abstract: In the emerging field of materials informatics, a fundamental task is to identify physicochemically meaningful descriptors, or materials genes, which are engineered from primary features and a set of elementary algebraic operators through compositions. Standard practice directly analyzes the high-dimensional candidate predictor space in a linear model; statistical analyses are then substantially hampered by the daunting challenge posed by the astronomically large number of correlated predictors with limited sample size. We formulate this problem as variable selection with operator-induced structure (OIS) and propose a new method to achieve unconventional dimension reduction by utilizing the geometry embedded in OIS. Although the model remains linear, we iterate nonparametric variable selection for effective dimension reduction. This enables variable selection based on ab initio primary features, leading to a method that is orders of magnitude faster than existing methods, with improved accuracy. To select the nonparametric module, we discuss a desired performance criterion that is uniquely induced by variable selection with OIS; in particular, we propose to employ a Bayesian Additive Regression Trees (BART)-based variable selection method. Numerical studies show superiority of the proposed method, which continues to exhibit robust performance when the input dimension is out of reach of existing methods. Our analysis of single-atom catalysis identifies physical descriptors that explain the binding energy of metal-support pairs with high explanatory power, leading to interpretable insights to guide the prevention of a notorious problem called sintering and aid catalysis design.

Session H	Sponsored session: Bayes in Finance and Insurance Chair: Antonio Canale
Room	10A
Sponsors	Generali

Eurizon

Anna Illiano and Marco Momo

Adopting a Bayesian approach to address today's business challenges: use-cases in the Insurance Industry

Abstract: Generali is one of the major Italian players in Insurance and asset management. In our team, we work to manage data flows and build data-driven processes to assist and enhance our operations across several business lines.

In this talk, we will discuss two instances in which the application of the Bayesian approach helped improve how we addressed the issues at hand.

The first use case consists of the application of Bayesian model averaging to improve the predictive performance of our anti-fraud model. The second use case represents an application of Bayesian Demography techniques to a non-strictly demographic domain: to forecast the life expectancy of insurance policies.

Adrian Raftery

Discussant

Andrea Trovato

Bayesian time series modelling and tactical asset allocation

Abstract: In the first section of the paper I propose a framework for modelling investment strategies on volatility control. Precisely, an early warning method for tactical asset allocation is developed in order to deal with volatility clustering and fat-tail distributions of financial returns. The early warning signals are given by a two-state Markov switching model with high-volatility and low-volatility states and time-varying transition probabilities. Using Bayesian inference and Gibbs Sampling for posterior approximation, I apply the model to the S&P500 Index for daily data from December 2000 up to the present. The results suggest that a measure of the gap between implied volatility (VIX Index) and realised volatility has important leading properties on switches between volatility regimes. The paper proposes to calibrate strategies of tactical asset allocation conditionally on the predominant volatility regime. Innovative examples of portfolio allocation strategies are Bongaerts et al. (2020), Chan et al. (2000) and Rasmussen (2003). Notable references on Markov Switching Models are provided by Frühwirth- Schnatter (2006), Casella and Wells (2000) and Kim et al. (2008). Besides individual volatility of assets returns, estimating the covariance structure between assets in portfolios is of great importance to efficiently support decisions of tactical asset allocation. Increased availability of high-frequency data in the last decade emerges in development of new non-parametric approach for modelling volatility in a multivariate setup. While in the GARCH and stochastic volatility framework the volatility process is latent, the realized covariance methods employ high-frequency data for modelling the correlative structure of underlying assets, making it effectively observable. I contribute to literature by proposing a methodology for dynamic modelling and forecasting realized variance vectors based on a Bayesian Time-Varying Markov Switching VAR Model. The approach allows for flexible dependence patterns and investigates spillover effects across volatilities of different asset classes. The application to a global portfolio investing in bonds, equities and currencies provides that volatility spillovers follow a regime-based representation, with state transitions driven by dynamics of the 3- month US T-Bill rate. Notable contributions on Bayesian inference and Bayesian Markov Switching VAR models are provided by Robert (2007), Billio et al. (2016) and Agudze et al. (2022). On Bayesian forecasting and dynamic models pioneer references are provided by Zhou et al. (2014) and West and Harrison (1997).]

The second section of the paper provides a scenario-based framework for setting trading strategies on interest rates. Precisely, a data-enriched term- structure modelling based on non arbitrage arguments is proposed in a state-space representation. The model includes the latent instantaneous rate as a state process and it is augmented with further measurement equations on bonds at different maturities. In order to make the evaluation of the yield curve more reactive to market events, the model incorporates additional exogenous variables related to monetary policy, equity market volatility and macroeconomic fundamentals such as inflation and output gap. The paper presents an innovative and a more interpretable way to estimate the market premium, as a combination of explicit sources of risk. The model is applied to the US yield curve sampled at a monthly frequency from December 2000 up to the end of May 2024. The augmented framework and the Bayesian inference allow for generating projections of the yield curve in line with different macroeconomic scenarios. The proposed data-enriched approach allows for a more accurate estimate of the latent short-term rate. Such an improvement yields a notable gain in terms of goodness-of-fit and it leads to a more credible definition of future scenarios and profitable trading strategies on interest rates. Following the approach suggested by Jones (1998) and Polson et al. (2002), for instance, Sanford and Martin (2005), Frühwirth-Schnatter and Geyer (1998) and Hordal et al. (2006) proposed a Bayesian simulation-based method for estimating the term structure model with the

discrete observations on yields augmented by additional high-frequency latent data.	
Mike West	**Discussant**

ISBA World Meeting

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

Saturday July 6th, 2024

Schedule of the Day

8:30 am - 10:00 am	Multi-track session 9
10:00 am - 10:30 am	*Coffee break*
10:30 am - 12:00 pm	David Draper Memorial Session
10:30 am - 12:00 pm	Multi-track session 10
12:00 pm - 1:30 pm	*Lunch break*
1:30 pm - 2:30 pm	Keynote Lecture: Johannes Schmidt-Hieber
2:30 pm - 3:00 pm	*Coffee break*
3:00 pm - 4:30 pm	Multi-track session 11
4:30 pm - 5:30 pm	ISBA General Meeting
7:00 pm - 10:00 pm	*Banquet and Awards*

8.30 am Multi-track session 9

For invited sessions, the organizer will chair unless otherwise stated

Session A	Invited session: Junior Advances in Scalable Bayesian Computational Methods Organizers: Cecilia Balocchi and Alejandra Avalos-Pacheco
Room	9A
Yuexi Wang	**Horseshoe Priors for Sparse Dirichlet-Multinomial Models**

Abstract: Bayesian inference for Dirichlet-Multinomial (DM) models has a long and important history. The concentration parameter α is pivotal in smoothing category probabilities within the multinomial distribution and is crucial for the inference afterward. Due to the lack of a tractable form of its marginal likelihood, α is often chosen ad-hoc, or estimated using approximation algorithms. A constant α often leads to inadequate smoothing of probabilities, particularly for sparse compositional count datasets. In this paper, we introduce a novel class of prior distributions facilitating conjugate updating of the concentration parameter, allowing for full Bayesian inference for DM models. Our methodology is based on fast residue computation and admits closed-form posterior moments in specific scenarios. Additionally, our prior provides continuous shrinkage with its heavy tail and substantial mass around zero, ensuring adaptability to the sparsity or quasi-sparsity of the data. We demonstrate the usefulness of our approach on both simulated examples and on a real-world human microbiome dataset. Finally, we conclude with directions for future research.

Chuxuan (Jessie) Jiang	Bayesian Inference for Vertex-Series-Parallel Partial Orders

Abstract: Partial orders are a natural model for the social hierarchies that may constrain "queue-like" rank-order data. However, the computational cost of counting the linear extensions of a general partial order on a ground set with more than a few tens of elements is prohibitive. Vertex-series-parallel partial orders (VSPs) are a subclass of partial orders which admit rapid counting and represent the sorts of relations we expect to see in a social hierarchy. However, no Bayesian analysis of VSPs has been given to date. We construct a marginally consistent family of priors over

VSPs with a parameter controlling the prior distribution over VSP depth. The prior for VSPs is given in closed form. We extend an existing observation model for queue-like rank-order data to represent noise in our data and carry out Bayesian inference on "Royal Acta" data and Formula 1 race data. Model comparison shows our model is a better fit to the data than Plackett-Luce mixtures, Mallows mixtures, and "bucket order" models and competitive with more complex models fitting general partial orders.

Paolo Ceriani	**Dimension-free convergence of coordinate-ascent variational inference algorithms for large hierarchical models**

Abstract: Mean Field Variational Inference (MF-VI) is a popular computational technique for approximating posterior distribution in large-scale Bayesian models. The associated optimization task of finding the KL-optimal distribution within the family of fully factorized distributions is often performed via the Coordinate Ascent VI algorithm (CAVI). Despite its popularity, there are still few quantitative results on the convergence properties of CAVI. We investigate how the speed of convergence of CAVI is affected by the dimensionality of the problem for two-level hierarchical models (also called global-local models). In particular, we obtain dimension-free convergence results under random data generating assumptions in the context of hierarchical models in asymptotic settings where both the number of data and parameters grow, with minimal assumptions on the model. Such results are substantiated by extensive simulations on synthetic and real dataset.

Gareth Roberts	**Discussant**

↑ back to the schedule ↑

Session B	Invited session: Discrete Random Structures in Bayesian Nonparametrics Organizer: Igor Pruenster
Room	2A
Junyi Zhang	**Posterior sampling from truncated inverse-Lévy representation of normalised completely random measure mixtures**

Abstract: In this talk, we introduce the truncated inverse-Lévy representation of the completely random measures (CRMs). The representation is obtained by keeping the N largest atom weights of the CRM unchanged and combining the smaller atom weights into a single term. We develop the simulation algorithms for the representation and characterise its posterior distribution, for which a blocked Gibbs sampler is devised. We demonstrate the usage of the representation in two models. The first assumes such an approximation as the mixing distribution of a Bayesian nonparametric mixture model and leads to a finite approximation to the model posterior. The second concerns the finite approximation to the Caron-Fox model. Examples and numerical implementations are given based on the gamma, stable and generalised gamma processes.

Francesco Sanna Passino	**Bayesian nonparametric projected normal mixture models for spectral graph clustering with degree heterogeneity**

Abstract: Real-world networks commonly exhibit within-group degree heterogeneity. For example, in an enterprise computer network, users from the same department might have very different levels of activity depending on their job. In this work, improvements to existing methodologies for clustering graphs with within-group degree heterogeneity are proposed, under the degree-corrected stochastic blockmodel (DCSBM) framework. In previous work, it has been shown that, under the DCSBM, the performance of community detection algorithms based on spectral embedding could be improved by a transformation to spherical coordinates of a scaled spectral decomposition of the graph adjacency matrix, called spectral embedding. A weighted Bayesian nonparametric mixture of projected normals is proposed to perform clustering of nodes on the unit d-sphere resulting from the transformation. The methodology is demonstrated to outperform existing techniques and applied to real data from a university computer network.

Claudio Del Sole

A random partition model with dependence on covariates

Abstract:Bayesian nonparametric models often assume some kind of homogeneity among the observations, motivated by the exchangeability assumption in de Finetti’s representation theorem. In presence of multiple groups of observations, homogeneity within each group is usually modeled through partial exchangeability. Instead, including continuous covariates within a fully nonparametric regression model represents a more challenging task, and existing models in the literature face a trade-off between flexibility in modelling the latent partition structure, its analytical tractability, and its consistency for new observations. This work introduces a novel class of covariate-dependent random probability measures, arising from the normalization of suitable random measures, which depend on covariates through a kernel structure: specifically, the jumps of a common discrete random measure are rescaled via multiplication by a similarity kernel. A noteworthy example arises when the distribution of such random measure is a specific transformation of the distribution of a stable completely random measure. This construction induces a random partition model with dependence on covariates, which is characterized by great flexibility while retaining some analytical tractability, thanks to the introduction of suitable latent variables; moreover, it is inherently consistent for new observations. Both the partition probability function and the posterior distribution of the common random measure are derived in closed form, conditionally on such latent variables. A marginal Gibbs sampler, based on a generalized Pólya urn scheme, is also developed for posterior computation, together with a conditional slice sampling algorithm. Our proposal can be effectively exploited as a clustering or species sampling model which incorporates information available through both discrete and continuous covariates; in addition, it may represent the building block for the construction of nonparametric regression models. This is a joint work with Antonio Lijoi and Igor Prünster.

Session C

Invited session: Bayesian Empirical Likelihood and Related Techniques
Organizer: Mario Peruggia

↑ back to the schedule ↑

<table>
<tr><td>Room</td><td>6A</td></tr>
<tr><td>Nicole Lazar</td><td>Bayesian Empirical Likelihood for Ridge and Lasso Regression</td></tr>
<tr><td colspan="2">Abstract: Empirical likelihood (EL) is a nonparametric analog of standard likelihood, and inherits many of its properties. Accordingly, Bayesian EL, which replaces the ordinary likelihood function with an empirical likelihood function, has been developed. In this talk, I will briefly introduce empirical likelihood in both its standard and Bayesian versions. A recent development is Bayesian EL for regularized methods such as ridge regression. The theoretical background for Bayesian ridge and Bayesian lasso empirical likelihoods will be presented, along with discussion of computational issues. This is joint work with Adel Bedoui.</td></tr>
<tr><td>Steven N. MacEachern</td><td>Regularized Empirical Likelihood for Bayesian Inference: Theory and Applications</td></tr>
<tr><td colspan="2">Abstract: Bayesian inference with empirical likelihood faces a challenge. The domain of the posterior distribution is a proper subset of the original parameter space due to the convex hull constraint that is implicit in empirical likelihood. Oddly, the proper subset grows as more data are collected. This stands in sharp contrast to traditional Bayesian methods where one expects the domain of the posterior to remain unchanged or to shrink. To resolve this problem, we propose a novel regularization of exponentially tilted empirical likelihood. The regularization involves a continuous exponential family distribution. This allows the domain of the ``posterior” distribution to be the same as the original parameter space for samples of all sizes. We show that this regularized exponentially tilted empirical likelihood retains certain desirable asymptotic properties of (exponentially tilted) empirical likelihood with improved finite sample performance. Simulations and data analysis demonstrate that the proposed method provides a suitable pseudo-likelihood for Bayesian inference. This is joint work with Eunseop Kim and Mario Peruggia</td></tr>
<tr><td>Ryan Thompson</td><td>Familial Inference for Regression Models</td></tr>
<tr><td colspan="2">Abstract: Familial inference is a principled framework for testing statistically ambiguous scientific hypotheses. To date, familial inference has only been developed</td></tr>
</table>

for hypotheses relating to unconditional measures of center. This work generalizes familial inference to the regression setting. The procedure evaluates a continuum of regression models, beginning at the conditional mean and ending at the conditional median. To test hypotheses on this continuum, the method sets forth a new Bayesian nonparametric test. This is joint work with Catherine Forbes, Steven MacEachern and Mario Peruggia

Session D	Invited session: Recent Development of Bayesian methods for complex spatial and imaging data analysis Organizer: Jian Kang
Room	Magna
Huiyan Sang	**GS-BART: Graph split additive decision trees for classification and nonparametric regression of spatial and network data**

Abstract: Ensemble decision tree methods such as XGBoost, RF, and BART have gained enormous popularity in data science for their superior performance in machine learning regression and classification tasks. In this paper, we develop a new Bayesian graph-split-based additive decision trees method, called GS-BART, to improve the performance of Bayesian additive decision trees for complex dependent data with graph relations. The new method adopts a highly flexible split rule complying with graph structure to relax the axis-parallel split rule assumption in most existing ensemble decision tree models. We design a scalable informed MCMC algorithm leveraging a gradient-based recursive algorithm to sample the graph-split-based decision tree. The superior performance of the method over conventional ensemble tree models and Gaussian process regression models is illustrated in various regression and classification tasks for spatial and network data analysis.

David Dunson	**Identifiable and interpretable nonparametric factor analysis**

Abstract: Factor models have been widely used to summarize the variability of high-dimensional data through a set of factors with much lower dimensionality. Gaussian linear factor models have been particularly popular due to their interpretability and ease of computation. However, in practice, data often violate the multivariate Gaussian assumption. To characterize higher-order dependence and nonlinearity, models that include factors as predictors in flexible multivariate regression are popular, with GP-LVMs using Gaussian process (GP) priors for the regression function and VAEs using deep neural networks. Unfortunately, such approaches lack identifiability and interpretability and tend to produce brittle and non-reproducible results. To address these problems by simplifying the nonparametric factor model while maintaining flexibility, we propose the NIFTY framework, which parsimoniously transforms uniform latent variables using one-dimensional nonlinear mappings and then applies a linear generative model. The induced multivariate distribution falls into a flexible class while maintaining simple computation and interpretation. We prove that this model is identifiable and empirically study NIFTY using simulated data, observing good performance in density estimation and data visualization. We then apply NIFTY to bird song data in an environmental monitoring application.

Weining Shen	**Bayesian biclustering and its application in education data analysis**

Abstract: We propose a novel nonparametric Bayesian item response theory model to estimate clusters at the question level while simultaneously allowing for heterogeneity at the examinee level under each question cluster, characterized by the mixture of Binomial distributions. We present some theoretical results that guarantee the identifiability of the proposed model and show that the model can correctly identify question-level clusters asymptotically. We also provide a tractable sampling algorithm to obtain valid posterior samples from the proposed model. Compared to the existing methods, the model manages to reveal the multi-dimensionality of the examinee's proficiency level in handling different types of questions parsimoniously by imposing a nested clustering structure. The proposed model is evaluated via a series of simulations as well as applied to an English proficiency assessment data set. This data analysis example nicely illustrates how the model can be used by test makers to distinguish different types of students and aid in the design of future tests.

Michele Guindani	**Discussant**

Session E	Invited session: Bayesian Methods: Pushing the Boundaries of Data Science Organizer: Piyali Basak Chair: Garritt Page
Room	5A
Debajyoti Sinha	**Soft Bayesian Additive Regression Trees (SBART) for analysis of complex studies with clustered survival, multiple responses, spatial effects and survey weights**

Abstract: Using a Poisson thinning based latent variable based data augmentation step, we have recently developed a computationally feasible and methodologically flexible method for analyzing various types of survival studies under a paradigm of Bayesian ensemble learning called Soft Bayesian Additive Regression Trees (SBART in short). Our computational tool can accommodate unknown functional forms of the main effects, unknown interactions of various covariates, unobserved cluster-specific effects, large number of clusters, and variable cluster sizes. We demonstrate our method first via a nonparametric Bayesian analysis of clustered survival study when some of the cluster-level covariates are scientifically augmented using a data source different from the survival study. Popular parametric and semi-parametric hazard regression methods for clustered survival data are inappropriate and inadequate for such studies where random cluster effects as well as the unknown cluster-level covariates are spatially correlated. We illustrate the practical implementation of our method for the assessment of the impacts of the interventions in various county-level covariates to eliminate the racial disparity in survival times of breast cancer patients in different counties (clusters) of Florida. We then further demonstrate the advantages of our SBART tools for dealing with large complex surveys with mixed-type responses where at least some of the responses are either count or highly skewed.

Candace Berrett	**Partially Ordering Prior Distributions**

Abstract: One major advantage of Bayesian methods for data analyses is the ability to elicit prior information from a subject matter expert. However, identifying how to elicit prior information from the expert can be a challenge. Consider the case of three treatment groups: placebo, low dose, and high dose. Rather than eliciting prior distributions on the means themselves, it may be easier for the expert to consider the relative treatment orderings: placebo < low dose < high dose. In this talk we explore how a statistician could illustrate the different choices of prior distributions on the treatment differences that an expert can use to choose the prior distribution that best reflects the current knowledge in the field. This is particularly important in cases where collecting large sample sizes is challenging (e.g., cancer treatment trials) and inference on treatments with little to no data -- relying solely on data for other treatment levels and expert knowledge -- is imperative.

Vanda Inacio De Carvalho	**Density regression via Dirichlet process mixtures of normal structured additive regression models**

Abstract: There is an increased interest in studying how the distribution of a univariate, real-valued, continuous response changes with a set of covariates. Within a Bayesian nonparametric framework, dependent Dirichlet process mixture of normal distributions provide a highly flexible approach for estimating the conditional density function. However, several formulations of this class of models involve intricate algorithms for posterior inference, thus preventing their widespread use unless user-friendly publicly available software is provided. Motivated by this problematic, we propose a flexible, versatile, and computationally tractable model for density regression based on a single-weights dependent Dirichlet process mixture of normals model where an additive structure is assumed for the mean of each component and the effects of continuous covariates are modelled trough smooth functions. The major modelling components are penalised B-splines and their bivariate tensor product extension. The resulting model can be regarded as a Dirichlet process mixture of normal structured additive regression models and can easily deal with discrete covariates, nonlinear effects of continuous covariates, interaction surfaces, spatial effects, and varying coefficient terms. A practically important feature of our method is that, since the full conditional distributions for all model parameters are available in closed form, it allows for ready posterior simulation through Gibbs sampling. The results of the simulation study show that our approach successfully recovers the true conditional densities in a variety of challenging scenarios. Applications to a toxicology, disease diagnosis, and agricultural study are provided.

Felipe Barrientos	**Bayesian nonparametric density estimation on the product of simplex spaces and the hypercube using multivariate Bernstein polynomials**

Abstract: We propose a Bayesian nonparametric model for density estimation on the product of simplex spaces and the hypercube. The model is particularly useful for cases where the available data consist of multiple compositional features alongside variables that take on values within bounded intervals. A compositional feature is a vector of non-negative components whose sum of values remains constant, such as the time an individual spends on different activities during the day or the fraction of different types of food consumed as part of a person's diet. Our approach relies on a generalization of random multivariate Bernstein polynomials and corresponds to a Dirichlet process mixture of products of Dirichlet and beta densities. Theoretical properties such as prior support and posterior consistency are studied. We evaluate the model's performance through a simulation study and a real-world application using data from the 2005–2006 cycle of the U.S. National Health and Nutrition Examination Survey (NHANES). Furthermore, the conditional densities derived under this modeling strategy can be used for regression analyses where both the response and predictors take values on the simplex space and/or hypercube. Joint work with Rufeng Liu, Claudia Wehrhahn, and Alejandro Jara.

Session F	Invited session: Markov Chain Monte Carlo: Methods and Applications Organizer: Radu Craiu
Room	4A
Nianqiao Ju	**Spectral gap bounds for reversible hybrid Gibbs chains**

Abstract: Hybrid Gibbs samplers represent a prominent class of approximated Gibbs algorithms that utilize Markov chains to approximate conditional distributions, with the Metropolis-within-Gibbs algorithm standing out as a well-known example. Despite their widespread use in both statistical and non-statistical applications, very little is known about their convergence properties. This article introduces novel methods for establishing bounds on the convergence rates of hybrid Gibbs samplers. In particular, we examine the convergence characteristics of hybrid random-scan Gibbs and data augmentation algorithms. Our analysis reveals that the absolute spectral gap of a reversible hybrid chain can be bounded based on the absolute spectral gap of the exact Gibbs chain and the absolute spectral gaps of the Markov chains employed for conditional distribution approximations. The new techniques are applied to four algorithms: a random-scan Metropolis-within-Gibbs sampler, a hybrid proximal sampler, random-scan Gibbs samplers with block updates, and a hybrid slice sampler. Joint work with Qian Qin and Guanyang Wang.

Alexandre Bouchard	**Fast, Distributed Bayesian Inference for Everyone**

Abstract: How can we scale Bayesian inference using distributed architectures? I will describe novel perspectives on this old problem coming from non-equilibrium, regenerative, and non-reversible perspectives. My group is working on making these complex Monte Carlo methods easy to use: check out https://pigeons.run/dev/ a package allowing the user to leverage clusters of 1000s of nodes to speed-up difficult Monte Carlo problems without requiring knowledge of distributed algorithms.

Galin Jones	**Metropolis-Hastings Markov chains in the High-Dimensional, Large Sample Size Regime**

Abstract: To avoid poor empirical performance in Metropolis-Hastings practitioners often tune them by trial and error, but this can be delicate in the setting with a high-dimensional parameter space and a large sample size. Lower bounds on the convergence rate are developed in both total variation and Wasserstein distances to identify how the simulations will fail so these settings can be avoided. Particular attention is paid to using the lower bounds to study the convergence complexity of

Metropolis-Hastings Markov chains and to constrain the rate of convergence for geometrically ergodic Markov chains. If the target density concentrates with a parameter n (e.g. posterior concentration, Laplace approximations), it is demonstrated that the convergence rate of a Metropolis-Hastings chain can be arbitrarily slow if the tuning parameters do not depend carefully on n. This is demonstrated with Bayesian logistic regression with Zellner's g-prior when the dimension and sample increase together and flat prior Bayesian logistic regression as n tends to infinity. Extensions to component-wise samplers are presented.

Session G	Contributed session: Gaussian Processes Chair: Alexandra Schmidt
Room	7A
Xinyi Xu	**Double-Score Gaussian Process Model for Robust Causal Inference in Observational Studies**

Abstract: The propensity score is a widely recognized tool for addressing confounding in causal inference within observational data. Recently, matching methodologies have evolved to incorporate both prognostic and propensity scores, offering consistent estimations when either model is correctly specified. However, these methods have centered on estimating Population Average Treatment Effects (PATE) but overlook the heterogeneity of treatment effects across subgroups, or Conditional Average Treatment Effects (CATE), which are crucial in many practical scenarios. Also, assessing estimation variability for matching estimators can be challenging and computationally demanding. In this work, we propose a Bayesian semi-parametric causal model that integrates both propensity and prognostic scores through Gaussian processes. This model exhibits a double robustness attribute, ensuring consistency if either score model is accurate. Furthermore, we derive the asymptotic distributions for both PATE and CATE under our model and a propensity-score-only model. This comparison reveals that our double score model offers reduced asymptotic variances in estimating both PATE and CATE. The implementation of a Markov Chain Monte Carlo (MCMC) algorithm within this framework facilitates sampling from posterior distributions, which allows straightforward calculations of the variances for both PATE and CATE estimators. Through extensive simulation studies, we demonstrate the superior performance of our double-score model over the propensity-score-only

↑ back to the schedule ↑

model, particularly when faced with model misspecification. This advancement represents a significant stride in the field of causal inference, offering more robust and nuanced insights in observational studies.

Francois Caron	**Large-width deep neural networks with dependent weights**

Abstract: In this talk I will discuss the infinite-width limit of deep feedforward neural networks whose weights are dependent, and modelled via a mixture of Gaussian distributions. Each hidden node of the network is assigned a nonnegative random variable that controls the variance of the outgoing weights of that node. We make minimal assumptions on these per-node random variables: they are iid and their sum, in each layer, converges to some finite random variable in the infinite-width limit. Under this model, we show that each layer of the infinite-width neural network can be characterised by two simple quantities: a non-negative scalar parameter and a Levy measure on the positive reals. If the scalar parameters are strictly positive and the Levy measures are trivial at all hidden layers, then one recovers the classical Gaussian process (GP) limit, obtained with iid Gaussian weights. More interestingly, if the Levy measure of at least one layer is non-trivial, we obtain a mixture of Gaussian processes (MoGP) in the large-width limit. The behaviour of the neural network in this regime is very different from the GP regime. One obtains correlated outputs, with non-Gaussian distributions, possibly with heavy tails. Additionally, we show that, in this regime, the weights are compressible, and some nodes have asymptotically non-negligible contributions, therefore representing important hidden features. Many sparsity-promoting neural network models can be recast as special cases of our approach, and we discuss their infinite-width limits; we also present an asymptotic analysis of the pruning error. We illustrate some of the benefits of the MoGP regime over the GP regime in terms of representation learning and compressibility on simulated, MNIST, and Fashion MNIST datasets.

Renato Berlinghieri	**Gaussian processes at the Helm(holtz): A more fluid model for ocean currents**

Abstract: Given sparse observations of buoy velocities, oceanographers are interested in reconstructing ocean currents away from the buoys and identifying divergences in a current vector field. As a first and modular step, we focus on the time-stationary case - for instance, by restricting to short time periods. Since we expect current velocity to be a continuous but highly non-linear function of spatial location, Gaussian processes (GPs) offer an attractive model. But we show that applying a GP with a standard stationary kernel directly to buoy data can struggle at both current reconstruction and divergence identification, due to some physically unrealistic prior assumptions. To better reflect known physical properties of currents, we propose to instead put a standard stationary kernel on the divergence and curl-free components of a vector field obtained through a Helmholtz decomposition. We show that, because this decomposition relates to the original vector field just via mixed partial derivatives, we can still perform inference given the original data with only a small constant multiple of additional computational expense. We illustrate the benefits of our method with theory and experiments on synthetic and real ocean data.

Francois-Xavier Briol	**Robust and Conjugate Gaussian Process Regression**

Abstract: To enable closed form conditioning, a common assumption in Gaussian process (GP) regression is independent and identically distributed Gaussian observation noise. This strong and simplistic assumption is often violated in practice, which leads to unreliable inferences and uncertainty quantification. Unfortunately, existing methods for robustifying GPs break closed-form conditioning, which makes them less attractive to practitioners and significantly more computationally expensive. In this talk, we demonstrate how to perform provably robust and conjugate Gaussian process (RCGP) regression at virtually no additional cost using generalised Bayesian inference. RCGP is particularly versatile as it enables exact conjugate closed form updates in all settings where standard GPs admit them. To demonstrate its strong empirical performance, we deploy RCGP for problems ranging from Bayesian optimisation to sparse variational Gaussian processes.

10.30 am David Draper Memorial Session

Room	2A
Speakers	**Dimitris Fouskakis, Ioannis Ntzoufras, William Brown, Erdong Guo, Edward George, Alexander Terenin**

This session will include an open mic component for attendees to share memories

10.30 am Multi-track session 10

For invited sessions, the organizer will chair unless otherwise stated

<table>
<tr><td>Session A</td><td>Invited session: Bayesian Privacy
Organizer: Christian P Robert
Chair: Joshua Bon</td></tr>
<tr><td>Room</td><td>6A</td></tr>
<tr><td>James Bailie</td><td>How does differential privacy limit disclosure risk? A precise prior-to-posterior analysis</td></tr>
<tr><td colspan="2">Abstract: Differential privacy (DP) is an increasingly popular standard for quantifying privacy in the context of sharing statistical data. It has numerous advantages – especially its composition of privacy loss over multiple data releases and its facilitation of valid statistical inference via algorithmic transparency – over previous statistical privacy frameworks. Yet one difficulty of DP in practice is setting its privacy loss budget. Such a choice is complicated by a lack of understanding of what DP means in connection to traditional notions of statistical disclosure limitation (SDL). In this talk, we trace the rich literature on SDL back to the foundational 1986 paper by Duncan and Lambert, which defines disclosure in a relative sense as an increase – due to the published data – in one’s knowledge of an individual record. We prove that DP is exactly equivalent to limiting this type of ‘prior-to-posterior’ disclosure, but only when the records are completely independent. More generally, DP is equivalent to controlling conditional prior-to-posterior learning, when conditioning on all other records in the dataset. This connects DP to traditional SDL while also highlighting the danger of viewing data variations as mechanistic – as does DP – rather than as statistical – in which one would explicitly acknowledge the variational dependencies between records.</td></tr>
<tr><td>Shenggang Hu</td><td>Privacy Guarantees in Posterior Sampling under Contamination</td></tr>
<tr><td colspan="2">Abstract: In recent years, differential privacy has been adopted by tech-companies</td></tr>
</table>

and governmental agencies as the standard for measuring privacy in algorithms. We study the level of differential privacy in Bayesian posterior sampling setups. As opposed to the common privatization approach of injecting Laplace/Gaussian noise into the output, Huber's contamination model is considered, where we replace at random the data points with samples from a heavy-tailed distribution. In bounding the differential privacy level, our approach removes the need for assuming bounded observation and parameter spaces. We further consider the effect of sample size on privacy level and the convergence rate of epsilon and delta to zero. Asymptotically, the contamination approach is fully private at no cost of information loss.

Joshua Bon	**A Bayesian decision-theoretic framework for measuring the privacy of data**

Abstract: The scientific and economic value of data continues to increase as our technology does. New hardware and software technologies allow, but also often require, richer and vaster data to operate. As the value of input data to these systems is recognised, so too is the loss of privacy for data providers. In this landscape, data privacy is a fundamental issue for data science disciplines such as statistics, and machine learning. More generally, it is an issue for scientific and industrial pursuits that use sensitive data. There are many existing definitions of privacy, the most popular being Differential Privacy (Dwork, 2006), which may be too strict in some contexts. In this work we develop a framework for measuring privacy from a Bayesian decision-theoretic point of view. With our framework we can generate new privacy principles that are rigorously justified, assess existing privacy definitions using decision-theory, and create new definitions that are fit for purpose. For example, we can evaluate the privacy of summary statistics without added noise, which is not possible with Differential Privacy. We also recover Differential Privacy as a special case.

Stanislas du Ché	**Discussant**

Session B	Invited session: Recent Advances in Incorporating Sample Design Effects and Spatial Structure into Small Area Estimation Organizer: Yajuan Si

Room	10A
Malay Ghosh	**Hierarchical Bayesian Analysis of Spatial Small Area Estimation Models**

Abstract: It is now well-recognized that small area estimation methods need to be model based. Up until now, most such analysis is based on a classical normal mixed effects model. However, very often, the small or local areas exhibit a spatial structure. The talk will aim at reviewing some of the recently developed hierarchical Bayesian methods for handling spatial small area estimation models.

Michael Elliott	**Advances in Incorporating Sample Designs in Bayesian Inference**

Abstract: Blending the Bayesian paradigm, with its emphasis on complex modeling, with the survey sampling paradigm, with its emphasis on non-parametrics and robustness, has been difficult. Here we adapt a weighted finite population Bayesian bootstrap procedure for generating synthetic populations (Dong et al. 2014; Zhou et al. 2016) to incorporate design effects into Bayesian analyses via importance weighting. We generate a synthetic population and draw single simple random sample (SRS) of the same size to which we fit our Bayesian model using Markov Chain or Hamiltonian Monte Carlo. We then generate L synthetic populations from which we draw SRSs using compute posterior mean estimates using importance weighting; these estimators are then combined using synthetic population theory (Raghunathan et al. 2003) to compute design-adjusted posterior means, variances, and credible intervals. We consider this approach in a few simulation settings, and discuss applications to accounting for complex sample design in the setting of small area estimation, using multiple datasets to develop county-level estimates of risky behavior.

Zhengyuan Zhu	**Modeling of Spatially Indexed Distributions and Application to Local Soil Erosion Distribution Prediction**

Abstract: Estimating local erosion distributions within small watersheds is crucial for making informed policy decisions and allocating resources to address severe erosion problems. Currently, obtaining this information is expensive and limited to only a few watersheds. To address this issue, this work proposes a novel nonparametric spatial prediction method which models a local erosion distribution in its entirety as a probability density in the Wasserstein space. Local erosion distributions are transformed into square-integrable trajectories and represented using basis expansion, which becomes a multivariate random field that encodes the spatial structure. Further application of local regression and Kriging enables flexible modeling and prediction of the spatially indexed density functions. Our method also provides improved prediction for functionals of the density/distribution, such as the mean and exceedance probability. Simulation studies suggest that the proposal substantially outperforms a parametric method if the latter is misspecified and has comparable performance even if the latter is well-specified. Our work features a detailed local erosion analysis over Shaanxi province in China based on field survey data, where local erosion distributions are obtained by conditioning on local land use type and covariates derived from a digital elevation model. The proposed method produces more accurate predictions than established approaches.

Jon Wakefield	**Discussant**

Session C	Invited session: Temporal Modeling in Bayesian Inference Organizer: Matteo Ruggiero
Room	7A
Beatrice Franzolini	**Conditional partial exchangeability: a probabilistic framework for longitudinal and multi-view clustering**

Abstract: Clustering longitudinal data requires adjusting the number of clusters, their frequencies, and shapes over time to accurately capture heterogeneity and temporal dynamics. However, many existing dynamic clustering techniques overlook within-subject dependence, thereby neglecting the individual identities of subjects over time. This issue is also encountered in stochastic block models for longitudinal and multiplex network data. To overcome these limitations, we propose a broad class of Bayesian mixture models capable of generating dependent random partitions, where the dependency is introduced at the subject level. At the core of our proposal is the concept of conditional partial exchangeability, a novel probabilistic paradigm that ensures analytical and computational tractability while defining a flexible law governing dependent random partitions of the same objects across time, space, or domains.

Filippo Ascolani

Conjugacy of Wright-Fisher diffusions given genealogical data

Abstract: We consider a statistical model where Wright-Fisher diffusions are interpreted as random parameters drawn from the prior on their path space, and data are given by the genealogy associated to a random sample from the underlying population. We characterize conditional Wright-Fisher diffusions given these type of data and show they are inhomogeneous Wright-Fisher diffusions with drift given by piecewise constant functions which depend on the data. We thus establish conjugacy for this Bayesian categorical model on the path space, which marginally in time reduces to the well-known conjugacy of the Dirichlet-categorical model.

Athanasios Kottas

Mixture modeling for temporal point processes with memory

Abstract: We will present a constructive approach to building temporal point processes that incorporate dependence on their history. The dependence is modeled through the conditional density of the duration, i.e., the interval between successive event times, using a mixture of first-order conditional densities for each one of a specific number of lagged durations. Such a formulation for the conditional duration density accommodates high-order dynamics, and it thus enables flexible modeling for point processes with memory. The implied conditional intensity function admits a representation as a local mixture of first-order hazard functions. By specifying appropriate families of distributions for the first-order conditional densities, with different shapes for the associated hazard functions, we can obtain either self-exciting or self-regulating point processes. From the perspective of duration processes, we develop a method to specify a stationary marginal density. The

↑ back to the schedule ↑

resulting model, interpreted as a dependent renewal process, introduces high-order Markov dependence among identically distributed durations. Furthermore, we provide extensions to cluster point processes, which can describe duration clustering behaviors attributed to different factors. The point process models are implemented within the Bayesian framework for inference, model assessment, and prediction. The methods will be illustrated with data examples from environmetrics and finance.
Joint work with Xiaotian Zheng (School of Mathematics and Applied Statistics, University of Wollongong) and Bruno Sanso (Department of Statistics, University of California, Santa Cruz).

Session D	Invited session: Are Likelihoods and Priors Necessary to Address Bayesian Inference? Organizers: Federico Camerlenghi & Michele Guindani
Room	Magna
Chris Holmes	**Bayesian inference without the need for priors**

Abstract: Bayesian inference can be viewed as a population missing data problem. Starting with the data that you have, Bayesian uncertainty in quantities of interest arise from the data that you don't have (from the remaining population that wasn't measured). Posterior distributions can be obtained by imputing the missing information via a joint predictive model built using the data you have. The starting point of the analysis includes the available data. This violates the usual Bayesian position of specifying beliefs a priori, which ensures coherency but at considerable cost. We will motivate why inference without a likelihood-prior specification is both interesting and Bayesian, and discuss recent extensions of ideas from "Martingale posterior distributions" (2022) by Fong, Holmes, and Walker.

Takuo Matsubara	**Generalised Bayesian Inference for Discrete Intractable Likelihood**

Abstract: Discrete state spaces represent a major computational challenge to statistical inference, since the computation of normalisation constants requires summation over large or possibly infinite sets, which can be impractical. This paper addresses this computational challenge through the development of a novel generalised Bayesian inference procedure suitable for discrete intractable likelihood.

Inspired by recent methodological advances for continuous data, the main idea is to update beliefs about model parameters using a discrete Fisher divergence, in lieu of the problematic intractable likelihood. The result is a generalised posterior that can be sampled from using standard computational tools, such as Markov chain Monte Carlo, circumventing the intractable normalising constant. The statistical properties of the generalised posterior are analysed, with sufficient conditions for posterior consistency and asymptotic normality established. In addition, a novel and general approach to calibration of generalised posteriors is proposed. Applications are presented on lattice models for discrete spatial data and on multivariate models for count data, where in each case the methodology facilitates generalised Bayesian inference at low computational cost.

Sonia Petrone	**Reversing the order: from prediction to inference**

Abstract: We are used to going from inference to prediction. We elicit a model and a prior on the model's parameters; or, non-parametrically, we elicit a prior on the unknown distribution of the data; and from the inferential setting we go to prediction by computing the predictive distribution. In a predictive approach, we avoid models and priors and directly reason on the predictive learning rule - how we learn on future observations given the past. As we discuss in the talk, one can then go from prediction to inference. What are the advantages of this reversed order of reasoning? I will point out intriguing implications, computational advantages, and challenges. Conceptually, rather than as a quantification of the uncertainty on a true distribution, the posterior distribution is understood in terms of "efficiency" of the predictive learning rule - as we formalize through novel predictive-based asymptotic Gaussian approximations. Computationally, we can obtain a simple predictive-based Monte Carlo scheme to simulate from the posterior distribution. A challenge is that, while the pressure for fast computations would call for manageable learning rules, natural and computationally attractive predictive rules - for example algorithms used in Bayesian statistics and machine learning - may break exchangeability constraints. In fact, we will see through known and novel examples that they may preserve exchangeability desiderata asymptotically, thus providing (computationally attractive) approximations of exchangeable Bayesian procedures "without the prior". The talk is based on joint work with Sandra Fortini.

Session E	Contributed session: Bayesian Computational Statistics Chair: Sinead Williamson
Room	4A
Kamélia Daudel	**Alpha-divergence Variational Inference Meets Importance Weighted Auto-Encoders: Methodology and Asymptotics**

Abstract: Variational Inference methods are optimization-based methods that have generated a lot of attention in Bayesian Statistics due to their applicability to high-dimensional machine learning problems. In particular, several algorithms involving the Variational Renyi (VR) bound have been proposed to optimize an alpha-divergence between a target posterior distribution and a variational distribution. Despite promising empirical results, those algorithms resort to biased stochastic gradient descent procedures and thus lack theoretical guarantees. In this paper, we formalize and study the VR-IWAE bound, a generalization of the Importance Weighted Auto-Encoder (IWAE) bound. We show that the VR-IWAE bound enjoys several desirable properties and notably leads to the same stochastic gradient descent procedure as the VR bound in the reparameterized case, but this time by relying on unbiased gradient estimators. We then provide two complementary theoretical analyses of the VR-IWAE bound and thus of the standard IWAE bound. Those analyses shed light on the benefits or lack thereof of these bounds. Lastly, we illustrate our theoretical claims over toy and real-data examples.

Brandon Carter	**A Mixture of Directed Acyclic Graphs Approximation for Discrete Hidden Markov Random Fields**

Abstract: Markov Random Fields (MRF) are an intuitive method to characterize spatial dependence for discrete data. Despite the simplicity of the model formulation, MRFs are computationally expensive to implement due to an intractable normalizing constant. Most often practitioners, especially in the Bayesian setting, use the psuedo-likelihood as a proxy; however, there is little exploration of the implications on posterior inference when the psuedo-likelihood is used. We propose a new class of prior distributions, mixtures of directed acyclic graphs (MDAG), to represent spatial dependence for discrete data. This new prior offers an alternate interpretation of spatial dependence for discrete data. We examine three cases of the MDAG prior and develop MCMC algorithms for each specific case. In the setting of discrete outcome data with a discrete hidden MRF, we compare the MDAG prior to exact inference

↑ back to the schedule ↑

using a MRF and inference using the psuedo-likelihood. We show superior performance of the MDAG prior compared to the pseudo-likelihood in estimating the spatial dependence of the latent variable.

Julie Zhang

Multi-resolution coalescent inference of effective population size

Abstract: Variation in a sample of molecular sequence data informs about the past evolutionary history of the sample. Traditionally, Bayesian modeling coupled with the standard coalescent, is used to infer the bifurcating phylogeny and evolutionary parameters such as effective population size, and mutation rates. However, it is often the case that there is not enough signal to infer the high-resolution bifurcating phylogeny. In this case, the posterior is concentrated in a large set of phylogenies which can be summarized as a mutifurcating phylogeny that reflects the uncertainty in the data. Here, we propose a Bayesian nonparametric method for inferring effective population size trajectories from a multifurcating phylogeny. We introduce the multi-resolution coalescent model of multifurcating phylogenies and compare it with the Lambda-coalescent model. In addition, we estimate the parameter of the Beta-coalescent, which is a particular type of Lambda-coalescent. Finally, we test our methods on simulations and apply them to study the growth of SARS-CoV-2 variants. This is joint work with Julia A. Palacios.

Florence Bockting

Simulation-Based Prior Knowledge Elicitation for Parametric Bayesian Models

Abstract: A central characteristic of Bayesian statistics is the ability to consistently incorporate prior knowledge into various modeling processes. In our work, we focus on translating domain expert knowledge into corresponding prior distributions over model parameters, a process known as prior elicitation. Expert knowledge can manifest itself in diverse formats, including information about raw data, summary statistics, or model parameters. A major challenge for existing elicitation methods is how to effectively utilize all of these different formats in order to formulate prior distributions that align with the expert,Äôs expectations, regardless of the model structure. To address these challenges, we develop a simulation-based elicitation method that can learn the hyperparameters of potentially any parametric prior distribution from a wide spectrum of expert knowledge using stochastic gradient descent. We validate the effectiveness and robustness of our elicitation method in representative case studies ranging from (generalized) linear models to hierarchical models. Our method is largely independent of the underlying model structure and adaptable to various elicitation techniques, including quantile-based, moment-based, and histogram-based methods.

Session F	Contributed session: Asymptotics in Bayesian Statistics Chair: Sergios Agapiou
Room	9A
Botond Szabo	**Theoretical guarantees for distributed Gaussian Processes**

Abstract: Gaussian Processes are highly popular in various applications. However, they scale poorly with the sample size, i.e. even in case of the regression model the training of the GP is O(n^3) and it requires memory of O(n^2), limiting their applicability for large data sets. Therefore in practice often approximation methods are used. One of the most popular approach is the distributed Bayesian method where the data is divided over local servers/cores/machines and the computations are carried out in parallel to each other before they are aggregated. We study such distributed framework and consider both randomly and spatially divided data sets. We show that by optimally scaling these approaches (depending on the regularity of the underlying truth) one can acieve optimal rates for both methods. However, if the regularity is not known, then standard methods using randomly divided data perform sub-optimally, while spatial methods achieve rate optimal inference. We also develop new aggregation techniques and demonstrate on synthetic and real world data sets their applicability. Based on joint works with Harry van Zanten, Aad van der Vaart and Amine Hadji.

Conor Moriarty-Osborne	**Convergence rates of non-stationary and deep Gaussian process regression**

Abstract: Gaussian processes have proved to be a powerful and flexible tool in the reconstruction of functions given a set of known training points, with applications in machine learning, optimisation and data assimilation. However, they can be limited when the functions being reconstructed are of a non-stationary or anisotropic nature. Deep Gaussian processes, constructed using a hierarchical process where the inputs to a Gaussian process are themselves Gaussian processes, aim to give a more flexible approach to function reconstruction. We look at convergence rates of these deep Gaussian processes in terms of the number of known training points. We also show that deep Gaussian process regression achieves considerably better results than standard Gaussian process regression when reconstructing non-stationary and anisotropic functions.

Paul Egels

Heavy-tailed priors for deep neural networks: nonparametric adaptation to smoothness and structure

Abstract: Deep neural networks are a flexible method whose expressivity is particularly helpful in complex settings, where it has the ability to recover unknown smaller-dimensional structural dependencies. However, the tuning of parameters in the architecture of the network is nontrivial. In particular, if one adopts a nonparametric Bayesian approach, some form of model-selection is often used, for instance via spike-and-slab priors; while effective theoretically, this can face computational difficulties. Building upon a recent idea of Agapiou and Castillo (2023), who showed that well-chosen heavy-tailed series priors lead to near-optimal posteriors that adapt to smoothness, we introduce a new prior distribution on neural network parameters in a nonparametric regression setting. This prior fits all coefficients of a well-chosen network structure, drawing each weight from the same heavy-tailed distribution. We prove that the corresponding (tempered) posterior distribution is rate-optimal in the minimax sense (up to log factors) and adapts both to unknown smoothness and hidden compositional structure of the regression function. Unlike many procedures, notably spike and slab priors, the method we propose does not involve any explicit model selection. Additionally, we establish that the variational approximation of tempered posteriors, defined within a heavy-tailed mean-field space of approximation, also achieves a minimax concentration rate. This is joint work with Ismael Castillo (Sorbonne Universite).

Alice L’Huillier

A variational Bayes approach to inference for low-dimensional parameters in high-dimensional linear regression

Abstract: We propose a scalable variational Bayes method for statistical inference for a single or low-dimensional subset of the coordinates of a high-dimensional parameter in sparse linear regression. Our approach relies on assigning a mean-field approximation to the nuisance coordinates and carefully modelling the conditional

distribution of the target given the nuisance. This requires only a preprocessing step and preserves the computational advantages of mean-field variational Bayes, while ensuring accurate and reliable inference for the target parameter, including for uncertainty quantification. We investigate the numerical performance of our algorithm, showing that it performs competitively with existing methods. We further establish accompanying theoretical guarantees for estimation and uncertainty quantification in the form of a Bernstein–von Mises theorem.

Session G	Contributed session: Graphical Modeling Chair: David Blei
Room	5A
Andrej Srakar	**Pitman-Yor Mixtures For BART: Novel Nonparametric Prior For Bayesian Causal Inference**

Abstract: Recently authors pointed to Bayesian causal inference as important pathway for future development of causal methodologies (Li, 2022; Imbens, 2022). Bayesian additive regression trees (BART) perspective has been developed by Chipman et al. (2010) and popularized in recent years in its usage in regression and causal inference problems (for example Tan and Roy, 2019; Hahn, Murray and Carvalho, 2020). Commonly, BART uses a specific regularization prior, sometimes combined with Gaussian, Dirichlet, Dirichlet Process Mixture (DPM) and semiparametric perspectives (Tan and Roy, 2019). Despite the success of BART, there has been a growing number of papers that point out its limitations. Hahn, Murray and Carvalho have developed Bayesian Causal Forests (BCF - Hahn, Murray and Carvalho, 2020) as a novel regularization approach for nonlinear models geared specifically towards situations with small effect sizes, heterogeneous effects, and strong confounding, to improve on the earlier BART perspective. We develop a novel nonparametric regularization prior for BART based on Pitman-Yor Mixture (PYM) partition-based process which has to date to our knowledge rarely been used in causal inference (but is suggested for classification and mixture modelling).

Pitman-Yor process mixtures (Ishwaran and James, 2001; 2003) are a generalization of DPMs based on the Pitman-Yor process, also known as the two-parameter Poisson‚ÄìDirichlet process. Our novel BART perspective is studied in more detail for several different potential outcome and graphical causal perspectives: regression discontinuity design; causal maximally partially directed acyclic graph (MPDAG); direct causal clause; and causal mediation. Study of asymptotic properties in a Bayesian framework follows Rockova and van der Pas (2017) and Rockova (2019). Our results on simulated and real data confirm improved properties as compared to earlier BART priors and the performance is similar to the Hahn et al. BCF model, but improved in the presence of strong confounding. Performance of the prior is different for causal mediation and we provide suggestions for future work in this perspective. We address computational issues by using importance sampling with the integrated nested Laplace approximation (Outzen Berild et al., 2021). In conclusion we discuss extensions to endogeneity corrections in the line of BCF-IV approach of Bargagli Stoffi et al. (2022) and extensions to Single World Intervention Graph perspective (Richardson and Robins, 2013; 2014).

Pariya Behrouzi	**Bayesian inference of microbiota systems from count metagenomic data**

Abstract: Metagenomics combined with high-resolution sequencing have enabled researchers to study the genomes of entire microbial communities. Revealing the underlying interactions between these communities is of vital importance to learn how microbes influence human health. Learning these interactions from microbiome data is challenging, due to the high dimensionality, discreteness, broad dispersion levels, compositionality and excess of zero counts that characterize these data. To tackle these issues, we develop a novel Gaussian copula graphical model with two key elements. Firstly, we model the marginal distributions via discrete Weibull regression, both to account for the typical features of microbiome data and to include the dependency from external covariates, often available in genomic studies but rarely used for network inference. Secondly, we advance a Bayesian structure learning framework, based on a computationally efficient search algorithm that is suited to high dimensionality. The approach returns simultaneous inference of the marginal effects and of the dependency structure, including graph uncertainty estimates. A simulation study and a real data analysis of microbiome data highlight the applicability of the proposed approach at inferring networks from high-dimensional

count data in general, and its relevance to microbiota data analyses in particular.

Xiaoyue Xi

A modelling framework for detecting and leveraging node-level information in Bayesian network inference

Abstract: Bayesian graphical models are powerful tools to infer complex relationships in high dimension yet are often fraught with computational and statistical challenges. If exploited in a principled way, the increasing information collected alongside the data of primary interest constitutes an opportunity to mitigate these difficulties by guiding the detection of dependence structures. For instance, gene network inference may be informed by the use of publicly available summary statistics on the regulation of genes by genetic variants. In this talk, we present a novel Gaussian graphical modelling framework to identify and leverage information on the centrality of nodes in conditional independence graphs. Specifically, we consider a fully joint hierarchical model to simultaneously infer (1) sparse precision matrices and (2) the relevance of node-level information for uncovering the sought-after network structure. We encode such information as candidate auxiliary variables using a spike-and-slab sub-model on the propensity of nodes to be hubs, which allows hypothesis-free selection and interpretation of a sparse subset of relevant variables. As efficient exploration of large posterior spaces is needed for real-world applications, we develop a variational expectation conditional maximisation algorithm that scales inference to hundreds of samples, nodes and auxiliary variables. We illustrate and exploit the advantages of our approach in simulations and in a gene network study which identifies hub genes involved in biological pathways relevant to immune-mediated diseases.

Deborah Sulem

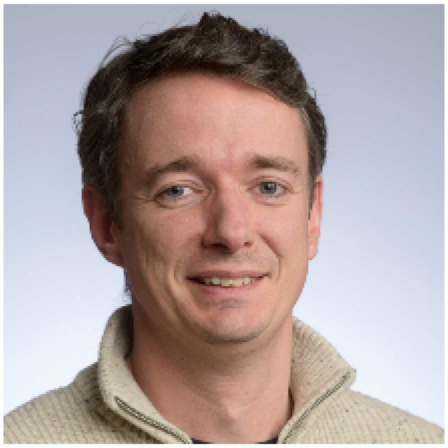
Almost-parallel Bayesian inference for high-dimensional Gaussian graphical models

Abstract: Gaussian graphical models are widely used to analyse the inter-dependence structure among variables such as gene expression and brain fMRI data. Under this model, conditional independence statements are encoded by zero entries in the model's precision matrix. However, the computational demands of estimating a high-dimensional precision matrix have limited the scope of applications when the number of observed variables is large. This work introduces a scalable, interpretable, and fully Bayesian method for estimating precision matrices in high-dimensional settings. Our method capitalises on the relationship between the conditional dependence structure and a linear regression model, and decomposes the high-dimensional estimation problem via row-wise computations. This further facilitates the use of Tempered Gibbs sampling and the Sticky Zig Zag sampler to conduct efficient inference with discrete spike-and-slab priors. Lastly, this approach enables us to parallelise some computations of posterior conditional distributions and fosters an efficient exploration of the network structure.

↑ back to the schedule ↑

1:30 pm **Keynote Lecture: Johannes Schmidt-Hieber**

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



Speaker

Johannes Schmidt-Hieber
Professor of Statistics,
University of Twente

Chair: Aad van der Vaart

Posterior contraction of deep Gaussian processes

Abstract: Deep neural networks are at the heart of the AI revolution and considerable progress has been made in building an underlying statistical foundation. Deep Gaussian process priors can be viewed as continuous analogues of Bayesian neural networks. This raises the question whether there is a closer link with deep learning and to what extend statistical convergence guarantees for deep artificial neural networks also hold for deep Gaussian process priors. For a carefully selected deep Gaussian process prior, we provide a positive answer for learning compositionally sparse functions. Similarly as the convergence rates for deep ReLU networks, deep Gaussian processes can achieve in this setting the fastest possible posterior contraction rates, which do not suffer from the curse of dimensionality. Moreover, we rigorously prove that deep Gaussian process priors can outperform Gaussian process priors in terms of contraction rates if the target function has a compositional structure. This is joint work with Gianluca Finocchio, Matteo Giordano and Kolyan Ray.

↑ back to the schedule ↑

3:00 pm Multi-track session 11

For invited sessions, the organizer will chair unless otherwise stated

<table>
<tr><td>Session A</td><td>Invited session: Bayesian Nonparametric Approaches for Causal Inference with Real-World Complexities
Organizer: Jason Roy</td></tr>
<tr><td>Room</td><td>6A</td></tr>
<tr><td>Liangyuan Hu</td><td>Bayesian nonparametric methods for inferring causal effects of longitudinal treatments amidst missing covariate data</td></tr>
<tr><td colspan="2">Abstract: Missing covariate data is a common issue in longitudinal studies, complicating the process of making causal inferences about longitudinal treatments. A widely recognized solution to this challenge is imputation. Most current imputation techniques rely on parametric models. These models explicitly define relationships among longitudinal responses, treatments, and covariates, which can be intricate in complex datasets. Biases can emerge if these parametric forms are incorrectly specified. To flexibly capture the underlying patterns of data structures, we introduce a novel Bayesian trees mixed-effects model tailored for longitudinal trajectories. Leveraging this model, we present an efficient MCMC algorithm for the sequential imputation of missing longitudinal covariate data. We then seamlessly integrate this innovative methodology for missing longitudinal data with g-computation to analyze the causal effects of longitudinal treatments. Comprehensive simulations are conducted to assess the efficacy of our proposed techniques. Lastly, we employ our methods on a cardiovascular disease study dataset, aiming to determine and validate the most effective rules for initiating antihypertensive treatment.</td></tr>
<tr><td>Jared D. Fisher</td><td>A Bayesian Classification Trees Approach to Treatment Effect Variation with Noncompliance</td></tr>
<tr><td colspan="2">Abstract: Estimating varying treatment effects in randomized trials with noncompliance is inherently challenging since variation comes from two separate</td></tr>
</table>

sources: variation in the impact itself and variation in the compliance rate. In this setting, existing frequentist and machine learning methods are quite flexible but are highly sensitive to the so-called weak instruments problem, in which the compliance rate is (locally) close to zero. Parametric Bayesian approaches, which account for noncompliance via imputation, are more robust in this case, but are much more sensitive to model specification. In this paper, we propose a Bayesian semiparametric approach that combines the best features of both approaches. Our main methodological contribution is to present a Bayesian Causal Forest model for binary response variables in scenarios with noncompliance. In this Bayesian noncompliance framework, we repeatedly impute individuals' compliance types, allowing us to flexibly estimate varying treatment effects among compliers. We then apply the method to detect and analyze heterogeneity in a study of workplace wellness, where there are a plethora of binary outcomes of interest.

Arman Oganisian	**Bayesian Models for Counterfactual Prediction and Optimization with Incomplete Information: Applications in HIV Care Retention**

Abstract: Human immunodeficiency virus (HIV) is managed over successive clinic visits. During each visit, patient features are assessed, treatments are prescribed, and a subsequent follow-up visit is scheduled. Retention, defined as returning within an acceptable window of the scheduled visit time, is essential for continuity of care and is associated with better health outcomes. To guide outreach efforts, there is clinical interest in models that can predict which patients have low retention probability. Moreover, predictions are desired under various counterfactual scheduling decisions to recommend an optimal scheduling time in the sense of maximizing retention probability. Estimation of these models using electronic health records (EHR) is complicated because competing and censoring events make it difficult to ascertain if a patient returned for their scheduled visit, leading to outcome missingness. At the same time, inconsistent monitoring of features such as viral load and CD4 count lead to covariate missingness. We present an all-in-one approach for predicting retention and recommending optimal scheduling times while accounting for these complexities. We first formulate and identify policy-relevant retention estimands. Flexible continuous-time transition models are then used to make posterior inferences about these estimands. The proposed models are fit using EHR from AMPATH, a consortium of clinics that treat HIV in Western Kenya. This work is part of a larger project developing decision support tools for AMPATH that can operate in real-time at the point-of-care.

Session B	Invited session: Sampling-Based Bayesian Computation Organizer: Nicolas Chopin
Room	8A
Francesca Romana Crucinio	**Towards a turnkey approach to unbiased Monte Carlo estimation of smooth functions of expectations**

Abstract: Using Monte Carlo samples with expectation m, we develop a general approach to unbiasedly estimate $f(m)$, where f is a smooth function. Specifically, we develop estimators that are based on randomly truncating the Taylor series expansion of f, and estimating the coefficients of the truncated series. We derive their properties, and propose a strategy to set their tuning parameters -- which depend on m -- automatically, with a view to make the whole approach simple to use. In this talk I will focus on $f(x)=1/x$ and on Bayesian inference for un-normalised models. Joint work with Nicolas Chopin and Sumeetpal S. Singh.

Hai-Dang Dau	**Sampling with diffusion models**

Abstract: Diffusion models (also called score-based generative models) have been hugely successful in the context of generating new samples given training data. However, the sampling setting, in which one wishes to conduct inference from an unnormalized probability density, receives less attention. In this talk, we cover recent advances in applying diffusion models to this problem and we highlight in particular the role of reference measures and sequential Monte Carlo techniques.

Sam Power	**Comparison Theorems for Practical Slice Sampling**

Abstract: Slice sampling is a popular gradient-free MCMC algorithm for approximate sampling from intractable probability distributions, implemented in software packages such as JAGS and TensorFlow Probability. This popularity stems from the intuitive geometric formulation, wide applicability, and general robustness of the algorithm, both in theory and in practice. An outstanding theoretical challenge has been that while the “ideal" slice sampler admits an elegant quantitative convergence theory, practical implementations typically involve additional approximations, which prevent the existing theory from applying as-is. In recent work, we advance a mathematical framework for the analysis of such “hybrid” slice samplers, facilitating novel convergence results for slice sampling as implemented in practice. We provide a number of concrete examples to illustrate the flexibility and practicality of our approach. No prior knowledge of the slice sampling algorithm will be assumed, and relevant theoretical concepts will be recalled as appropriate in the talk. This is joint work with Daniel Rudolf, Björn Sprung, and Andi Wang.

Session C	Invited session: Scalable Bayes for Omics Data Science Organizer: Himel Mallick Chair: Francesco Gaffi
Room	5A
Yiwei Gong	**Detecting State Changes in Dynamic Neuronal Networks**

Abstract: Our brain function depends on temporal modulation of neuronal networks. Neuronal developmental disorders (NDDs) such as autism spectrum disorder and Rett syndrome are associated with certain patterns of modulation. Our research goal is to identify state changes in neuronal activity associated with the onset of NDDs. Existing statistical models for spike-train data focus on changes of amplitudes and correlations, instead of node connectivity; exploratory analysis on spike train data from in-vitro mouse MEA recordings with known timings of light stimulation suggests that they are insufficiently flexible for our purposes. We present a flexible nonlinear model based on factorial hidden Markov Model that is designed to capture the interactive behavior among nodes while characterizing the changes in spike intensity. We propose a variational inference algorithm using Gumbel-Softmax relaxation to infer the latent states and model parameters. The model is designed but not limited to reveal the switching process of brain neurons.

Qiwei Li	**When Bayesian Statistics Meets AI - A Comprehensive Analysis of Spatially Resolved Transcriptomics Data**

Abstract: Recent technology breakthroughs in spatial molecular profiling have enabled the comprehensive molecular characterization of single cells while preserving their spatial and morphological contexts. This new bioinformatics scenario advances our understanding of molecular and cellular spatial organizations in tissues, fueling the next generation of scientific discovery. Bayesian statistics relies more on human analyses with computer aids, while AI relies more on computer algorithms with aids from humans. This talk will outline methodologies for synergizing AI capabilities with Bayesian frameworks, aiming to resolve key issues in this emerging field. Particularly, I will demonstrate how to integrate information from AI tools into Bayesian models and Bayesian nonparametric approaches to enhance spatial domain identification and enable gene expression reconstruction at the single-cell level.

Federico Ferrari	**Bayesian Methods for Multi-omics Data Integration and Biomarker Discovery in the Large P, Small N Setting**

Abstract: Motivated by the problem of inferring interactions among multiple omics features impacting human health outcomes, we propose a dependence-aware Bayesian multi-omics integration method that accounts for correlations in multi-omics data, both within and between omics layers. We demonstrate that the proposed Bayesian method achieves higher predictive accuracy on simulated data and real multi-omics datasets, outperforming published early and late integration approaches in estimation, prediction, and variable selection, while also facilitating uncertainty quantification. The open-source software implementation of our method is publicly available.

Session D	Invited session: Frontiers in Models Based on Gaussian Process Priors: Asymptotics, Applications, and Scalability Organizer: Sanvesh Srivastava Chair: Renato Berlinghieri
Room	10A

Yichen Zhu

Posterior Contraction Rates for Vecchia Approximations of Gaussian Processes

Abstract: Gaussian Processes (GP) are widely used to model spatial dependency in geostatistical data, yet the exact Bayesian inference has an intractable time complexity of O(n^3). Vecchia approximation has become a popular solution to this computational issue, where spatial dependency is characterized by a sparse directed acyclic graph (DAG) that allows scalable Bayesian inference. Despite the popularity in practice, little is understood about its theoretical properties. In this paper, we systematically study the posterior contraction rates of Vecchia approximations of GP. Under minimal regularity conditions, we prove that by appropriate selection of the underlying DAG, the Vecchia approximated GP possess the same posterior contraction rates as the mother GP. Therefore, by optimal choices of the tunning hyper-parameters, the Vecchia approximation can achieve the minimax contraction rate, providing strong frequentist guarantees to the procedure. Our theoretical findings are demonstrated numerically as well using synthetic and real world data sets.

Nan Wu

Inferring manifolds from noisy data using Gaussian processes

Abstract: We focus on the study of a noisy data set sampled around an unknown Riemannian submanifold of a high-dimensional space. Most existing manifold learning algorithms replace the original data with lower dimensional coordinates without providing an estimate of the manifold in the observation space or using the manifold to denoise the original data. We propose a Manifold reconstruction via Gaussian processes (MrGap) algorithm for addressing these problems, allowing interpolation of the estimated manifold between fitted data points. The proposed approach is motivated by novel theoretical properties of local covariance matrices constructed from noisy samples on a manifold. Our results enable us to turn a global manifold reconstruction problem into a local regression problem, allowing the application of Gaussian processes for probabilistic manifold reconstruction. In this talk, I will review the classical manifold learning algorithms and discuss the theoretical foundation of the new method, MrGap. Simulated and real data examples will be provided to illustrate the performance.

↑ back to the schedule ↑

Michael Minyi Zhang

Bayesian Non-linear Latent Variable Modeling via Random Fourier Features

Abstract: The Gaussian process latent variable model (GPLVM) is a popular probabilistic method used for nonlinear dimension reduction, matrix factorization, and state-space modeling. Inference for GPLVMs is computationally tractable only when the data likelihood is Gaussian. Moreover, inference for GPLVMs has typically been restricted to obtaining maximum a posteriori point estimates, which can lead to overfitting, or variational approximations, which mischaracterize the posterior uncertainty. Here, we present a method to perform Markov chain Monte Carlo (MCMC) inference for generalized Bayesian nonlinear latent variable modeling. The crucial insight necessary to generalize GPLVMs to arbitrary observation models is that we approximate the kernel function in the Gaussian process mappings with random Fourier features; this allows us to compute the gradient of the posterior in closed form with respect to the latent variables. We show that we can generalize GPLVMs to non-Gaussian observations, such as Poisson, negative binomial, and multinomial distributions, using our random feature latent variable model (RFLVM). Our generalized RFLVMs perform on par with state-of-the-art latent variable models on a wide range of applications, including motion capture, images, and text data for the purpose of estimating the latent structure and imputing the missing data of these complex data sets.

Alexander Terenin

Sampling from Gaussian Process Posteriors using Stochastic Gradient Descent

Abstract: The ability to deploy Gaussian-process-based decision-making systems such as Bayesian optimization at scale has traditionally been limited by computational costs arising from the need to solve large linear systems. The de-facto standard for solving linear systems at scale is via the conjugate gradient algorithm—in particular, stochastic gradient descent is known to converge near-arbitrarily-slowly on quadratic objectives that correspond to Gaussian process models’ linear systems. In spite of this, we show that it produces solutions which have low test error, and quantify uncertainty in a manner that mirrors the true posterior. We develop a spectral characterization of the error caused by finite-time non-convergence, which we prove is small both near the data, and sufficiently far from the data. Stochastic gradient descent therefore only differs from the true posterior between these regions, demonstrating a form of implicit bias caused by benign non-convergence. We

↑ back to the schedule ↑

conclude by showing, empirically, that stochastic gradient descent achieves state-of-the-art performance on sufficiently large-scale regression tasks, and produces uncertainty estimates which match the performance of significantly more expensive baselines on large-scale Bayesian optimization.

Session E	Invited session: Proposing potential next states in Bayesian computation Organizer: Akihiko (Aki) Nishimura
Room	Magna
Trevor Campbell	**General bounds on the quality of Bayesian coresets**

Abstract: Bayesian coresets speed up posterior inference in the large-scale data regime by approximating the full-data log-likelihood function with a surrogate log-likelihood based on a small, weighted subset of the data. But while Bayesian coresets and methods for construction are applicable in a wide range of models, existing theoretical analysis of the posterior inferential error incurred by coreset approximations only apply in restrictive settings—i.e., exponential family models, or models with strong log-concavity and smoothness assumptions. In this talk I will present general upper and lower bounds on the Kullback-Leibler (KL) divergence of coreset approximations that reflect the full range of applicability of Bayesian coresets. The lower bounds require only mild model assumptions typical of Bayesian asymptotic analyses, while the upper bounds require the log-likelihood functions to satisfy a generalized subexponentiality criterion that is weaker than conditions used in earlier work.

Akihiko (Aki) Nishimura	**MCMC using *bouncy* Hamiltonian dynamics: A unifying framework for Hamiltonian Monte Carlo and piecewise deterministic Markov process samplers**

Abstract: Piecewise-deterministic Markov process (PDMP) samplers constitute a state of the art Markov chain Monte Carlo (MCMC) paradigm in Bayesian computation, with examples including the zig-zag and bouncy particle sampler (BPS). Recent work on the zig-zag has indicated its connection to Hamiltonian Monte Carlo, a version of the Metropolis algorithm that exploits Hamiltonian dynamics.Here we establish that, in fact, the connection between the paradigms extends far beyond the specific instance. The key lies in (1) the fact that any time-reversible deterministic dynamics provides a valid Metropolis proposal and (2) how PDMPs' characteristic velocity changes constitute an alternative to the usual acceptance-rejection. We turn this observation into a rigorous framework for constructing rejection-free Metropolis proposals based on bouncy Hamiltonian dynamics which simultaneously possess Hamiltonian-like properties and generate discontinuous trajectories similar in appearance to PDMPs. When combined with periodic refreshment of the inertia, the dynamics converge strongly to PDMP equivalents in the limit of increasingly frequent refreshment. We demonstrate the practical implications of this new paradigm, with a sampler based on a bouncy Hamiltonian dynamics closely related to the BPS. The resulting sampler exhibits competitive performance on challenging real-data posteriors involving tens of thousands of parameters.

Minh-Ngoc Tran	**Quantum natural gradient for variational Bayes**

Abstract: Variational Bayes (VB) is a critical method in machine learning and statistics, underpinning the recent success of Bayesian deep learning. The natural gradient is an essential component of efficient VB estimation, but it is prohibitively computationally expensive in high dimensions. We propose a computationally efficient regression-based method for natural gradient estimation, with convergence guarantees under standard assumptions. The method enables the use of quantum matrix inversion to further speed up VB. We demonstrate that the problem setup fulfills the conditions required for quantum matrix inversion to deliver computational efficiency. The method works with a broad range of statistical models and does not require special-purpose or simplified variational distributions. Joint work with Anna Lopatnikova.

Session F	Contributed session: Climate Modeling Chair: Toryn Shafer
Room	7A

Marina Silva Paez

Dynamic Deformation: Modeling Nonstationary Covariance in Spatiotemporal Structures

Abstract: In various fields of study, such as health and environmental sciences, researchers frequently encounter data that spans both time and space. Modeling spatiotemporal processes is challenging because the dependency structure in space and time is complex and often non-stationary. To address this problem, our research expands upon anisotropic models rooted in spatial deformation by accommodating temporal variations in spatial deformation. The idea is to map the geographic coordinates of the original space to a new latent space where the assumptions of isotropy and stationarity hold, and deformations are linked over time, allowing for smooth changes. We apply the proposed model to the average temperature of the Southern Region of Brazil, considering the varying spatial dependency structure across different seasons of the year, with the aim of improving predictive accuracy and gaining a better understanding of spatiotemporal interactions in the data.

Gabriel Huerta

A Multivariate Space-time Dynamic Model for Characterizing Pathways Following the Mt. Pinatubo Eruption

Abstract: The effects of solar climate intervention (SCI) efforts are confounded with climate change and natural variability. The June 1991 Mt. Pinatubo eruption resulted in a massive increase of aerosols (in the form of sulfur dioxide) in the atmosphere, absorbing radiation and thus serving as a natural analog for SAI which presents a chance to develop tools for climate attribution of regionalized sources. Our goal is to characterize the multivariate and dynamic nature of the climate impacts following the Mt Pinatubo eruption. To achieve this goal, we introduce a space-time multivariate model that captures correlations between climate effects following an event. Specifically, spatial variation is modeled using a flexible set of basis functions for which the basis coefficients are allowed to vary in time thru a vector autoregressive (VAR) structure. We show how this novel model can be formulated in the Dynamic Linear Model framework and estimated via a customized MCMC approach. Furthermore, we demonstrate how the model characterizes the observed pathways following the Mt. Pinatubo eruption with reanalysis data from MERRA-2.

Lucas Darius Konrad

Bayesian Structural Break Detection for the Identification of Effective Climate Policies

Abstract: Tackling climate change requires swift and decisive actions to reduce emissions. However, choosing the right policies requires knowledge about which of them proved most effective in the past. Most existing approaches in economics evaluate a narrow set of policies, requiring pre-exisitng knowledge about them. In this paper, we introduce new tools to tackle the question of reverse causality in environmental panel data. Despite recent advancements in the field, the estimation of step indicator saturation in a Bayesian manner remains unexplored to this point. Our proposed setup complements methods in the classical statistical literature, and improves upon them. Additionally to being competitive to classical approaches, they allow for a probabilistically sound approach to quantify not only the presence of a break at a given time point, but also the uncertainty surrounding them. Furthermore, they naturally allows for an extension of this notion to intervals in which breaks can occur, which is novel in the literature. With our methods we aim to contribute to this literature and hand applied researchers and policy-makers tools that allows tackling relevant questions in the realm of environmental economics and beyond.

Session G	Contributed session: Advances in Variational Inference Chair: Sameer Deshpande
Room	4A
Alisa Sheinkman	**Variational Bayesian Neural Networks with Shrinkage**

Abstract: While Deep Learning is known for its expressive and powerful models, these models are often overconfident and tend to overfit. The Bayesian approach to deep models is a natural framework to overcome such obstacles by creating ensembles of deep models: Bayesian neural networks provide improved uncertainty quantification with better calibration and robustness to gradient-based attacks. We study a deep generative model generalizing rectified linear networks, where we employ a stochastic relaxation of the activation function and a Polya-Gamma data augmentation trick to render the model which is conditional linear and Gaussian. Additionally, we consider sparsity-inducing global-local Normal-generalized inverse Gaussians (N-GIG) priors on the weights of the network known to provide improvement in the prediction performance of Bayesian deep models.

Diana Cai	**Black-box variational inference with Gaussian score matching**

Abstract: Black-box variational inference (BBVI), a class of algorithms that only requires a differentiable unnormalized target distribution, has enabled fast and flexible approximate posterior inference. Classical approaches to BBVI, which optimize a stochastic evidence lower bound (ELBO), may converge slowly due to high variance gradient estimates. A recent alternative approach---known as score-matching VI---estimates a variational distribution that matches the scores, or the gradients of the log density, between the variational and target densities. For Gaussian variational families, this approach leads to closed-from parameter updates that were observed to converge faster than those based on ELBO optimization. However, when the target is not exactly Gaussian, this exact score-matching approach attempts to solve an infeasible problem, and additional heuristics are needed to ensure its convergence. In this work, we propose a score-matching VI approach that makes these heuristics unnecessary; our framework is based on minimizing a novel score-based divergence and includes the exact score-matching approach as a special limiting case. Within this general framework, we again show that the variational updates can be computed in closed form for Gaussian approximations. In addition, we prove that these updates converge to the target mean and covariance when the target is Gaussian. Finally, we evaluate the performance of these updates on a variety of Gaussian and non-Gaussian target distributions and find that our method typically converges in fewer gradient evaluations than exact score-matching and a leading implementation of BBVI based on ELBO maximization.

Zuheng (David) Xu	**Embracing the chaos: analysis and diagnosis of numerical instability in variational flows**

Abstract: In modern applications, flow-based variational inference methods can require hundreds or even thousands flow layers in order to achieve enough expressiveness. With such a deep architecture, numerical instability can become a concern. In this work, we investigate the impact of numerical instability on the reliability of sampling, density evaluation, and evidence lower bound (ELBO) estimation in variational flows. We first empirically demonstrate that common flows can exhibit a catastrophic accumulation of error: the numerical flow map deviates significantly from the exact map -- which affects sampling -- and the numerical inverse flow map does not accurately recover the initial input -- which affects density and ELBO computations. Surprisingly though, we find that results produced by flows are often accurate enough for applications despite the presence of serious numerical instability. In this work, we treat variational flows as dynamical systems, and leverage shadowing theory to elucidate this behavior via theoretical guarantees on the error of sampling, density evaluation, and ELBO estimation. Finally, we develop and empirically test a diagnostic procedure that can be used to validate results produced by numerically unstable flows in practice.

Daniel Dodd	**Learning-Rate-Free Natural Gradients for Variational Inference**

Abstract: In recent years, the popularity of stochastic natural-gradient-based variational inference methods has surged as scalable solutions for Bayesian inference. Notably, the Bayesian learning rule stands out as a versatile variational inference method, accommodating a wide array of existing learning algorithms and fostering the development of innovative approaches. However, a significant challenge arises from the reliance of these methods on hyperparameters, particularly the learning rate. The need for practitioners to meticulously fine-tune the learning rate is crucial for ensuring convergence to the optimal variational posterior at a suitable rate. In this work, we introduce an innovative learning-rate-free algorithm for stochastic natural-gradient variational inference, eliminating the need for meticulous tuning and offering a more robust and user-friendly approach. We substantiate the efficacy of our approach through a comprehensive set of numerical experiments encompassing various high-dimensional models and datasets, demonstrating competitive performance with existing methods with no need to tune a learning rate.

4.30 pm ISBA General Meeting

Room Magna