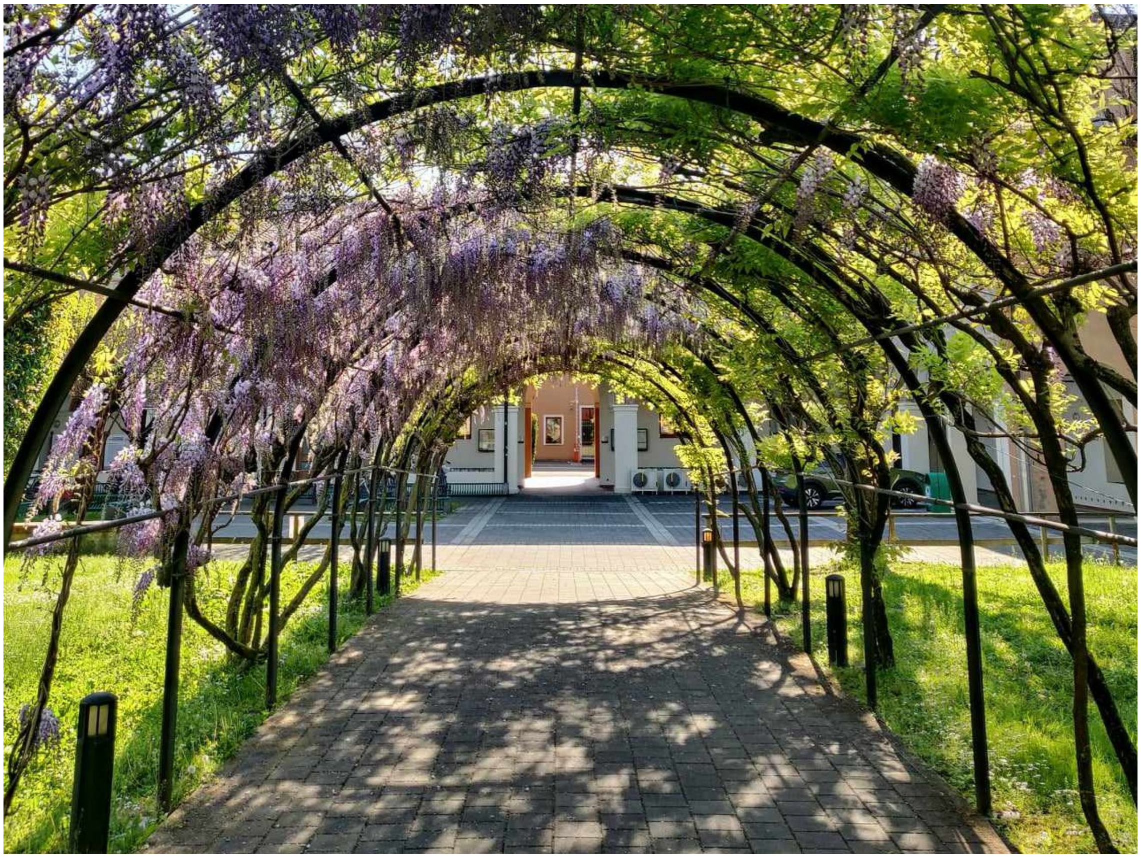

SISBAYES 2025 Workshop

Department of Statistical Sciences
University of Padova

4-5 September 2025

ABSTRACT BOOKLET

siSbayes

Scientific Committee

Raffaele Argiento (chair)
Isadora Antoniano-Villalobos
Veronica Ballerini
Federico Camerlenghi
Antonio Canale
Federico Castelletti
Leonardo Egidi
Brunero Liseo

Organizing Committee

Antonio Canale (chair)
Emanuele Aliverti
Francesco Denti
Stefano Rizzelli

With the support of

Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Università degli Studi di Bergamo

Dipartimento di Scienze Economiche

Endorsed by

ISBA

SIS Società Italiana di Statistica

Contents

Fundational sessions

From Statistical to Causal Models

Guido Consonni
Università Cattolica

The remarkable success of modern Machine Learning (ML) has been driven by access to large datasets, expressive models such as neural networks, and high-performance computing. At its core, this progress has relied heavily on statistical learning, which provides a coherent theoretical foundation. Many of the most celebrated achievements in ML and Artificial Intelligence (AI) are grounded in exploiting statistical dependencies under the assumption that data are independent and identically distributed. However, current challenges in AI—such as robustness to distributional shifts, transferability, and generalization—highlight the limitations of purely statistical approaches. In this context, causal models offer a principled alternative framework. In this talk, I will introduce the key ideas behind causal modeling, focusing on both graphical and structural approaches. I will explain how causal models can be used to evaluate the effects of interventions and answer counterfactual queries. I will also cover the dual challenges of causal discovery (learning causal structure from data) and causal reasoning (answering causal questions once the structure is known). Finally, I will highlight recent contributions from my own research in Bayesian causal inference.

TBA

Sonia Petrone
University Bocconi

TBA

Keynote sessions

A Bayesian theory for estimation of biodiversity

Tommaso Rigon
Università di Milano Bicocca

Statistical inference on biodiversity has a rich history going back to RA Fisher. An influential ecological theory suggests the existence of a fundamental biodiversity number, denoted alpha, which coincides with the precision parameter of a Dirichlet process. Motivated by this theory, we develop Bayesian nonparametric methods for statistical inference on biodiversity, building on the literature on Gibbs-type priors. We argue that sigma-diversity is the most natural extension of the fundamental biodiversity number and discuss strategies for its estimation. Furthermore, we develop novel theory and methods starting with an Aldous-Pitman process, which serves as the building block for any Gibbs-type prior with a square-root growth rate. We propose a modeling framework that accommodates the hierarchical structure of Linnean taxonomy, offering a more refined approach to quantifying biodiversity. The analysis of a large and comprehensive dataset on Amazon tree flora provides a motivating application.

Bayesian models for misreported counts data: theoretical and applied issues.

Serena Arima

Data quality underpins the effective use of information in all data-driven processes. A quality issue typical of surveillance, notification, and other official registers is the misreporting of events. Failing to account for missing cases may lead to severe under/overestimation. When dealing with sanitary data, missing cases may impact vital statistics, affecting the incidence and prevalence of diseases and morbidity and mortality rates in turn, thus jeopardizing the aims of a National Health System (NHS). Defective reporting may occur either because not all cases seek healthcare (under-ascertainment) or due to the inadequacy to detect symptomatic cases that have sought medical advice. We propose a nonparametric compound Poisson model for underreported count data that introduces a latent clustering structure for the reporting probabilities in observation groups. Estimates of the latter are based on model parameters exploiting experts' opinions a proxy for the reporting process. The proposed model is used to estimate the prevalence of the Chronic Kidney Disease (CKD) in Apulia, Italy, based on a unique statistical database covering information on $m = 258$ municipalities obtained by integrating multi-source register information. Accurate prevalence estimates are required for monitoring, surveillance, and management purposes; yet, counts are deemed to be considerably underreported, especially in some areas of Apulia, one of the most deprived and heterogeneous regions in Italy.

We extend the aforementioned idea to count data that might be affected by two sources of measurement error: measurement error leading to extra counts that are not supposed to be counted (add-in or overreporting) and measurement error causing the loss of counts that should have been counted (leave-out or underreporting). We extend the compound Poisson model by incorporating both sources of error: The proposed model is applied to investigate the distribution of the counts of wildfire occur-

rences in the municipal areas of two neighboring Italian regions for the 2021 summer season, which may be affected by both underreporting and double-counting when multiple information sources are employed. Fire counts are obtained by processing MODIS satellite data, while several socio-economic and environmental-driven potential risk factors are also considered in the model formulation. Data from multiple sources with different spatial support are processed to comply with the municipal units. Results suggest the appropriateness of the approach and provide some insights on the features of wildfire occurrences.

Invited Session: Model Based Clustering

Model-Based Clustering of Categorical Data Based on the Hamming Distance

Lucia Paci
Università Cattolica Milano

A model-based approach is developed for clustering categorical data with no natural ordering. The proposed method exploits the Hamming distance to define a family of probability mass functions to model the data. The elements of this family are then considered as kernels of a finite mixture model with an unknown number of components. Conjugate Bayesian inference has been derived for the parameters of the Hamming distribution model. The mixture is framed in a Bayesian nonparametric setting, and a transdimensional blocked Gibbs sampler is developed to provide full Bayesian inference on the number of clusters, their structure, and the group-specific parameters, facilitating the computation with respect to customary reversible jump algorithms. The proposed model encompasses a parsimonious latent class model as a special case when the number of components is fixed. Model performances are assessed via a simulation study and reference datasets, showing improvements in clustering recovery over existing approaches.

Dependent Dirichlet-Multinomial Processes with Random Number of Components

Andrea Cremaschi
IE University

Over the past two decades, Bayesian nonparametrics has expanded to include flexible dependent prior distributions for mixture models, extending beyond univariate species sampling processes to effectively capture dependencies in grouped data under partial exchangeability. While most research has focused on nonparametric priors with almost surely infinite support points, much less attention has been given to almost surely finite-dimensional dependent mixture models under partial exchangeability, despite their strong theoretical properties and promising performance in the exchangeable case. In this work, we explore alternative strategies for defining a multivariate extension of the finite Dirichlet-Multinomial process and its counterpart incorporating a prior on the number of components. Specifically, we introduce a class of flexible dependent Dirichlet-Multinomial processes based on Generalised Wishart unnormalised weights. We analyse their theoretical properties and demonstrate that, unlike existing alternatives, the proposed prior can achieve any desired level of dependence for any fixed number of components. Additionally, our approach allows for efficient posterior computation without the need for costly variable augmentation schemes. We also illustrate the practical advantages of our model through extensive simulation studies and an application to sex-specific gene expression differentiation in the human brain, showcasing its flexibility and computational efficiency in capturing complex dependence structures.

A Bayesian nonparametric approach to discriminant analysis

Bernardo Nipoti
Università degli Studi di Milano-Bicocca

We introduce a Bayesian nonparametric framework to improve classical discriminant analysis, particularly in scenarios with sparse data. Our method provides a flexible approach that encompasses both linear and quadratic discriminant analysis as special cases. The key innovation lies in allowing information sharing across groups to improve the estimation of group-specific covariance matrices. This is accomplished through a scale-only nonparametric mixture model defined on the space of positive definite matrices. The use of a conjugate nonparametric prior ensures tractability and remarkable ease of implementation. Applications to both simulated and real datasets demonstrate the adaptability and effectiveness of the proposed methodology.

BNP4BNP: Bayesian Nonparametric Models for Biomarkers and Neuronal Patterns

Francesco Denti
Università degli Studi di Padova

This talk discusses two Bayesian nonparametric models designed to address the challenges posed by large, complex biological datasets generated by state-of-the-art imaging technologies. The first application focuses on a MALDI-mass spectrometry imaging dataset, which quantifies the abundance of numerous specific molecules across multiple locations within a biological tissue sample. We propose a model for nested, separate exchangeable data, inducing a biclustering to simultaneously group spatial locations with similar abundance profiles and molecules with similar expression patterns. A hidden Markov random field prior is incorporated to enable precise image segmentation, ensuring that clusters correspond to biologically meaningful regions. The resulting biclustering structure facilitates the detection of molecular activation patterns and provides interpretable segmentations of the analyzed image. The second model addresses calcium imaging data, which record the activity of individual neurons over time in freely moving animals. We develop a spatiotemporal mixture model to identify co-activating neurons, detecting groups of cells with a similar firing activity over time. This is achieved through a multivariate time-series framework that detects spikes in calcium traces and captures recurring temporal patterns. We employ a Dependent Dirichlet Process to incorporate information on the anatomical proximity between neurons. Simultaneously, spikes' amplitudes are segmented using a Dirichlet process, providing clusters based on their magnitudes.

Invited Session: Prior elicitation for complex problems

Objective Priors for Measures of Evidence

Laura Ventura
Università degli Studi di Padova

To test a precise (or sharp) null hypothesis on a scalar parameter of interest, Bayesian measures of evidence are the e-value and the Bayesian Discrepancy Measure. In the framework of these measures of evidence, for a parameter of interest we discuss the role of objective matching priors, which are based on higher-order asymptotic expansions. Connections of the e-value and the Bayesian Discrepancy Measure with frequentist inference are highlighted when using these objective matching priors.

Closed-Form Power and Sample Size Calculations for Bayes Factors

Samuel Pawel
University of Zurich

Determining an appropriate sample size is a critical element of study design, and the method used to determine it should be consistent with the planned analysis. When the planned analysis involves Bayes factor hypothesis testing, the sample size is usually desired to ensure a sufficiently high probability of obtaining a Bayes factor indicating compelling evidence for a hypothesis, given that the hypothesis is true. In practice, Bayes factor sample size determination is typically performed using computationally intensive Monte Carlo simulation. Here, we summarize alternative approaches that enable sample size determination without simulation. We show how, under approximate normality assumptions, sample sizes can be determined numerically, and provide the R package bfpwr for this purpose. Additionally, we identify conditions under which sample sizes can even be determined in closed-form, resulting in novel, easy-to-use formulas that also help foster intuition, enable asymptotic analysis, and can also be used for hybrid Bayesian/likelihoodist design. Furthermore, we show how power and sample size can be computed without simulation for more complex analysis priors, such as Jeffreys-Zellner-Siow priors or non-local normal moment priors. Case studies from medicine and psychology illustrate how researchers can use our methods to design informative yet cost-efficient studies. [Pawel, S., Held, L. (2025). Closed-Form Power and Sample Size Calculations for Bayes Factors. The American Statistician, ¡https://doi.org/10.1080/00031305.2025.2467919¿]

Filtering procedures for dynamic multinomial probit models

Augusto Fasano
Università degli Studi di Torino

The multinomial probit constitutes a widely-used model for categorical data in many applications, especially in the econometrics and discrete-choice literature. The computational challenges encountered when fitting this model still motivate ongoing research both from the frequentist and Bayesian viewpoints. In this contribution, we consider a dynamic formulation based on a state-space model where at each time one observes a sample from a multinomial probit with time-specific parameter value. Dependence across time is then induced by the Markovian dynamics of the latent parameter. We show that the filtering and predictive distribution of the latent parameter belong to the unified skew-normal family, developing an associated i..i.d. sampler to approximate quantities of interest via Monte Carlo. Motivated by the computational bottlenecks of the sampler encountered already for moderate sample sizes, we also develop approximate methods for online inference based on assumed density filtering and expectation propagation. This gives more scalable, yet accurate, algorithms for online inference about the latent state and prediction of future observations. Results are shown oversimulated data and a real dataset regarding reservations made from a list of results from an online booking platform.

Invited Session: Bayesian methods for ecological applications and beyond

Taxonomic and covariate-dependent feature allocation models

Federica Stolf
Duke University

Indian Buffet Processes (IBPs) are widely used Bayesian nonparametric models designed for binary latent feature matrices with a potentially infinite number of columns. In biodiversity studies, where features correspond to observed species, this approach is particularly powerful as it allows for the inclusion of an ever-growing number of species. However, current IBP-based models rely on unrealistic assumptions that limit their applicability to ecological data. In particular, they assume that data are exchangeable, ignoring habitat-specific characteristics. Additionally, they model species only at a single taxonomic level, discarding the full taxonomic structure of the data. To address these limitations, we propose a taxonomic and covariate-dependent extension of the IBP that leverages the hierarchical tree structure inherent in the data and incorporates heterogeneity across sampling sites. We discuss theoretical properties of the proposed modeling paradigm and implement efficient algorithms for posterior computation. Crucially, our framework also enables to clearly define the so-called beta-diversity, i.e. the taxonomic heterogeneity of species across different sampling regions, under a coherent and elegant probabilistic framework.

A Bayesian approach to capture-recapture models with misidentification

Andrea Tancredi
Sapienza Università di Roma

The absence of identification errors is a fundamental prerequisite in capture-recapture for population size estimation and species abundance estimation. However, such errors have been reported and studied in both contexts. The most common case is the failure to recognize a previously detected entity, i.e., a false negative record linkage error. This results in artificious entities, sometimes referred to as "ghosts", which typically constitute spurious singletons. We present a Bayesian parametric approach to the problem, which is applicable when data are summarized as number of captures. We develop a Markov chain Monte Carlo algorithm to estimate the proposed model and illustrate the performance of our approach on some datasets available in the microbial diversity literature.

Multivariate species sampling models

Beatrice Franzolini
Bocconi University

Species sampling models provide a structural framework for understanding random discrete distributions in the context of exchangeable observations, as they correspond to the class of exchangeable partitions. However, they do not account for more general structures, such as those that naturally arise in the context of heterogeneous data collected from related sources or under distinct experimental conditions. For instance, this is the case in problems involving the sequential sampling of species across multiple sites, where often the goal is to maximize the number of species discoveries. To address this limitation, we introduce multivariate species sampling models (mSSMs), a general class of models characterized by their partially exchangeable partition probability function. These models encompass most existing Bayesian nonparametric approaches for partially exchangeable data and help elucidate their core distributional properties and the learning mechanisms they induce. Specifically, mSSMs facilitate the study of general properties of random partitions under partial exchangeability assumptions. We demonstrate that the dependence structure is determined by the probability of ties occurring across groups, with independence across sources corresponding to a zero probability of such ties. Furthermore, mSSMs admit three equivalent representations: in terms of the law of the vector of random probability measures, the partially exchangeable random partition, or the sequence of observations. We provide three characterization theorems describing the laws governing these objects. The results presented offer a comprehensive understanding of the dependence structures induced by a wide range of nonparametric models under partial exchangeability assumptions and beyond.

Invited Session: Challenging Posteriors

Full Bayesian Reinforcement Learning via LF-IBIS

Cecilia Viscardi
Università di Salerno

Reinforcement Learning (RL) is a widely used class of methods to support decision-making. Traditional RL focuses on maximizing cumulative rewards based on the feedback an agent receives while interacting with an environment. These methods often suffer from data scarcity — i.e., a lack of knowledge about the dynamics of the environment — as data comes from costly interactions that occur gradually over time. Bayesian Reinforcement Learning (BRL) mitigates this problem by incorporating prior knowledge about the environment and updating it as data is collected. However, this approach requires explicit formalization of the environment's behaviour through a likelihood function, which is often unavailable in real-world scenarios. We propose a full Bayesian Reinforcement Learning (fBRL) strategy aimed at evaluating the posterior distributions for both environment parameters and optimal policies. Our approach is based on Likelihood-free Iterated Batch Importance Sampling (LF-IBIS), a novel algorithm that combines Approximate Bayesian Computation with Iterated Batch Importance Sampling. This hybrid sampling scheme overcomes the definition of an explicit likelihood function and allows an online updating of the posterior distributions as data coming from new interactions becomes available. Finally, we test the effectiveness of our proposal addressing the problem of Response-Adaptive Randomization in clinical trials.

Concentration of discrepancy-based approximate Bayesian computation via Rademacher complexity

Sirio Legramanti
Università d Bergamo

There has been an increasing interest on summary-free solutions for approximate Bayesian computation (ABC) that replace distances among summaries with discrepancies between the empirical distributions of the observed data and the synthetic samples generated under the proposed parameter values. The success of these strategies has motivated theoretical studies on the limiting properties of the induced posteriors. However, there is still the lack of a theoretical framework for summary-free ABC that (i) is unified, instead of discrepancy-specific, (ii) does not necessarily require to constrain the analysis to data generating processes and statistical models meeting specific regularity conditions, but rather facilitates the derivation of limiting properties that hold uniformly, and (iii) relies on verifiable assumptions that provide more explicit concentration bounds clarifying which factors govern the limiting behavior of the ABC posterior. We address this gap via a novel theoretical framework that introduces the concept of Rademacher complexity in the analysis of the limiting properties for discrepancy-based ABC posteriors, including in non-i.i.d. and misspecified settings. This yields a unified theory that relies on constructive arguments and provides more informative asymptotic results and uniform concentration bounds, even in those settings not covered by current studies. These key advancements are obtained by relating the asymptotic properties of summary-free ABC posteriors to the behavior of the Rademacher complexity associated with the chosen discrepancy within the family of integral probability semimetrics (IPS). The IPS class extends summary-based distances, and also includes the widely implemented Wasserstein distance and maximum mean discrepancy (MMD), among others. As clarified in specialized theoretical analyses of popular IPS discrepancies and via illustrative simulations, this new perspective improves the understanding of summary-free ABC.

Sampling on constrained spaces

Elena Bortolato
Universitat Pompeu Fabra

Sampling probability distributions on submanifolds is a relevant task in various problems in statistics, especially for dealing with models defined upon constraints. Specialized Markov Chain Monte Carlo (MCMC) methods were originally designed to address similar challenges in statistical physics and were recently adopted in various statistical contexts, as hypotheses testing, ABC, overparameterized models, Generalized fiducial inference. However, these MCMC methods have proven to be adaptable to a variety of sampling problems in Bayesian statistics which are not naturally defined on submanifolds and offer promising avenues for efficiently sampling probability distributions.

Invited Session: Bayesian Causal Inference

Multivariate Causal Effect: a Bayesian Regression Factor Mode

Dafne Zorzetto
Brown University

The impact of wildfire smoke on air quality is a growing concern, contributing to air pollution through a complex mixture of chemical species with important implications for public health. Although previous studies have focused mainly on its association with total particulate matter ($\mathrm{PM}_{2.5}$), the causal relationship between wildfire smoke and the chemical composition of $\mathrm{PM}_{2.5}$ remains largely unexplored. Exposure to these chemical mixtures plays a critical role in shaping public health, but capturing their relationships requires advanced statistical methods capable of modeling the complex dependencies among chemical species. To fill this gap, we propose a Bayesian causal regression factor model that estimates the multivariate causal effects of wildfire smoke on the concentration of 27 chemical species in $\mathrm{PM}_{2.5}$ across the United States. Our approach introduces two key innovations: (i) a causal inference framework for multivariate potential outcomes, and (ii) a novel Bayesian factor model that employs a probit stick-breaking process as prior for treatment-specific factor scores. By focusing on factor scores, our method addresses the missing data challenge common to causal inference and enables a flexible, data-driven characterization of the latent factor structure, which is crucial to capture the complex correlation between multivariate outcomes. Through Monte Carlo simulations, we show the accuracy of the model in estimating the causal effects in multivariate outcomes and characterizing the treatment-specific latent structure. Finally, we apply our method to US air quality data, estimating the causal effect of wildfire smoke on 27 chemical species in $\mathrm{PM}_{2.5}$, providing a deeper understanding of their interdependencies.

The I-MAP Parameterization of Gaussian DAG Models

Alessandro Mascaro
Universitat Pompeu Fabra

We introduce a novel parameterization of Gaussian Directed Acyclic Graph (DAG) models, called the I-MAP parameterization. Unlike the conventional parameterization through the LDL decomposition of the precision matrix, the I-MAP parameterization always includes a single parameter corresponding to a pre-specified causal effect of interest. This feature proves especially valuable in Bayesian inference of causal effects when the DAG is unknown and must be learned from data. By incorporating this single parameter, our parameterization enables full control over the prior on the causal effect and ensures its consistent specification across all the DAGs that may be considered in a model averaging strategy. We illustrate the utility of the I-MAP parameterizations for Bayesian hypothesis testing of causal effects, and show how it enables consistent prior specification in two-step procedures that first derive the posterior distribution over DAGs and then produce Bayesian point estimates of causal effects under different loss functions.

Evaluating causal effects on time-to-event outcomes in an RCT in oncology with treatment discontinuation

Fabrizia Mealli
European University Institute

In clinical trials, patients may discontinue treatments prematurely, breaking the initial randomization. The ICH E9(R1) Addendum provides guidelines for handling such ”intercurrent events;” the right strategy to adopt depends on the questions of interest. Our study is motivated by a randomized controlled trial (RCT) in oncology, where patients assigned the investigational treatment may discontinue it due to adverse events. We propose adopting a principal stratum strategy and decomposing the overall ITT effect into principal causal effects for groups of patients defined by their potential discontinuation behaviour. We first show how to implement a principal stratum strategy to assess causal effects on a survival outcome in the presence of continuous time treatment discontinuation, its advantages, and the conclusions one can draw. Our strategy deals with the time-to-event intermediate variable that may not be defined for patients who would not discontinue; moreover, discontinuation time and the primary endpoint are subject to censoring. We employ a flexible model-based Bayesian approach to tackle these complexities, providing easily interpretable results. We apply this Bayesian principal stratification framework to analyze synthetic data of the motivating oncology trial. We simulate data under different assumptions that reflect real scenarios where patients' behaviour depends on critical baseline covariates. Supported by a simulation study, we shed light on the role of covariates in this framework: beyond making structural and parametric assumptions more credible, they lead to more precise inference and can be used to characterize patients' discontinuation behaviour, which could help inform clinical practice and future protocols.

Invited Session: Bayesian graphical models

Inference of multiple high-dimensional networks with the Graphical Horseshoe prior

Claudio Busatto
Università di Firenze

We develop a novel full-Bayesian approach for multiple correlated precision matrices, called multiple Graphical Horseshoe (mGHS). The proposed approach relies on a novel multivariate shrinkage prior based on the Horseshoe prior that borrows strength and shares sparsity patterns across groups, improving posterior edge selection when the precision matrices are similar. On the other hand, there is no loss of performance when the groups are independent. Moreover, mGHS provides a similarity matrix estimate, useful for understanding network similarities across groups. We implement an efficient Metropolis-within-Gibbs for posterior inference; specifically, local variance parameters are updated via a novel and efficient modified rejection sampling algorithm that samples from a three-parameter Gamma distribution. The method scales well with respect to the number of variables and provides one of the fastest full-Bayesian approaches for the estimation of multiple precision matrices. Finally, edge selection is performed with a novel approach based on model cuts. We empirically demonstrate that mGHS outperforms competing approaches through both simulation studies and the application to a bike-sharing and a genomic dataset.

Bayesian inference of multiple Ising models for heterogeneous public opinion survey networks

Alejandra Avalos-Pacheco
Johannes Kepler Universität Linz

Public opinion studies show that relationships between opinions shift based on respondent characteristics. Understanding these complexities requires methods that account for heterogeneity across groups. We adopt a class of multiple Ising models that use graphs to analyse how external factors—such as time spent online or generational differences—shape joint dependence relationships between opinions. A Bayesian methodology is proposed based on a Markov Random Field prior, allowing information sharing across groups to encourage common edges when supported by data. A spike-and-slab prior induces sparsity and identifies shared graph structures across subgroups. Specifically, we develop two Bayesian approaches for inferring multiple Ising models, focusing on model selection: (i) a Fully Bayesian method for low-dimensional graphs using conjugate priors and exact likelihood and (ii) an Approximate Bayesian method for high-dimensional graphs based on a quasi-likelihood approach, avoiding computational intractability. These methods are applied to two US public opinion studies: one examining how time spent online affects confidence in political institutions, and another exploring intergenerational differences in opinions on public spending. Our results balance identifying significant edges (both shared and group-specific) with sparsity while quantifying uncertainty, ultimately revealing how external factors shape public opinion dynamics.

Learning block structures in Gaussian graphical models for spectrometric data analysis

Alessandro Colombi
Università degli Studi di Milano-Bicocca

Probabilistic graphical modeling serves as a robust framework for capturing the conditional dependencies among variables that follow a Gaussian distribution. Within such models, each node represents a variable, and the absence of an edge between nodes indicates conditional independence given all other variables. Previous studies have applied this methodology to spectrometric data analysis, aiming at discovering the relationships among substances within a compound by analyzing their spectra. Such a goal has been achieved by coupling smoothing techniques for functional data analysis with a Bayesian Gaussian graphical model on basis expansion coefficients, hence simultaneously smoothing the data and providing an estimate of their conditional independence structure. Empirical evidence from real-world applications has shown that the adjacency matrix describing the underlying graph often presents a block structure. This implies a natural clustering of variables into disjoint groups. In this work, a new prior for Gaussian graphical models is introduced to learn the underlying clustering structure of the nodes. The method builds upon stochastic block models while accounting for the natural ordering of the nodes. The model is employed to analyze fruit purees and discover groups of portions of their spectra.