Bayesian Inference in Stochastic Processes 13

Book of Abstracts

Real Academia de Ciencias
Valverde 22, Madrid, Spain

May 22-24, 2023

With the support of

https://www.icmat.es/congresos/2023/BISP/index.php

Invited senior papers

Alan Gelfand, Shin Shirota, Jorge Castillo-Mateo, Erin Schliep, Rob Schick

Title: Three spatial data fusion vignettes

Abstract:

With increased collection of spatial (and spatio-temporal) datasets, we often find multiple sources that are capable of informing about features of a process of interest. Through suitable fusion of the data sources, we can learn at least as much about the process features of interest than from any individual source. For three different illustrative ecological/environmental applications, this talk will propose suitable coherent stochastic modeling to implement a fusion of these sources. We focus exclusively on approaches that arise through generative hierarchical modeling; the specification could produce the data sources that have been observed. Such modeling enables full inference both with regard to estimation and prediction, with implicit incorporation of uncertainty.

We consider the general setting of points and marks, modeled as [points][marks|points], points in **D**, marks in **Y**. The process can model the points themselves, the marks themselves (ignoring any randomness in the points), or the points and marks jointly. This results in four data types: (i) a point pattern, **S** = (s_1, s_2, ..., s_n), (ii) a vector of counts for sets, {$N(B_k)$, k = 1, 2, ..,K}, (iii) a vector of observations at points, {$Y(s_i)$, i = 1, 2, ..., n}, (iv) a vector of averages for sets, {$Y(B_1)$, $Y(B_2)$, ..., $Y(B_k)$}. We illustrate with two data sources; each can be any one of the four data types. Regardless of how the data are observed, we imagine the process operates at point level. Further, we imagine a stochastic process over **D** which links the two data sources.

The first vignette considers presence/absence data over **D** with one dataset being presence/absence of a species collected at a set of chosen locations. The other data source is in the form of museum/citizen science data, recording random locations where the species was observed. The goal is to better understand the probability of presence surface over **D**. The second vignette considers zooplankton abundance data gathered through two different towing mechanisms. One mechanism is calibrated while the other is not. The goal is to better understand zooplankton abundance over **D**. The third, and most challenging vignette seeks to learn about whale abundance. Here, the two sources are aerial distance sampling data for whale sightings and passive acoustic monitoring data (using monitors on the ocean floor) for whale calls.

Janine Illian

Title: Complex spatio-temporal modelling — an emphasis on communicating with users

Abstract:

There is a strong interest within ecological research in understanding how individuals – plants, animals or other organisms – interact with each other and with the environment. The spatial pattern formed by the locations of individuals in space can reflect both the inter- actions among individuals as well preferences of different species for specific environmental conditions or habitats. A statistical analysis based on spatial or spatio-temporal point pro- cess methodology can analyse these patterns and – as a result – reveal, e.g. specific habitat preferences in a changing environment. Due to the development of computationally efficient model fitting methodology such as integrated nested Laplace approximation (INLA), realistically complex spatial and spatio-temporal models my be formulated and fitted to complex data structures within feasible time. These are now easily accessible to non- statisticians through a large number of different R packages. We as developers of complex statistical methods have a responsibility to support the adequate use of the methods we develop – in particular when users are quantitatively trained, yet non-specialist, scientists.

When introducing these users to our methodology we need to strike the right balance between treating methodology as a mere black box and explaining every single technical detail, while providing an adequate understanding of the methodology that allows users to independently decide on appropriate model choices. This is needed to encourage the use of our methods as well as to establish a fruitful dialogue with the users to improve and successfully build on exciting methods.

In the talk we will briefly discuss the capabilities of inlabru, but we will put a strong emphasis on exploring the need for – as well as approaches to – communicating the methodology well to potential users. This is picking up on the responsibility of model developers to support scientists who use the methodology. In other words, this talk uses the example of the software package inlabru and the associated broad range of statistical methodology to outline an approach to addressing the issue of juggling the right balance between treating an approach as a black box and explaining the every single mathematical detail of a modelling approach. In particular, we will discuss our recent thoughts on and attempts to finding a level of explanation that summarises what the methodology does through focusing on the role of the different model components and how they are reflected in the syntax of the package.

Athanasios Kottas, Chunyi Zhao

Title: Nonparametric Bayesian inference for spatial point patterns over irregular domains

Abstract:

We will present Bayesian nonparametric modeling approaches for spatial point processes recorded over irregular domains. The key building block is a prior model for spatial intensities, based on structured weighted combinations of beta densities, designed to balance inferential flexibility with computational efficiency. The model can be used directly for the Poisson process intensity or for the background intensity in spatial Hawkes processes and space-time Hawkes processes. Combined with the Hawkes process branching structure, the prior models allow for full inference for point process functionals, avoiding the need for approximations to the point process likelihood or to the posterior distribution. The methods will be illustrated with synthetic data sets, as well as with crime data from the city of Boston.

Theodore Kypraios

Title: Bayesian nonparametric inference for stochastic infectious disease models

Abstract

Infectious disease transmission models require assumptions about how the pathogen spreads between individuals. These assumptions may be somewhat arbitrary, particularly when it comes to describing how transmission varies between individuals of different types or in different locations and may in turn lead to incorrect conclusions or policy decisions.

In this talk, we will present a novel and general Bayesian nonparametric framework for transmission modelling which removes the need to make such specific assumptions with regards to the infection process. We use multi-output Gaussian process prior distributions to model different infection rates in populations containing multiple types of individuals. Further challenges arise because the transmission process itself is unobserved, and large outbreaks can be computationally demanding to analyse. We address these issues by data augmentation and a suitable efficient approximation method. Simulation studies using synthetic data demonstrate that our framework gives accurate results. Finally, we use our methods to enhance our understanding of the transmission mechanisms of the 2001 UK Foot and Mouth Disease outbreak.

Oliver Ratmann

Title: Deciphering the true patterns and time trends in human social contacts

Abstract

Social contact studies have been crucial in informing public health policy and parameterizing mathematical infectious disease models during the COVID-19 pandemic. Here, we present a Bayesian non-parametric modelling framework for estimating human social contact patterns by age, gender, time and other factors, the Bayesian rate consistency model. The primary aim of the model framework is to accurately estimate the marked structure in age-specific contact patterns at a high resolution of 1-year age bands in a computationally scalable manner, and we achieve this with particular approximations to multi-dimensional Gaussian process priors. The most attractive component of this fully model-based approach are mathematical consistency constraints on the population-level target quantities, which enable us to decipher contact patterns by 1-year age bands even if the data are coarsely reported by 5-year age bands, or are subject to reporting biases, or longitudinal reporting fatigue. We have applied the Bayesian rate consistency model to several large-scale data sets, from reconstructing cryptic and often under-reported sexual contact patterns in East Africa to ongoing longitudinal contact surveys during and following the COVID-19 pandemic, and we will highlight the central scientific insights from our modelling work in the past year.

Nalini Ravishanker, Namitha Pais, Sanguthevar Rajasekaran

Title: Models for High-Frequency Time Series

Abstract

We describe high-frequency time series modeling, such as intra-day asset prices, in two interesting contexts. The first is a hierarchical framework with latent level correlation to model dependent transaction counts defined at different user specified risk levels. We show how the R-INLA package is useful for approximate Bayesian estimation. Next, we describe univariate and multivariate stochastic volatility models for irregularly-spaced time series, distinguishing between scenarios where the gaps between events (transactions) are assumed to be either fixed or random. We show how the NIMBLE package in R can be used for estimation and forecasting. We illustrate the proposed approaches using intra-day returns from multiple assets.

Invited junior papers

Alejandra Avalos Pacheco

Title: Fast Integrative Factor Models: Applications from Nutritional Epidemiology to Cancer Genomics

Abstract:

Data-integration of multiple studies can be key to understand and gain knowledge in statistical research. However, such data present artifactual sources of variation, also known as covariate effects. Covariate effects can be complex, leading to systematic biases, that if not corrected could lead to unreliable inference. In this talk I will present novel sparse latent factor regression (FR) and multi-study factor regression (MSFR) models to integrate such heterogeneous data. The FR model provide a tool for data exploration via dimensionality reduction and sparse low-rank covariance estimation while correcting for a range of covariate effects. MSFR are extensions of FR that enable us to jointly obtain a covariance structure that models the group-specific covariances in addition to the common component. I will discuss the use of several sparse priors (local and non-local) to learn the dimension of the latent factors. Our approaches provide a flexible methodology for sparse factor regression which is not limited to data with covariate effects. Our models are fitted leveraging novel scalable EM and ECM algorithms as well as variational inference methods. I will present several examples, with a focus on bioinformatics applications. We show the usefulness of our methods in two main tasks: as an unsupervised dimension reduction task to give a visual representation of the latent factors of the data; and as a supervised tool to: (i) provide survival predictions leveraging the obtained factors, or (ii) obtain dietary patterns, associating each factor with a measure of overall diet quality related to cardiometabolic disease risk for a hispanic community health nutritional-data study.

Xenia Miscouridou

Title: Series representations and finite approximations of completely random measures

Abstract:

Infinite-activity completely random measures (CRMs) have become important building blocks of complex Bayesian nonparametric models. They have been successfully used in various applications such as clustering, density estimation, latent feature models, survival analysis or network science. Popular infinite-activity CRMs include the (generalised) gamma process and the (stable) beta process. However, except in some specific cases, exact simulation or scalable inference with these models is challenging and finite-dimensional approximations are often considered. In this work, we propose a general and unified framework to derive both series representations and finite-dimensional approximations of CRMs. Our framework can be seen as a generalisation of constructions based on size-biased sampling of Poisson point process. It includes as special cases several known series representations and finite approximations as well as novel ones. In particular, we show that one can get novel series representations for the generalised gamma process and the stable beta process. We show how these constructions can be used to derive novel algorithms for posterior inference, including a generalisation of the slice sampler for normalised CRMs mixture models. We also provide some analysis of the truncation error.

Ioannis Papageorgiou

Title: Exact Bayesian inference with effective time series models based on context trees

Abstract:

We introduce a collection of statistical ideas and algorithmic tools for modelling and performing exact inference with both discrete and continuous-valued time series. For discrete time series, we describe a novel Bayesian framework based on variable-memory Markov chains, called Bayesian Context Trees (BCT). This is a rich class of higher-order Markov chains that admit parsimonious representations by allowing the memory length of the process to depend on the values of the most recent observations. A general prior structure is introduced, and a collection of methodological and algorithmic tools is developed, allowing for efficient, exact Bayesian inference. It is shown that the prior predictive likelihood (averaged over all models and parameters) can be computed exactly, and that the a posteriori most likely models can be precisely identified. The relevant algorithms have only linear complexity in the length of the data and can be updated sequentially, facilitating efficient online forecasting. The proposed approach is then extended to real-valued time series, where it is employed to develop a general hierarchical Bayesian framework for building mixture models. At the top level, a discrete context, or "state", is extracted from quantised versions of some of the most recent samples preceding it. The set of all relevant contexts are represented as a context tree. At the bottom level, a different real-valued time series model is associated with each state. This defines a very general framework that can be used in conjunction with any existing model class to build flexible and interpretable mixture models. We show that, again, effective computational tools can be developed that allow for efficient, exact Bayesian inference. The proposed methods are found to outperform several state-of-the-art techniques on both simulated and real-world data from a wide range of applications.

Tommaso Rigon

Title: The Stirling-gamma process and its application to Bayesian nonparametrics

Abstract:

We introduce a flexible family of priors for the precision parameter of a Dirichlet process. These priors belong to a novel class of Stirling-gamma distributions, which allow us to define a Stirling-gamma process. Stirling-gamma processes are a particular instance of Gibbs-type processes, and we present their properties in detail. We show how a special case of our Stirling-gamma prior is conjugate to the distribution of the number of distinct values arising from the Dirichlet process. This eases prior elicitation and interpretability when clustering via Dirichlet process mixture models. Finite and asymptotic implications are thoroughly discussed, and we show how using a conjugate Stirling-gamma counterbalances the logarithmic growth of the number of clusters. Finally, we show the computational and practical advantages of a mixture of Dirichlet processes with a Stirling-gamma prior over common alternatives proposed in the literature in applied clustering settings.

Simón Rodríguez Santana

Title: Approximate inference with implicit stochastic processes

Abstract:

When performing Bayesian inference in complex models, Implicit stochastic processes (IPs), a generalization of Gaussian processes (GPs), represent a promising approach to conduct inference. Examples of IPs include, among others, Bayesian neural networks or neural samplers. Although IPs may lack a closed-form expression, samples can be easily extracted from them. This allows IPs to be used as priors over functions to conduct approximate inference directly in function-space, circumventing many degenerate problems related to parameter-space inference. However, although promising, this has proven to be a challenging task. Here we will explore some of the most recent techniques that make use of IPs and point out novel research directions in function-space approximate inference.

Alexander Volfovsky

Title: Stochastic processes, dynamic networks and causal inference

Abstract:

Dynamic network data have become ubiquitous in social network analysis, with new information becoming available that captures when friendships form, when corporate transactions happen and when countries interact with each other. Moreover, data are available about individual actors in the network, including information about the spread of viral (disease or otherwise) processes between individuals in the network. We argue that the dynamics of these processes should be coupled with those of the network evolution in order to improve downstream inference. We then present a case study of stochastic epidemic models where the joint dynamics are modeled as a continuous-time Markov chain such that disease transmission is constrained by the contact network structure, and network evolution is in turn influenced by individual disease statuses. We propose a likelihood-based inference method for this model and describe several innovations in missing data imputation that allows us to perform this inference. We conclude by describing how these types of stochastic network models can be leveraged for design of experiments and policy evaluation.

Yanxun Xu

Title: A Bayesian Decision Framework for Optimizing Sequential Combination Antiretroviral Therapy in People with HIV

Abstract:

Numerous adverse effects (e.g., depression) have been reported for combination antiretroviral therapy (cART) despite its remarkable success on viral suppression in people with HIV (PWH). To improve long-term health outcomes for PWH, 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 Bayesian reinforcement learning framework for optimizing sequential cART assignments. Applying the proposed approach 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.

Xiaotian Zheng

Title: Geospatial Modeling for Non-Gaussian Data Using Nearest-Neighbor Mixture Processes

Abstract:

Non-Gaussian geospatial data are prevalent in various fields such as ecology, climate, and environmental science. In this talk, we introduce a class of geospatial models that provide direct, computationally efficient, probabilistic modeling for non-Gaussian data. We build models from local mixtures of conditional distributions with adaptive weights, using a directed graphical model, with a directed acyclic graph that summarizes spatial nearest-neighbor structure. The resulting class of models is referred to as nearest-neighbor mixture process (NNMP). We study model construction and properties analytically, providing a general strategy for modeling general types of non-Gaussian data. This is achieved through specification of bivariate distributions that define the local transition kernels, with appropriate marginal distributions. The NNMP framework emphasizes the description of non-Gaussian dependence at the data level, in contrast with approaches that introduce a spatial process for transformed data, or for functionals of data probability distribution. Thus, it facilitates efficient, full simulation-based inference. Regarding computation, the framework lays out a new approach to handling spatial data sets, leveraging a mixture model structure to avoid computational issues that arise from large matrix operations. We illustrate NNMP's modeling capacity for different non-Gaussian characteristics such as skewness, heavy tails, compact support, and counts, with synthetic data examples, and with analyses of sea surface temperature observations from the Mediterranean Sea, and species counts from the North American Breeding Bird Survey.

Contributed papers

Hugalf Bernburg, Clemens Elster, Katy Klauenberg

Title: Flexible Bayesian Reliability Demonstration Testing

Abstract:

The aim of this work is to test the reliability of a population: at consecutive times t_i and for a population of devices, we want to decide if at least 100p% of the devices are going to function with probability of at least $1 - \alpha$ until the next predefined time t_{i+1}. For that decision, we will determine minimal sampling plans for the discrete time points $t_1, t_2, \ldots$

For reliability demonstration testing it is common to assume that a lifetime distribution f_θ with few parameters $\theta = (\theta_1, \ldots, \theta_m)$ from some parameter space Θ sufficiently describes the failure times. A popular model is the Weibull distribution including the exponential distribution. We will extend the model of a fixed lifetime distribution to allow more flexibility.

When the family of lifetime distribution $\{f_\theta\}_{\theta \in \Theta}$ is too simple to reflect reality, inference can be misleading. To counteract this, we allow time dependent $\theta = \theta(.)$ which are inferred only from the previous $M \geq m$ quantities $k_{i-M-1}, \ldots, k_{i-1}$, the numbers of defective devices. A large M is adequate if the parameters of the lifetime model are believed to be constant for a longer time period. Inferring and inserting a posterior distribution on the parameter θ leads to a self-exciting counting process generating the probabilities for the increment of the number of defective devices, $k_i - k_{i-1}$. That is, we model k as a Markov chain of order M with transition probabilities coming from the counting process. Bayesian inference on k_i is made on the basis of samples drawn without replacement at time t_i and by using prior knowledge gained from all previous tests, together with the assumption of a Markov chain. In our setting k_i is never observed. Therefore, the model is a hidden Markov model. The developed model allows for temporal variations in θ and the fully Bayesian inference of k_i accounts for the uncertainty in θ.

This idea of generalizing lifetime distributions is quite general. For illustration, we will focus on the case of the exponential lifetime distribution. An example of application is a population of utility meters which are subject to section 35 of the *Mess- und Eichverordnung* (German Weights and Measures Act). For that example, we will provide minimal sampling plans that limit the posterior consumers' risk α of wrong decisions on the population's future state. We will apply the minimal sampling plans on simulated Weibull data to demonstrate the balance between cost-effectiveness and flexibility of the proposed adaptive reliability demonstration test.

Rachel Binks, Sarah Heaps, Darren Wilkinson

Title: Determining the order of stationary vector autoregressions with application to modelling brain activity of patients with epilepsy

Abstract:

All vector autoregressive processes have an associated order *p*; conditional on observations at the preceding *p* time points, the variable at time *t* is conditionally independent of the earlier history of the process. Learning the order of the process is therefore important for its characterisation and subsequent use in forecasting. For example, the order can serve as a point of comparison between different data sets and informs the decomposition of the time series into latent processes, which provides information about the underlying dynamics. It is common to assume that a vector autoregressive process is stationary. A vector autoregression is stable if and only if the roots of its characteristic equation lie outside the unit circle, which constrains the autoregressive coefficient matrices to lie in the stationary region. Unfortunately, the geometry of the stationary region can be very complicated, and specification of a prior distribution over this region is difficult. In this work, the autoregressive coefficients are mapped to a set of transformed partial autocorrelation matrices which are unconstrained, allowing for easier prior specification, routine computational inference, and meaningful interpretation of their magnitude. The multiplicative gamma process is used to build a prior distribution for the unconstrained matrices, which encourages increasing shrinkage of the partial autocorrelation parameters as the lag increases. Posterior inference is performed using Hamiltonian Monte Carlo via the probabilistic programming language Stan. Samples from the posterior distribution of the order of the process use a truncation criterion which is motivated by classical theory on the sampling distribution of the partial autocorrelation function. The work is applied in a simulation study to investigate the agreement between the posterior distribution for the order of the process and its known value, with promising results. The model and inferential procedures are then applied to EEG data from epilepsy patients in order to assess differences in the structure of the time series across different frequency bands.

Jose Manuel Camacho, William N. Caballero, Tahir Ekin, Roi Naveiro

Title: Manipulation of Hidden-Markov-Model Inferences through Poisoned Batch Data

Abstract:

Time-series models assume that the data they receive is valid and trustworthy. Nevertheless, an adversary may be motivated to manipulate the data, causing incorrect inferences by decision-makers. This study focuses on an area of adversarial machine learning that has received less attention: poisoned hidden Markov models. The research uses Adversarial Risk Analysis to develop poisoning problems for filtering, smoothing, and decoding inferences. We designed multiple stochastic programming models that account for realistic uncertainties and different attacker objectives. In particular, we developed solution methods using frequentist and Bayesian approaches to solve these problems. We tested the presented methods extensively and evaluated their efficacy based on their solution quality and computational effort. The study reveals the vulnerabilities of hidden Markov models to adversarial activities, emphasizing the need for robust techniques to ensure their security.

Hieu Cao, Simon Wilson, Reme Sillero

Title: Bayesian approaches to covariate shift in Gaussian process regression, with application to estimating galaxy red-shift

Abstract:

In the regression setting, it is typically assumed that the distributions of covariates in the training and test sets are similar, but this assumption is not always valid. Where they are different, but the same model for the response is assumed, is known as the covariate shift problem. In the presence of covariate shift, for data with complex, non-Gaussian distributions, it is also crucial to estimate the full conditional density accurately in order to quantify prediction uncertainty.

An example arises in photometric surveys of galaxies, where the brightness of galaxies is measured in 4 or 5 colour bands, which are then used to estimate the galaxy's red-shift, or speed at which it is receding from us. This is used as a proxy for its distance, allowing for studies of the large-scale structure of the universe. Such models are trained on galaxies from which a full spectroscopic analysis has been obtained, allowing for accurate determination of the red-shift. The covariate shift arises because the galaxies in this training set are brighter than those for which photometry is the only observation.

To address these challenges, we propose a Bayesian approach that employs observation weighting to estimate the conditional density under covariate shift. We consider three different methods for reweighing observations: kernel means matching, optimal transport, and neighbourhood component analysis. Furthermore, we investigate how to incorporate additional features, such as the spatial information of galaxies, into a Gaussian Process model while integrating our reweighting methods. We evaluate our approach using simulated data and demonstrate its effectiveness in mitigating the impact of covariate shift and improving prediction accuracy.

Lorenzo Cappello, Oscar Hernan Madrid Padilla

Title: Variance change point detection with credible sets.

Abstract:

The detection of change points - when and how many times the distribution underlying an ordered data stream experiences a change - is a field with a long history. The collection of large quantities of data enabled by new technologies - e.g., wearable devices, telecommunications infrastructure, and genomic data - has fostered a renaissance of the field. To analyze these data sets, we cannot assume relatively rigid structures where parameters are shared across all observations. Change points define partitions of the data where, within each segment, assumptions like exchangeability are not violated; hence, standard methods can be used. Given the growing needs discussed above, lots of work has been dedicated to this problem lately. The recent research efforts have delivered tremendous advances, including algorithms with low complexity, new mathematical tools to analyze their theoretical properties, and methods suitable for multivariate and non-Euclidean data. Most of these methodologies provide point estimates of the changes' number and locations. An underdeveloped aspect of change point detection is uncertainty quantification, in the sense of being able to provide a set of times instances containing the location of the change at a prescribed level of significance. Bayesian change point methods offer a natural way to quantify uncertainty. Despite this obvious benefit, the Bayesian literature has not kept pace with recent advances in change point detection, and practitioners do not commonly employ these methods. The main reasons are the high computational burden required and the limited literature on statistical guarantees available for these methods. If one resorts to an existing Bayesian change-point method, with few exceptions, one will use a methodology slower and with fewer guarantees than most state-of-the-art competitors. Indeed, practitioners do not commonly employ these methods, and the Bayesian literature has not kept the pace of what has been done in statistics and machine learning. In the talk, we will focus on the specific task of detecting changes in the variance of a Gaussian sequence model, and quantifying the uncertainty in the change point locations and providing a scalable algorithm for inference. The need to add a measure of uncertainty associated with changes in variance is motivated by data from an experiment that studies a new technique to determine whether a liver is viable for transplant. In our proposal, we frame the model as a product of multiple single changes in the scale parameter. We fit the model through an iterative procedure similar to what is done for additive models. The novelty is that each iteration returns a probability distribution on time instances, which captures the uncertainty in the change point location. Leveraging a recent result in the literature, we can show that our proposal is a variational approximation of the exact model posterior distribution. We study the convergence of the algorithm and the change point localization rate. We are able to show that our proposal is optimal in a minimax sense. In simulations, we show that our method offers point estimates as accurate as state-of-the-art methodologies (at times even more accurate), and the uncertainty underestimation typical of VB is mild. Overall, at little additional computational costs, our proposal provides points estimates as precise as those of competitors and a measure of uncertainty that competitors lack. While we will focus on changes in variance, many of the ideas discussed in the talk generalize to other parametric settings. We discuss a few of these generalizations and conclude with some open problems, illustrating our next research steps.

Jorge Castillo-Mateo, Alan E. Gelfand, Jesus Asin, Ana C. Cebrian

Title: Quantile autoregression

Abstract:

Quantile regression (QR) models the relationship between a set of predictor variables and specific quantiles of a response variable. The usual approach called multiple QR fits a separate regression for each quantile of interest, leading to the possibility of crossing of the regression across quantiles. To solve this limitation, the so-called joint QR avoids quantile crossing by modeling all quantiles simultaneously. Working with time-series data, we substitute the set of predictor variables for the response variable on its own previous values, leading to quantile autoregression (QAR) modeling. In a fully Bayesian setting, we propose a characterization of the non-crossing QAR(1) model using a decomposition into two monotone curves. We model these curves using the flexible class of Kumaraswamy cumulative distribution functions. We extend this modeling to the QAR(p) case and the multivariate QAR(1) case. Finally, we extend the modeling to the spatial QAR(1) case, we add spatially varying coefficients to obtain a spatially varying QAR and we capture the remaining spatial dependence in the quantile levels through a copula process. We illustrate these methods with a study of persistence in daily maximum temperature series collected in the northeast of Spain.

Allan Clark, Res Altwegg

Title: A Gibbs sampler for multi-species occupancy models

Abstract:

Multi-species occupancy (MSO) models use detection-nondetection data from several species observed at different locations to estimate the probability that a particular species occupies a particular geographical region. The models are particularly useful for estimating the occupancy probabilities associated with rare species since they are seldom observed when undertaking field surveys. In this paper, we develop Gibbs sampling algorithms that can be used to fit various Bayesian MSO models to detection-nondetection data. Bayesian analysis of these models can be undertaken using statistical packages such as JAGS, Stan and NIMBLE, however, since these packages were not developed specifically to fit occupancy models, one often experiences long run-times when undertaking analysis.

In a single season (single species) nonspatial and spatial occupancy modelling context, Clark and Altwegg (2019), show that special-purpose Gibbs samplers can produce posterior chains that mix faster and have larger expected sampling rates (Holmes and Held (2006)) than those obtained using JAGS and Stan. These results suggest that such algorithms could potentially lead to significant reductions in the run times of MSO models.

This paper illustrates how to fit MSO models when the detection and occupancy processes are modelled using logistic link functions and apply these methods to a camera-trapping study undertaken by Drouilly et al. (2018). Variable selection is undertaken using a reversible-jump Markov chain Monte Carlo (Barker and Link (2013)) algorithm. We found that the Gibbs sampling algorithm developed produces posterior samples that are identical to those obtained when using Stan, resulting in faster run times and has a larger expected sampling rate than Stan when analysing the above-referenced data set.

Sonali Das, Najmeh Nakhaei Rad

Title: A Conditional Bayesian Cylindrical Approach to Predict Direction of Seasonal Extreme Wind Speed

Abstract:

In this paper we propose a Bayesian approach to predict wind direction for extreme seasonal wind speeds. From a deployment of resources perspective during high fire danger index seasons that include extreme wind speed, it is vital to know the direction of these high-speed winds for efficient and effective response. With this aim, we extend a conditional approach to a cylindrical model, which is a combination of Gumbel distribution, as the extreme value distribution, and the circular sine-skewed von Mises distribution (Kalaylioglu, 2022), to capture the direction for extreme wind speeds. Using the posterior predictive distribution, we predict the direction of extreme wind speed for four seasons using hourly wind data from Polokwane, South Africa.

Reference:

Kalaylioglu, Z., 2022. Analysis of correlated circular and extremal data with a flexible cylindrical distribution. *Environmental and Ecological Statistics*, *29*(1), pp.207-222.

Ramon Diaz-Uriarte, Adrián López López

Title: Fitting cancer progression models using Bayesian approaches and probabilistic programming languages

Abstract:

Cancer progression models (CPMs) try to identify restrictions and stochastic dependencies in the order of accumulation of mutations from cross-sectional data. These procedures have also been used to try to estimate short- and long-term tumor evolution and predictability and might be used to choose therapeutic targets. The evolutionary and sampling assumptions of these models are, however, unrealistic and overly restrictive (which also affects their interpretation); moreover, it is difficult to do model selection/expansion, model assessment, and to incorporate uncertainties in derived predictions. We will argue that these issues could be dealt with by using general-purpose probabilistic programming languages (PPLs) in a Bayesian (or approximate Bayesian) framework; this would allow us to flexibly modify and relax the evolutionary and sampling assumptions, and explore a wider range of better scientific models. We will present some results of our use of PPLs to fit CPMs; this work takes advantage of the fact that many CPMs, under specific scenarios, directly specify a transition rate matrix between genotypes. We will end with open questions and future directions. Please note that this is still early work, so expect much more speculation than solid conclusions and results.

Tahir Ekin, R. Muzaffer Musal, Tevfik Aktekin

Title: Bayesian Spatial Analysis of Socioeconomic Determinants on COVID-19 Mortality

Abstract:

This paper introduces statistical modeling strategies for assessing the effects of socioeconomic factors such as poverty, income level, and income inequality on COVID-19 mortality across the different phases of the pandemic. In doing so, we consider Bayesian spatial and non-spatial models, discuss relevant inference results, and introduce spatio-temporal extensions. Our findings indicate that deteriorating socioeconomic factors lead to higher mortality rates when we accurately account for spatial effects across neighboring units. In addition, we investigate the effects of temporal variations in socioeconomic covariates on relevant spatial units over time. We provide insights that can be useful for policy makers and public health decision makers. Our numerical analysis focuses on publicly available data merged from various federal as well as state level sources, with an emphasis on the state of California.

Daniel Fernandez-Sanchez, Daniel Hernandez-Lobato

Title: Joint Entropy Search for Multi-Objective Bayesian Optimization with Constraints and Multiple Fidelities

Abstract:

Many optimization problems involve multiple objectives and constraints. For example, when designing a robot, we may want to maximize its speed and minimize its power consumption. We may also want to make it robust to pass some stress tests. To get a faster robot we may need more power supplies. For less power consumption we may need lighter and weaker materials. Finally, for robustness we may need stronger and heavier materials. An extra challenge is that we do not have the analytical form of the objectives nor the constraints, thus we may have to build the robot for their evaluation. Thus, the evaluations are expensive and they can be contained with noise. When the objectives are conflictive (i.e., they prefer different configurations), the solution to the problem is an infinite set of optimal points in the feasible space called the Pareto set X^*. Evaluating X^* at the objectives yields the Pareto front Y^*, an infinite set of points that shows the best objective trade-off. To solve these problems more efficiently, one can often reduce the evaluation cost of the objectives and the constraints in exchange of reducing fidelity by a certain degree. In the previous robot example, to reduce the evaluation cost, we could perform simulations instead of building new robots each time. Fidelities of different levels correlate with each other and hence, the optimization cost can be reduced if we take advantage of it. Multi-objective Bayesian optimization with constraints and multiple fidelities (MOBOCMF) can be used to solve these problems efficiently while minimizing the number of evaluations and taking advantage of fidelity correlations. MOBOCMF relies on two key pieces: (i) a probabilistic model that estimates the potential values of objectives and constraints at unknown locations, and (ii) an acquisition function that defines the expected benefit of performing an evaluation at each location for each fidelity. To model the correlations among fidelities we use Multi-fidelity Deep Gaussian Processes (MFDPGs). MFDPGs are multi-layer networks with several GPs connected in a row, one per each fidelity level. Correlations are introduced by considering the output of fidelity level *i* when computing the output of fidelity level *i + 1* and so on. We propose a novel acquisition function Joint entropy search for multi-objective Bayesian optimization with constraints and multiple fidelities (JESMOC-MF) that is able to choose not only at which location to perform the next evaluation, but also which fidelity level to use. Specifically, JESMOC-MF, at each iteration, chooses the input and fidelity level that is expected to reduce the most the entropy of the problem's solution $\{X^*, Y^*\}$. This entropy computation is intractable. Nevertheless, we propose an efficient approximate inference algorithm that can be used to approximate it based on variational inference.

Mario Figueira, Xavier Barber, David Conesa, Antonio López-Quílez, Joaquín Martínez-Minaya, Iosu Paradinas, Maria-Grazia Pennino

Title: Improving ecological modeling using Bayesian feedback

Abstract:

In ecology we may find scenarios where the same phenomenon (species occurrence, species abundance, etc.) is observed using two different types of samplers. For instance, species data can be collected from scientific surveys with a completely random sample pattern, but also from opportunistic sampling (e.g., whale or bird watching fishery commercial vessels), in which observers tend to look for a specific species in areas where they expected to find it.

Species Distribution Models (SDMs) are a widely used tool for analyzing this kind of ecological data. Specifically, we have two geostatistical models available for the above data: an independent model (IM) for the data coming from a complete random sampler and a preferential model (PM) for data from opportunistic sampling.

In this work, we propose a sequential Bayesian procedure to connect these two models through the update of prior distributions. Although, the analysis is restricted to this scheme because of its strong presence in ecology, we also expose the general structure for performing Bayesian feedback between an arbitrary number of models. Implementation of the Bayesian paradigm was done through the integrated nested Laplace approximation (INLA) methodology. Additionally, we have also used the stochastic partial differential equations (SPDE) approach, a good option to make inference and prediction in spatial models with high performance and low computational costs. The sequential procedure has been evaluated by simulating several scenarios and comparing the results of sharing information from one model to another using different criteria.

Our main results imply that, in general, it is better to share information from the independent (completely random) to the dependent model than the alternative way. However, it depends on different factors such as spatial range or the location resulting from sample distributions.

Rafael Jiménez Llamas, Emilio Carrizosa Priego, Pepa Ramírez Cobo

Title: New approaches to Fairness in a Bayesian Setting

Abstract:

A new approach to the problem of Fairness is pursued in a Linear Regression setting using Bayesian tools. Fairness in Machine Learning aims to avoid discriminatory predictions that may be induced because of the bias present in the training data. In our work, we create a method which selects the hyper-parameters, analogously to Empirical Bayes, in order to create a tradeoff between a certain measure of Fairness and the accuracy. Furthermore, our considered measure of Fairness avoids privacy concerns by considering the average differences between sensitive and non-sensitive populations. We also explore an adaptation of this approach to the problem of classification. An approach using Variational Bayes and a penalization term due to Fairness will be used to induce Fairness into the solution, in order to obtain a tradeoff between accuracy and Fairness.

Chaitanya Joshi, Paul Brown, Stephen Joe

Title: Computational Bayesian inference using Low Discrepancy Sequences within Integrated Nested Laplace Approximations

Abstract:

Integrated Nested Laplace Approximations (INLA) provides computationally fast and accurate approximate Bayesian inference for latent Gaussian models (LGM). LGMs are typically hierarchically constructed, with an assumed Gauss Markov dependence structure for latent parameters. Since the development of INLA, there has been an interest in developing computationally efficient alternatives to the numerical integration methods INLA employs to compute posterior marginal distributions, such as grid-based approximations. Although grid-based methods work very well in low-dimensional (hyper)parameter space, the approximations become computationally prohibitive with LGMs that have a hyperparameter space generally greater than five.

Novel marginalisation approaches have been developed as an alternative to grid-based methods. The developers of INLA constructed an algorithm to bypass numerical integration and construct unimodal half-Gaussian approximations to posterior marginals (Martins *et al.,* 2013). Sparse grids have also been proposed by Hewitt and Hoeting (2019) as a more computationally efficient and accurate approach.

Recently, Joshi *et al.,* (2021) proposed the use of low discrepancy sequences (LDS) in the marginalisation process. These LDS-based methods were shown to be more efficient than grid-based approaches and were proven to converge to the true posterior marginal. Brown *et al.,* (2020) provided several approaches for implementation of LDS-based methods into INLA to provide approximations to posterior marginals of hyperparameters. Further work on this front has shown that more computational gains can also be made when using LDS to estimate posterior marginals of latent parameters.

Juan Maroñas, Daniel Hernández-Lobato

Title: Efficient Transformed Gaussian Processes for Non-stationary Dependent Multiclass Classification

Abstract:

This work introduces the Efficient Transformed Gaussian Process (ETGP), a new way of creating C stochastic processes characterized by: 1) the C processes are non-stationary, 2) the C processes are dependent by construction without needing a mixing matrix, 3) training and making predictions is very efficient since the number of Gaussian Processes (GP) operations (e.g. inverting the inducing point's covariance matrix) do not depend on the number of processes. This makes the ETGP particularly suited for multi-class problems with a very large number of classes, which are the problems studied in this work. ETGP exploits the recently proposed Transformed Gaussian Process (TGP), a stochastic process specified by transforming a Gaussian Process using an invertible transformation. However, unlike TGP, ETGP is constructed by transforming a single sample from a GP using C invertible transformations. We derive an efficient sparse variational inference algorithm for the proposed model and demonstrate its utility in 5 classification tasks which include low/medium/large datasets and a different number of classes, ranging from just a few to hundreds. Our results show that ETGP, in general, outperforms state-of-the-art methods for multi-class classification based on GPs, and has a lower computational cost (around one order of magnitude smaller).

Jan Münch, Ralf Schmauder, Klaus Benndorf

Title: Bayesian filtering for macroscopic hidden Markov Models (HMM): Potential and limits of minimal informative to improve parameter inference

Abstract:

Inferring the complex functional dynamics of ion channels from ensemble currents or combined with fluorescence data is a daunting task because the chemical-reaction network is only partially observable and the signal is usually highly aggregated. We previously addressed this problem by applying a parallelized Bayesian filter to specify kinetic schemes for macroscopic current and fluorescence data leading to a more accurate likelihood than previous gold-standard algorithms (Muench 2022 eLife 11:e62714). Our algorithm is based on an Fokker-Planck approximation of the first-order chemical reaction network. We generalized the signal model of the previous work of (Moffat 2007 Biophys. J) and are able to report closed-form filtering equations beyond the Kalman filter solution to filter data from a first-order chemical reaction network (Moffat 2007 Biophys. J). Thus, we derived a likelihood which is correct up to the second statistical moment even if state dependent noise contributions in the signal are included into the likelihood. We show that including the details of the single-channel Markov dynamics by using the derived Gaussian Markov kernel for the macroscopic signal supersedes classical approaches with rate equations in two ways. Firstly, the sampled posterior is closer located to the true parameter values. Secondly, the Markov kernel produces a correct uncertainty quantification while a deterministic algorithm, even for large ensembles of chemically reacting molecules fails to cover the true values. The likelihood misspecification caused by the typically used deterministic approximation by a rate equation causes the posterior to lose a large part of its meaning. It is too confident for any finite data amount. Using Bayesian statistics requires defining a prior distribution. When little information about the parameter is known, especially when the information content in the data is poor, the prior is crucial to make the posterior as sensitive as possible to the data. For ion-channel HMMs, a minimal informative prior may consist of a log uniform distribution for the inverse dwell times of a state and a Dirichlet distribution for the probability of each transition out of the state. Applying this prior reduces the number of ion channels required for a reasonable inference by one order of magnitude compared to the standard uniform prior, which is often considered mistakenly as uninformative. However, Ion currents from patch-clamp experiments observe only partially the dynamics of interest in the chemical network. We show by simulated patch-clamp data that this partial observability causes the likelihood to become flat in many directions in the parameter space, causing a degree of practical parameter nonidentifiability for most non-trivial hidden Markov models. The log uniform distribution of the minimal informative prior desensitizes the posterior to the non-identifiability problem of the likelihood for some part of the parameter space. Nevertheless, the posterior will always be dominated by the structure of the prior in the rest of the parameter space. Thus, all posteriors of HMM models will be improper if equipped with minimal informative improper prior distributions. Here, we discuss how to recognize and treat this practical parameter non-identifiability problem, inherent in most HHMs and any statistical framework used. Including other physically justified constraints may tremendously increase the inference quality and decrease the parameter non-identifiability problem. We conclude that: First for the poster to be meaningful the likelihood for macroscopic data needs to reflect the

Markov behavior of the single molecules. Second, given a finite data amount the quality of the model inference can be significantly improved by selecting a minimal informative prior.

Luis A. Ortega, Simón Rodríguez Santana, Daniel Hernández-Lobato

Title: Variational Linearized Laplace Approximation for Bayesian Deep Learning

Abstract:

Pre-trained deep neural networks can be adapted to perform uncertainty estimation by transforming them into Bayesian neural networks through methods such as the Laplace approximation (LA). However, these approximations do not retain the previous fine-tuned solution as the mean of its predictive distribution, changing the model's predictions. The linearized LA (LLA) tackles this issue by linearizing the deep model over its parameters, maintaining the pre-trained predictions. To make these methods more tractable, the generalized Gauss-Newton (GGN) approximation is often used. However, due to other inefficiency difficulties, LA and LLA rely on further approximations, e.g. Kronecker-factored or diagonal approximate GGN matrices, which affect the results. Here, we propose a new method for scaling LLA using a variational sparse Gaussian Process (GP). The main counterpart of this approach is that, as in LA, the mean of the sparse approximation does not coincide with the pre-trained solution. However, making use of the dual formulation for the sparse approximation of GPs, we generalize the original approximation in terms of sparse GPs to a family of models where the predictive mean can be fixed to a particular function, e.g. the pre-trained solution. As a result, our method is capable of retaining the predictive mean of the original model while allowing for efficient stochastic optimization and scalability in both the number of parameters and the size of the training dataset. Intuitively, a set of inducing locations is employed to tune the uncertainty of the predictive distribution of the model while maintaining the predictive mean of the original pre-trained model. The variational optimization of the ELBO allows for optimizing the model with a cost independent of the number of training points, which represented a clear bottleneck of other methods. Preliminary experiments indicate that such an approach outperforms already existing efficient variants of LLA based on alternative sparse approximations, such as accelerated LLA (ELLA).

Jaime Perez, Edmond Awad, Mario Castro, Gregorio Lopez

Title: Expert vs. data-driven causal discovery: A case study on Cyberbullying

Abstract:

Surveys are commonly used in social science research, but analyzing the results can be problematic when implying causal relationships. This is especially true in studies where randomized control trials are not possible. Cyberbullying is a significant issue for young people, with 10% of European youth on average being victims of it and 49% experiencing it at least once. Preventing cyberbullying faces three main challenges: a lack of data on the motivations of minors online, the under-reporting of events, multiple contributing factors, and the need to consider ethical constraints to avoid harm. Here we present a hybrid methodology where we use Bayesian Networks (BN) [3] to combine insights from literature expert knowledge and our survey data. The structure of the BN is represented as a directed acyclic graph (DAG) that encodes the statistical and causal dependencies between variables (i.e., joint probability distributions). Causal inference is a newer scientific philosophy that can help researchers better understand causal relationships between variables. Although data-driven learning is the most popular approach nowadays, it is unclear whether learning algorithms to discover causal connections produce optimal solutions. We also automatically generate new data-driven BN using k-fold cross-validation and evaluating the BN structures through the (log) likelihood $L(\vartheta \mid X)$, where ϑ represents the BN with its fitted parameters, and X represents unobserved data samples. Table 1 shows a comparison between expert-driven and data-driven BNs. The best results, both in training and testing, were obtained with one of the structures proposed by the CB experts. The results obtained are aligned with the BN literature, which suggests that causal discovery algorithms can be helpful for the researcher to consider alternative structures but do not usually produce optimal solutions. This highlights the importance of experts discussing their theories openly rather than relying on mechanical means. In this work, we also discuss the implications of using expert vs. data-based DAGs to choose the best counterfactual analysis and recommendations for policy.

References

[1] Smahel, D., *et al*. EU Kids Online 2020: Survey Results from 19 Countries. London School of Economics and Political Science (2020).

[2] Lobe, B., *et al*. How Children (10-18) Experienced Online Risks during the Covid-19 Lockdown: Spring 2020. Publications Office of the European Union (2021). DOI: 10.2760/066196

[3] Pearl, J., 1988. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann

	Naive-BN	Bayesian Search-BN	PC-BN	Greedy Thick Thinning-BN	1st Experts proposal-BN	2nd Experts proposal-BN
Mean Log Likelihood [Training sets]	-4103.8	-4034.9	-3987.8	-4003.9	-3990.8	**-3933.9**
Mean Log Likelihood [Testing sets]	-1089.1	-1073.1	-1064.9	-1067.4	-1071.8	**-1061.8**

Table 1. BN architecture comparison results. A k-fold cross validation was performed with k=5. The mean of the (log) likelihoods is shown. The higher the (log) likelihood, the better. The experts' second proposal obtains the best results in both training and testing sets.

Dung Tien Pham, Simon Wilson

Title: Exploiting Special Structures in Large Sparse Matrix for Efficient Matrix-Vector Multiplication

Abstract:

Computing the product of a large, sparse matrix with a vector is a fundamental operation in a wide variety of statistical tasks, such as solving large-scale linear systems that appear in many Bayesian algorithms. While it is straightforward to compute the product in time proportional to the non-zero entries of the matrix, in many cases, the scale of the problem is prohibitively large that accessing these values directly is impractical. In such cases, exploiting the properties beyond the matrices' mere sparsity is key for an efficient matrix-vector multiplication operation. Inspired by the idea of graph compression for adjacency-matrix multiplication [1], in this paper, we propose a method to speed up the computation for large, sparse, real-valued matrices that cannot be represented explicitly. Our experiments on a particular linear system that arises in the context of fitting a Markov random field model in the Cosmic Microwave Background (CMB) source separation problem [2] show significant improvements in terms of computational time and memory required compared to the original solution. More concretely, let $A \in R^{n \times n}$ be an n×n matrix and $x = (x_1, x_2, ..., x_n) \in R^n$ be a vector. The matrix-vector product $x \cdot A$ (or $A \cdot x^T$) is an important operation in many analysis tasks, including solving large, sparse linear systems. A direct computation of the product can only exploit the sparsity of the matrix by operating on all the non-zero values while ignoring other properties that might appear in the matrix construction. In this paper, we take advantage of the neighborhood and cluster properties of the mixing matrix in the CMB source separation problem to design an efficient matrix-vector procedure. We show that the operation can be executed much faster with less memory than the direct method, opening up applications of graph-based methods to large, sparse matrices that cannot be represented explicitly as adjacency matrices.

References

[1] Francisco, A.P., Gagie, T., Koppl, D. et al. (2022), Graph Compression for Adjacency-Matrix Multiplication. SN COMPUT. SCI. 3, 193.

[2] Soodhalter, K., Wilson, S., Pham, D.. (2022). Fast solution of Sylvester-structured systems for spatial source separation of the Cosmic Microwave Background. arXiv:2204.08057v2

Maria de los Remedios Sillero Denamiel, Pepa Ramírez-Cobo, Brani Vidakovic

Title: On Bayesian estimation of multinomial probabilities under incomplete experimental information

Abstract:

In this work, we discuss Bayesian estimation of multinomial probabilities associated with a finite alphabet **A**, under incomplete experimental information. Two types of sample information are considered: (i) number of letters needed to see a particular pattern for the first time, and (ii) the fact that for two fixed words one appeared before the other. An application of the method to a reliability problem will be shown.

Manel Rodríguez-Soto, Maite López-Sánchez, Juan Antonio Rodríguez-Aguilar

Title: Multi-objective reinforcement learning for designing ethical environments

Abstract:

The challenge of guaranteeing that autonomous agents act value-aligned (in alignment with human values) is becoming critical as agents increasingly populate our society. Hence, it is of great concern to design ethically aligned trustworthy AI capable of respecting human values in a wide range of emerging application domains (e.g. social assistive robotics, self-driving cars, conversational agents).

Here we present our approach to tackle the open problem of value alignment. We propose an approach to build an ethical environment that guarantees that an agent learns an ethically aligned behaviour while pursuing its individual objectives. Our contributions are founded in the framework of Multi-Objective Reinforcement Learning. Firstly, we characterise a family of Multi-Objective Markov Decision Processes (MOMDPs), the so-called ethical MOMDPs, for which we can formally guarantee the learning of ethical behaviours. Secondly, based on our characterisation we specify the process for building single-objective ethical environments that simplify the learning of the agent.

During our presentation we will enumerate and discuss the open research questions that could lead to more general results: can we generalise our results for more realistic environments with continuous state spaces and/or partial observability? What are the computational challenges faced by reinforcement learning agents despite the theoretical guarantees of our approach?

Finally, we briefly discuss the generalisation of our results for multi-agent systems, in which we have more challenging open research questions, such as finding dominant equilibria in stochastic games. This presentation is based on the following publications:

[1] Manel Rodriguez-Soto, Maite Lopez-Sanchez, Juan A. Rodriguez-Aguilar, “Multi-Objective Reinforcement Learning for Designing Ethical Environments”. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI 2021, Core A* conference).

[2] Manel Rodriguez-Soto, Marc Serramia, Maite Lopez-Sanchez, Juan A. Rodriguez-Aguilar, “Instilling moral value alignment by means of multi-objective reinforcement learning” at the Ethics and Information Technology journal. Ed. Springer (2020 IF JCR: 4.449, Q1 Cat. Philosophy (19/317)). Vol 24:9. pp 1-17. 24 Jan 2022.

[3] Manel Rodriguez-Soto, Juan A. Rodriguez-Aguilar, Maite Lopez-Sanchez, “Building Multi-Agent Environments with Theoretical Guarantees on the Learning of Ethical Policies”. Adaptive and Learning Agents Workshop (Workshop of AAMAS 2022).

Robert Richardson, Brian Hartman, Chris Groendyke, Zoe Gibbs, Michael Shull

Title: Spatio-Temporally Varying Stick Breaking Weights in Dependent Dirichlet Processes for Multivariate Spatio-Temporal Models

Abstract:

Dirichlet processes require a multiplicity of exchangeable observations to estimate parameters. This makes fully correlated data sets with no exchangeability difficult to analyze with Bayesian non-parametric methods. This is why very little Bayesian non-parametric work has been done on multivariate spatio-temporal models where data is correlated in multiple ways and exchangeability is difficult to justify. We propose a model where dependent Dirichlet processes are constructed on one source of dependence using stick breaking weights that vary based on the other two sources of dependence. In this way we can build flexible non-Gaussian, non-linear, and non-separable relationships into a multivariate spatio-temporal model. There still are a variety of restrictions we explore and compare with other state of the art methods and show scenarios using simulated data where it works both better and worse. We present several properties and develop methods for forecasting, kriging, and smoothing with this model. Using a proprietary data set containing mortality rates for every U.S. county for 20 years we apply our model to seek to understand mortality trends over time. The data set bins individuals by age and sex which we use to separate the data creating a multivariate spatio-temporal model. Using our methodology, we model mortality curves as a function of age using a dependent Dirichlet process where stick breaking weights are spatially and temporally varying

Alexandra M. Schmidt, Dirk Douwes-Schultz, Yannan Shen, David Buckeridge

Title: A three-state coupled Markov switching model for COVID-19 outbreaks across Quebec based on hospital admissions

Abstract:

Recurrent COVID-19 outbreaks have placed immense strain on the hospital system in Quebec. We develop a Bayesian three-state coupled Markov switching model to analyze COVID-19 outbreaks across Quebec based on admissions in the 30 largest hospitals. Within each catchment area we assume the existence of three states for the disease: absence, a new state meant to account for many zeroes in some of the smaller areas, endemic and outbreak. Then we assume the disease switches between the three states in each area through a series of coupled nonhomogeneous hidden Markov chains. Unlike previous approaches, the transition probabilities may depend on covariates and the occurrence of outbreaks in neighboring areas, to account for geographical outbreak spread. Additionally, to prevent rapid switching between endemic and outbreak periods we introduce clone states into the model which enforce minimum endemic and outbreak durations. We make some interesting findings such as that mobility in retail and recreation venues had a strong positive association with the development and persistence of new COVID-19 outbreaks in Quebec. Based on model comparison our contributions show promise in improving state estimation retrospectively and in real-time, especially when there are smaller areas and highly spatially synchronized outbreaks, and they offer new and interesting epidemiological interpretations.

Deborah Sulem, Vincent Rivoirard, Judith Rousseau

Title: Scalable Variational Bayes methods for Hawkes processes

Abstract:

Multivariate Hawkes processes are temporal point processes extensively applied to model event data with dependence on past occurrences and interaction phenomena, e.g., neuronal spike trains, online messages, and financial transactions. In the nonparametric setting, learning the temporal dependence structure of Hawkes processes is often a computationally expensive task, all the more with Bayesian estimation methods. In the generalised nonlinear Hawkes model, the posterior distribution is non- conjugate and doubly intractable, and existing Monte-Carlo Markov Chain methods are often slow and not scalable to high-dimensional processes in practice. Recently, efficient algorithms targeting a mean-field variational approximation of the posterior distribution have been proposed. In this work, we unify existing variational Bayes inference approaches under a general framework, that we theoretically analyse under easily verifiable conditions on the prior, the variational class, and the model. Then, in the context of the popular sigmoid Hawkes model, we design adaptive and sparsity-inducing mean-field variational methods. In particular, we propose a two-step algorithm based on a thresholding heuristic to select the connectivity graph parameter of the Hawkes model. Through an extensive set of numerical simulations, we demonstrate that our approach enjoys several benefits: it is computationally efficient, can reduce the dimensionality of the problem by selecting the graph parameter, and is able to adapt to the smoothness of the underlying parameter.

Szymon Urbas, Donagh Berry, Isobel Claire Gormley

Title: Predictions from spectral data using Bayesian partial least squares with a multiplicative gamma process prior

Abstract:

Partial least squares (PLS) regression is a popular method for making predictions in chemometrics. It is based on maximising the covariance between the predictor and response vectors through a latent-space decomposition. However, the method does not arise from the statistical inference of any particular probabilistic mode. Thus, it cannot account for parameter uncertainty and does not easily lend itself to modifications. Furthermore, the choice of the latent dimension, q, is usually left to the user's judgement, and common metrics guiding the choice can be sensitive to the training dataset. We introduce a Bayesian latent variable model which emulates the desirable properties of PLS whilst being a more user-friendly prediction tool. We illustrate three variants of the model, each allowing for a different sparsity structure in the regression of the final responses. The latent dimension is automatically chosen through the multiplicative gamma process prior which enforces shrinkage on the columns of the loading matrices in the model. Inference is carried out using a Gibbs sampling algorithm, and the predictions from the model incorporate the uncertainty in the parameters and the latent dimension. The final point predictions are at least as accurate as the industry-standard PLS method on various examples, performing particularly well in "small N, large P" scenarios. Additionally, the model provides prediction intervals with the correct coverage.

Alessandro Viani, Adam M. Johansen, Alberto Sorrentino

Title: Two Birds with one stone: hyper-parameter selection and averaging in Bayesian inverse problems by SMC samplers

Abstract:

The Bayesian approach to inverse problems is based on the approximation of the posterior distribution for a set of unknowns given some measured data. When we consider non-linear inverse problems an exact characterization of the posterior distribution is not always straightforward, therefore its approximation is required; Sequential Monte Carlo (SMC) samplers represent a good choice to approximate complex distributions. This class of Monte Carlo algorithms is based on the iterative approximation of a sequence of distributions that smoothly reaches the target one. One problem often encountered is to properly set the scalar hyper-parameter of the algorithm, that could strongly affect the approximation; for instance the noise standard deviation for additive Gaussian noise inverse problems.

In this work we show how to build an SMC sampler that provides an approximation of the marginal likelihood associated with this scalar hyper-parameter for free, i.e. at a negligible additional computational cost. The key point is to construct a sequence of auxiliary distributions such that each one is a conditional posterior distribution with respect to a different value of the hyper-parameter. By this trick, one can kill two birds with one stone by performing both selection of the hyper-parameter in Empirical Bayes (EB) approaches or/and hyper-parameter averaging by fixing some hyper-prior distribution in Fully Bayesian (FB) approaches.

We show numerical results for two distinct cases where the hyper-parameter of interest affects only the likelihood function: a toy example, where we use an SMC sampler to approximate the full posterior distribution; and a brain imaging example, where the interest is in estimating the activity within the brain as a spatiotemporal process, using a Rao-Blackwellized SMC sampler to approximate the posterior distribution for a subset of the unknowns.

The results show that the proposed method gives multiple benefits: (i) allows to perform prior sensitivity analysis on the hyper-parameter, (ii) permits to recycle particles at all the relevant iterations avoiding the discard of intermediate SMC samplers iterations, (iii) guarantees an approximation of both the unknown and the hyper-parameter; all these features at a negligible computational cost.

Audrone Virbickaite, Hoang Nguyen, Minh-Ngoc Tran

Title: Bayesian predictive distributions of oil returns using Mixed Data Sampling volatility models

Abstract:

With the financialization of the commodity markets, it is now straightforward to include commodities in one's investment portfolio. According to the Futures Trading Association, oil is the most traded commodity in the world by far: apart from being an excellent hedging instrument, it is also a major input in industrial production and transportation. Therefore, modeling and predicting the movements of the oil price, returns and volatility is of major importance. In this article we consider standard Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Stochastic Volatility (SV) and Generalized Autoregressive Score (GAS)-type models for the volatility. In particular, we are interested in the extensions that incorporate the leverage effect, excessive kurtosis of the unconditional return distribution, and the effect of exogenous variables. Generally, the volatility asymmetry and excessive kurtosis can be incorporated based on the model assumption, while the use of exogenous variables implies the extension of the available information set. In many cases, the exogenous predictors are some economic factors that are observed at a lower frequency (monthly, quarterly, yearly) than the crude oil prices (daily). In order to account for the frequency mismatch, we rely on a mixed data sampling method, or MIDAS in short. In the MIDAS setting, the effects of the lagged macro-finance variables are regularized by a polynomial weighting scheme to ensure a parsimonious model specification. Inference, model selection, and prediction are carried out using a novel Bayesian estimation strategy based on Density Tempered Sequential Monte Carlo (DTSMC) sampler. The method produces marginal log-likelihoods as a by-product, which allows for consistent model comparison even for non-nested models, as is our case. Also, it is well suited for non-linear non-Gaussian state space models, such as SV-type specifications. The Bayesian estimation approach is also important because the estimation uncertainty plays an important, if not crucial, role in the return density prediction as it punishes pointlessly complicated models and is often referred to as a fully automatic Occam's razor. We find that incorporating asymmetries, fat tails, and exogenous predictors leads to superior multi-step-ahead point volatility and entire density forecasts. The employed DTSMC estimation strategy provides a possibility to parallelize the estimation, leading to low computation costs even for non-linear non-Gaussian state space models.

Martina Danielova Zaharieva

Title: Infinite sparse factor stochastic volatility model

Abstract:

This paper proposes a sparse factor multivariate stochastic volatility model, in which the sparsity of the loading matrix is achieved by introducing the Indian buffet process, a Bayesian nonparametric prior defining a distribution over infinite binary matrices. The benefit of the infinite dimensional latent process is twofold. First, inducing sparsity prior reduces the dimensionality of the problem and second, the number of active factors is determined by the data itself and a priori set to infinity. Both, the diagonal elements of the covariance matrix of the idiosyncratic term, and the active factors follow univariate stochastic volatility processes. Each latent volatility is sampled independently and in parallel by means of a particle filtering and smoothing technique, based on a simulated likelihood. The model is applied to a cross section of five international stock market indices

Title: Infinite sparse factor stochastic volatility model

Abstract:

A significant part of the UK rail infrastructure is nearing 200 years in age whilst being built on high-plasticity soils that are prone to weathering and deterioration. Deterioration processes have been studied through computer simulation experiments of individual cuttings or embankments, but these are computationally-expensive and time-consuming, and therefore impractical to use directly for understanding the state of a rail network containing thousands of uniquely parameterised slopes. Instead we use surrogate statistical models, which can be used to approximate computationally-burdensome simulators, based on a relatively small number of simulator runs at well-selected input parameters.

The simulation models produce large amounts of output, but emulation strategies focus on (derived) outputs of direct practical interest. One such output is the simulated time to failure of a given slope. An interesting issue with this output is that the computer experiments are terminated after around 200 years of simulated time, and so the data set used for emulation contains right-censored observations. An MCMC-based Bayesian modelling framework can accommodate such censored data, and can also be adapted to other derived outputs, such as deterioration curves. These emulators can be embedded in a Bayesian hierarchical model of a (section of) rail network for characterisation of network state and evaluation of cost-effectiveness of potential intervention strategies.

This work is supported by ACHILLES (https://www.achilles-grant.org.uk/), a UKRI-EPSRC programme grant.

Reference

Svalova, A., Prangle, D., Wilkinson, D. J., Helm, P., Rouainia, M., Glendinning, S. (2021) Emulating computer experiments of infrastructure slope stability in the London-Bristol rail corridor using Gaussian processes and Bayesian inference, Data-Centric Engineering, 2:e12.

https://doi.org/10.1017/dce.2021.14