Bayesian Non Parametrics 12

Conference booklet

12th International Conference on Bayesian Nonparametrics

Oxford, United Kingdom, June 24-28, 2019

ISBA

erc European Research Council

Google

amazon

ims

The electronic version of this booklet can be found at:
http://www.stats.ox.ac.uk/bnp12/programme.html

Contents

About

12th International Conference on Bayesian Nonparametrics

Welcome to Oxford!

The Bayesian nonparametrics (BNP) conference is a bi-annual international meeting bringing together leading experts and talented young researchers working on applications and theory of nonparametric Bayesian statistics.

It is an official section meeting of the Bayesian nonparametrics section of the International Society for Bayesian Analysis (ISBA) and is co-sponsored by the Institute of Mathematical Statistics (IMS).

Attendees are expected to share our commitment to safeISBA, and to adhere to the ISBA code of conduct.

Scientific committee

Antonio Lijoi, Bocconi (Chair)
Catherine Forbes, Monash
Erik Sudderth, UC Irvine
Harry van Zanten, Amsterdam
Igor Prünster, Bocconi
Jaeyong Lee, Seoul
Li Ma, Duke
Maria de Iorio, NUS/UCL
Peter Müller, UT Austin
Richard Nickl, Cambridge
Steven MacEachern, Ohio State

Organizing committee

Fabrizio Leisen, Kent (co-chair)
François Caron, Oxford (co-chair)
Beverley Lane, Oxford (conf. officer)
Chris Holmes, Oxford
Jim Griffin, UCL
Maria de Iorio, NUS/UCL
Mark Steel, Warwick
Richard Nickl, Cambridge
Yee Whye Teh, Oxford/Deepmind
Zoubin Ghahramani, Cambridge/Uber

Sponsors

Timetable

KL: Keynote Lecture, IS: Invited Session, CS: Contributed Session.

The keynote talks, invited sessions and the BNP section meeting all take place in the lecture theatre 2 (L2) at the Mathematical Institute. The contributed sessions take place in L2 and L3.

Sunday 23 June

19:00–21:30	**Welcome Reception - Pitt Rivers Museum**

Monday 24 June

8:30–9:15	**Registration**		
9:15–9:30	**Opening**		
9:30–10:30	KL	**Aad van der Vaart**	Chair: R. Nickl
10:30–11:00	**Coffee break**		
11:00-12:30	IS	**BNP Computations I** Sinead Williamson John Paisley Mike Hughes	Chair: E. Sudderth
12:30–14:00	**Lunch**		
14:00-15:30	IS	**High dimensions and sparsity** Yanxun Xu Kshitij Khare Roberto Casarin	Chair: C. Holmes
15:30–16:00	**Coffee break**		
16:00–17:30	CS	**Asymptotics I** Chair: S. van der Pas – Room: L2 G. Di Benedetto S. Ghosal C. Li S. Wang	**Random measures & mixtures** Chair: A. Ongaro – Room: L3 F. Ayed R. Giordano D. Kowal J. Lee
19:30–21:30	**Poster session I**		

Tuesday 25 June

09:00–10:30	CS	**Computational Methods** Chair: V. Rao – Room: L2 R. Corradin A. Nishimura R. Ryder G. Zanella	**Asymptotics II** Chair: S. Agapiou – Room: L3 M. Kasprzak H. Kekkonen B. Ning
10:30–11:00	**Coffee break**		
11:00-12:30	IS	**Asymptotics & credible sets** Stephanie van der Pas Botond Szabo Ismaël Castillo	Chair: H. van Zanten
12:30–14:00	**Lunch**		
14:00-15:30	IS	**Foundational aspects** Jared Murray Chris Holmes Natalia Bochkina	Chair: J. Griffin
15:30–16:00	**Coffee break**		
16:00–17:30	CS	**Dependence structures in BNP** Chair: M. Ruggiero – Room: L2 T. Campbell J. Palacios F. Quintana T. Rigon	**Survival and healthcare data** Chair: F. Camerlenghi – Room: L3 F. Donaghy S. Filippi Y. Luo A. Riva-Palacio
19:30–21:30	**Poster session II**		

Wednesday 26 June

9:00–10:00	KL	**Tamara Broderick**	Chair: P. Müller
10:00–10:30	**Coffee break**		
10:30-12:30	IS	**Inverse problems** Andrew Stuart Judith Rousseau Kolyan Ray Sergios Agapiou	Chair: R. Nickl
12:30–14:00	**Lunch**		
14:00-15:30	IS	**Biostatistics** Yang Ni George Karabatsos Veera Baladandayuthapani	Chair: P. Müller
15:30–16:00	**Coffee break**		
16:00–17:30	CS	**Prediction/Random partitions I** Chair: Y. Xu – Room: L2 C. Balocchi F. Camerlenghi L. Elliott R. Warr	**Time dependent modeling** Chair: R. Casarin – Room: L3 L. Cappello G. Kon Kam King R. Meyer X. Miscouridou
17:45–19:00	**Bayesian Nonparametrics Section Meeting**		
19:15–21:00	**Junior ISBA reception - Dept of Statistics**		

Thursday 27 June

09:00–10:30	CS	**BNP modeling & robustness** Chair: S. Wade – Room: L2 A. Kottas J. Miller L.E. Nieto-Barajas S. Tokdar	**Prediction/Random partitions II** Chair: T. Savitsky – Room: L3 S. Deshpande A. Diana S. Favaro M. Zhang
10:30–11:00	**Coffee break**		
11:00-12:30	IS	**Random measures & predictive inference** Mike West Dario Spanó Alejandro Jara	Chair: J. Lee
12:30–14:00	**Lunch**		
14:00-15:30	IS	**BNP for discrete structures** Dan Roy Peter Orbanz Diana Cai	Chair: S. MacEachern
15:30–16:00	**Coffee break**		
16:00–17:30	CS	**Asymptotics III** Chair: K. Ray – Room: L2 E. Dolera D. Li A. Norets Z. Naulet	**BNP Applications** Chair: Y. Ni – Room: L3 A. Avalos-Pacheco M. Shiffman A. Sottosanti A. Verma
20:00–23:00	**Conference dinner - Lady Margaret Hall**		

Friday 28 June

9:00-10:30	IS	**BNP Computations II** Terrance Savitsky Vinayak Rao Maria Kalli	Chair: C. Forbes
10:30–11:00	**Coffee break**		
11:00–12:00	KL	**Long Nguyen**	Chair: I. Prünster
12:00–12:15	**Closing**		

List of Talks

Monday 24 June

Semiparametric Bayesian estimation, with or without bias Mo KL
Aad Van der Vaart, Leiden University, Netherlands

Distributed inference for Bayesian Nonparametrics Mo IS
Sinead Williamson, University of Texas at Austin/Amazon, USA

Population Random Measure Embedding Mo IS
John Paisley, Columbia University, USA

Scalable and Reliable Variational Inference for Dirichlet Process Clustering with Sparse Assignments Mo IS
Michael C. Hughes, Tufts University, Medford, MA, USA

Bayesian Estimation of Sparse Spiked Covariance Matrices in High Dimensions Mo IS
Yanxun Xu, Johns Hopkins University, United States

A Bayesian approach for joint estimation of multiple networks Mo IS
Kshitij Khare, University of Florida, USA

A Bayesian Approach for Inference on Probabilistic Surveys Mo IS
Roberto Casarin, University Ca' Foscari of Venice, Italy

Posterior contraction rates for Bayesian Functional Regression Mo CS
Giuseppe Di Benedetto, University of Oxford, United Kingdom

Coverage of credible intervals for monotone regression Mo CS
Subhashis Ghoshal, North Carolina State University, USA

Posterior Consistency of Tail Index for Bayesian Kernel Mixture Models Mo CS
Cheng Li, National University of Singapore, Singapore

Convergence rate results for PDE-constrained statistical inverse problems Mo CS
Sven Wang, University of Cambridge, United Kingdom

Double power law completely random measures with application to clustering Mo CS
Fadhel Ayed, University of Oxford, UK

Evaluating Sensitivity to the Stick Breaking Prior in Bayesian Nonparametrics Mo CS
Ryan Giordano, UC Berkeley, USA

Dynamic Shrinkage Processes Mo CS
Daniel Kowal, Rice University, USA

Arrival time augmentation for series representations of completely random measures Mo CS
Juho Lee, AITRICS, Republic of Korea

Tuesday 25 June

Importance conditional sampler for nonparametric mixture models Tu CS
Riccardo Corradin, University of Milano-Bicocca, Italy

Scalable Bayesian sparse survival analysis and generalized linear models via curvature-adaptive Hamiltonian Monte Carlo for high-dimensional log-concave distributions Tu CS
Akihiko Nishimura, University of California - Los Angeles, USA

A Bayesian non-parametric methodology for inferring grammar complexity Tu CS
Robin Ryder, Université Paris-Dauphine, France

Informed proposals for local MCMC in discrete spaces Tu CS
Giacomo Zanella, Bocconi University, Italy

Probabilistic methods for proving error bounds in GP posterior approximation Tu CS
Mikołaj Kasprzak, University of Luxembourg, Luxembourg

Bernstein–von Mises Theorems for linear inverse problems Tu CS
Hanne Kekkonen, University of Cambridge, UK

Bayesian high-dimensional analyses for a multivariate linear regression model and a sparse spiked-covariance model. Tu CS
Bo Ning, Yale University, USA

Uncertainty quantification for survival analysis Tu IS
Stéphanie van der Pas, Leiden University, The Netherlands

On distributed Bayesian computation Tu IS
Botond Szabo, Leiden University, Netherlands

On multiscale properties of some Bayesian tree methods Tu IS
Ismael Castillo, Sorbonne University, Paris, France

Recent developments in model specification, regularization, and summarization for nonparametric Bayesian models of heterogeneous treatment effects. Tu IS
Jared Murray, UT Austin, USA

Bayesian nonparametric learning through randomized loss functions and posterior bootstraps Tu IS
Christopher Holmes, University of Oxford, UK

Testing for model misspecification Tu IS
Natalia Bochkina, University of Edinburgh and the Alan Turing Institute, UK

Local Exchangeability Tu CS
Trevor Campbell, UBC, Canada

Statistically efficient coalescent inference Tu CS
Julia Palacios, Stanford University, USA

Nonparametric Bayesian Modeling of Correlation Functions for Global Data Tu CS
Fernando Quintana, Pontificia Universidad Católica de Chile, Chile

Bayesian inference for finite-dimensional discrete priors Tu CS
Tommaso Rigon, Bocconi University, Italy

Dependent neutral-to-the-right priors for software reliability data Tu CS
Fearghal Donaghy, Trinity College, Dublin, Ireland

Bayesian Non Parametric approaches for Survival analysis Tu CS
Sarah Filippi, Imperial College London, UK

Clustering via Dirichlet Process Mixture Models for Trajectories with Fixed Effects Continuous-Time Hidden Markov Models Tu CS
Yu Luo, McGill University, Canada

Survival analysis regression Tu CS
Alan Riva-Palacio, University of Kent, United Kingdom

Wednesday 26 June

Title TBA We KL
Tamara Broderick, MIT, USA

Posterior Consistency In Large Graph Limits of Learning Algorithms We IS
Andrew Stuart, California Institute of Technology, USA

Mixtures and inverse or semi-parametric problems We IS
Judith Rousseau, University of Oxford, UK

Nonparametric Bayesian drift estimation for multidimensional diffusions We IS
Kolyan Ray, Kings College London, United Kingdom

Rates of contraction: some non-Gaussian priors and some nonlinear inverse problems We IS
Sergios Agapiou, University of Cyprus, Cyprus

Double Feature Allocation for Phenotyping with Electronic Health Records Data We IS
Yang Ni, Texas A&M University, USA

Fast Search for General Bayesian Nonparametric Mixture Models We IS
George Karabatsos, University of Illinois-Chicago, U.S.A.

Bayesian Nonparameteric Models for Richly Structured Data in Biomedicine We IS
Veera Baladandayuthapani, University of Michigan, USA

Clustering data at multiple resolutions We CS
Cecilia Balocchi, University of Pennsylvania, US

Bayesian prediction in feature models We CS
Federico Camerlenghi, University of Milano - Bicocca and Collegio Carlo Alberto, Italy

The Multidimensional Partitioning Tree Process We CS
Lloyd T. Elliott, Simon Fraser University, Canada

The Attraction Indian Buffet Distribution We CS
Richard Warr, Brigham Young University, USA

Bayesian nonparametric inference of population trajectories via Tajima heterochronous n-coalescent We CS
Lorenzo Cappello, Stanford University, USA

Computable inference for a class of non-linear state-space models We CS
Guillaume Kon Kam King, University of Turin and Collegio Carlo Alberto, Italy

Beyond Whittle: Nonparametric Correction of a Parametric Likelihood with a Focus on Bayesian Time Series Analysis We CS
Renate Meyer, University of Auckland, New Zealand

Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data We CS
Xenia Miscouridou, University of Oxford, UK

Thursday 27 June

Approximate multiple shrinkage for clustered regressions Th CS
Sameer Deshpande, CSAIL, MIT, USA

A Hierarchical Dependent Dirichlet process prior for modelling bird migration patterns in the UK Th CS
Alex Diana, University of Kent, UK

A Bayesian nonparametric approach to disclosure risk assessment Th CS
Stefano Favaro, University of Torino and Collegio Carlo Alberto, Italy

A New Class of Time Dependent Latent Factor Models with Applications Th CS
Michael Zhang, Princeton University, USA

Bayesian Quantile Mixture Regression Th CS
Athanasios Kottas, University of California, Santa Cruz, USA

Flexible perturbation models for robustness to misspecification Th CS
Jeffrey Miller, Harvard T.H. Chan School of Public Health, USA

Projected Polya tree Th CS
Luis Nieto-Barajas, ITAM, Mexico

Joint Quantile Regression under Dependency Th CS
Surya Tokdar, Duke University, USA

Bayesian Predictive Synthesis Th IS
Mike West, Duke University, USA

Dualities and genealogies for time-dependent nonparametric models Th IS
Dario Spanó, University of Warwick, United Kingdom

Models for related probability measures on nonstandard domains Th IS
Alejandro Jara, Pontificia Universidad Catolica de Chile, Chile

Nonstandard Nonparametrics Th IS
Daniel Roy, University of Toronto and Vector Institute, Ontario, Canada

Bayesian aspects of preferential attachment networks Th IS
Peter Orbanz, Columbia University, USA

A Bayesian Nonparametric View on Count-Min Sketch Th IS
Diana Cai, Princeton University, USA

A Berry-Esseen theorem for Pitman's α-diversity Th CS
Emanuele Dolera, University of Pavia, Italy

Density Estimation with Mixture of Spherelets Th CS
Didong Li, Duke University, USA

Adaptive Bayesian Estimation of Mixed Discrete-Continuous Distributions under Smoothness and Sparsity Th CS
Andriy Norets, Brown University, USA

Asymptotic analysis of the posterior distribution in the Caron and Fox model Th CS
Zacharie Naulet, University of Toronto, Canada

Cross-study Bayesian Factor Regression Analysis in High-dimensional Biological Data Th CS
Alejandra Avalos-Pacheco, Princeton University, USA

Reconstructing probabilistic trees of cellular differentiation from single-cell RNA-seq data Th CS
Miriam Shiffman, MIT & Broad Institute, USA

Astronomical source detection and background separation via hierarchical Bayesian nonparametric mixtures Th CS
Andrea Sottosanti, University of Padova, Italy

Robust, Nonparametric Manifold Learning for Single Cell RNA Sequencing Th CS
Archit Verma, Princeton University, USA

Friday 28 June

Bayesian Pseudo Posterior Synthesis for Data Privacy Protection Fr IS
Terrance Savitsky, U.S. Bureau of Labor Statistics, USA

Nonparametric mixture modeling on constrained spaces Fr IS
Vinayak Rao , Purdue University, USA

Bayesian nonparametric methods for analysing macroeconomic time series Fr IS
Maria Kalli, University of Kent, UK

Posterior contraction of parameters and interpretability in Bayesian mixture modeling Fr KL
Long Nguyen, University of Michigan, USA

List of Posters

Monday 24 June

Geometric Sensitivity Measures for Bayesian Nonparametric Density Estimation Models Mo PS
Abhijoy Saha, The Ohio State University, USA

Cross-study Bayesian Factor Regression Analysis in High-dimensional Biological Data Mo PS
Alejandra Avalos-Pacheco, Harvard Medical School, USA

Bayesian nonparametric modeling for large spatio-temporal data: an application to mobile networks Mo PS
Alessandra Guglielmi, Politecnico di Milano, Italy

Population Random Measure Embedding: A Genetic Method for Discrete Random Measures, via Distributional Symmetry Mo PS
Aonan Zhang, Columbia University, United States

Nonparametric Bayesian Functional Regression with application to shot put data Mo PS
Alessandro Lanteri, University of Turin, Italy

Optimize, Learn, Sample Mo PS
Alfredo Garbuno-Inigo, Caltech, US

Investigating a Bayesian semi-parametric model for the study of synergistic interaction effects in in-vitro drug combination experiments Mo PS
Andrea Cremaschi, Universitetet i Oslo, Norway

On posterior contraction of parameters and interpretability in Bayesian mixture modeling Mo PS
Aritra Guha, University of Michigan, USA

Closed Form Bayesian Filtering for Multivariate Binary Time Series Mo PS
Augusto Fasano, Bocconi University, Milan, Italy

A nonparametric Bayesian prediction approach for modelling small data Mo PS
Azizur Rahman, Charles Sturt University, Australia

Poisson Process Radial Basis Bayesian Neural Networks Mo PS
Beau Coker, Harvard University, USA

Models for Networks with Core-Periphery Structure Mo PS
Cian Naik, University of Oxford, United Kingdom

Dependent Random Measures Indexed by a Functional Covariate Mo PS
Emmanuel Bernieri, University of Edinburgh, Scotland

Generalised Polya urn for a class of dependent Dirichlet Processes Mo PS
Filippo Ascolani, University of Torino and Collegio Carlo Alberto, Italy

Sparse Spatial Random Graphs Mo PS
Francesca Panero, University of Oxford, United Kingdom

Modeling Human Microbiome Data via Latent Nested Nonparametric Priors Mo PS
Francesco Denti, University of Milan-Bicocca, Italy and Università della Svizzera italiana, Switzerland

Joint Species Distribution Modelling: Dimension reduction using Bayesian nonparametric priors Mo PS
Giovanni Poggiato, Inria(Grenoble INP), France,

Hybrid BNP priors for clustering Mo PS
Giovanni Rebaudo, Bocconi University, Italy

Non-echangeable prior for feature models, a generalization of the three-parameter Indian Buffet Process Mo PS
Giuseppe Di Benedetto, University of Oxford, United Kingdom

Some developments of the generalized species sampling sequences Mo PS
Hristo Sariev, Università Commerciale Luigi Bocconi, Italy

Nonparametric temporal sequence alignment Mo PS
Ieva Kazlauskaite, University of Bath, UK

Monotonic random processes Mo PS
Ivan Ustyuzhaninov, University of Tübingen, Germany

Improving Inference for the Non-Stationary Contextual Bandit via Iterative Moment-Matching Algorithms Mo PS
Jack McKenzie, The University of Manchester, UK

Posterior contraction of non-parametric Bayesian inference on non-homogeneous Poisson processes Mo PS
James Grant, Lancaster University, UK

Detection of common-variance subspace and its application to classification Mo PS
Jiae Kim, The Ohio State University, USA

A Generalization of Hierarchical Exchangeability on Trees to Directed Acyclic Graphs Mo PS
Jiho Lee, KAIST, South Korea

Bayesian semi-parametric density estimation for nonregular models Mo PS
Johan Van Der Molen Moris, University of Edinburgh, UK

Bayesian non-parametric methods for malware classification Mo PS
Jose Antonio Perusquia Cortes, University of Kent, UK

Bayesian nonparametric priors for hidden Markov random fields: Application to image segmentation Mo PS
Julyan Arbel, Inria Grenoble Rhône-Alpes, France

Bayesian Nonparametric Unsupervised Concept Drift Detection for Data Stream Mining Mo PS
Junyu Xuan, University of Technology Sydney, Australia

Post-Processed Posteriors for Band-Structured Covariances Mo PS
Kwangmin Lee, Seoul National University, South Korea

Consistent reconstruction of electrical impedance tomography images Mo PS
Kweku Abraham, University of Cambridge, United Kingdom

Tuesday 25 June

Policymaking and Statistical Estimates: A Bayesian Decision-Analytic Model for a Binary Outcome Tu PS
Akisato Suzuki, University College Dublin, Ireland

Multiple kernel learning with structured Gaussian processes: an application to drug interaction prediction Tu PS
Leiv Rønneberg, University of Oslo, Norway

A Bayesian nonparametric testing procedure for paired samples Tu PS
Luis Gutierrez, Pontificia Universidad Catolica de Chile, Chile

Bayesian neural network priors at the level of units Tu PS
Mariia Vladimirova, Inria Grenoble Rhone-Alpes, France

Bayesian Nonparametric Vector Auto Regressive models via a logit stick-breaking prior Tu PS
Mario Beraha, Politecnico di Milano and Universita degli Studi di Bologna, Italy

Measuring the sensitivity to prior specification for time-to-event data through the Wasserstein distance Tu PS
Marta Catalano, Bocconi University, Italy

Turing.jl: Probabilistic programming with discrete random probability measures. Tu PS
Martin Trapp, Graz University of Technology, Austria

Bayesian nonparametric graphical models for time-varying parameters VAR Tu PS
Matteo Iacopini, Cao Foscari University of Venice & Scuola Normale Superiore of Pisa, Italy

A Data-driven Posterior for Uncertainty Exploration in Bayesian Variable Selection via Hopfield Networks Tu PS
Matteo Vestrucci, University of Texas at Austin, USA

Fast Bayesian Hazard Regression Under General Censoring via Monotone P-Splines Tu PS
Matthias Kaeding, RWI - Leibniz Institute for Economic Research, Germany

Updating Variational Bayes for Online Inference of a Dirichlet Process Mixture Tu PS
Nathaniel Tomasetti, Monash University, Australia

Bayesian Nonparametrics for Circular Statistics and Density Estimation on Compact Metric Spaces Tu PS
Olivier Binette, University of Quebec at Montreal, Canada

Efficient Bayesian shape–restricted function estimation with constrained Gaussian process prior Tu PS
Pallavi Ray, Texas A&M University, USA

Whittle approximation for locally stationary time series Tu PS
Patricio Maturana-Russel, University of Auckland, New Zealand

Modeling data in simplexes Tu PS
Rayleigh Lei, University of Michigan, USA

Generalized modes in Bayesian inverse problems Tu PS
Remo Kretschmann, Universität Duisburg-Essen, Germany

Hierarchical Species Sampling Models Tu PS
Roberto Casarin, University Ca' Foscari of Venice, Italy

Bayesian Non-Parametric Inference for Stochastic Epidemic Models Tu PS
Rowland Seymour, University of Nottingham, UK

Evaluating Sensitivity to the Stick Breaking Prior in Bayesian Nonparametrics Tu PS
Ryan Giordano, UC Berkeley, USA

Bayesian Varying Coefficients Models Based on Gaussian Process Priors Tu PS
Sanvesh Srivastava, The University of Iowa, USA

Bayesian Hierarchical Modeling on Covariance Valued Data Tu PS
Satwik Acharyya, Texas A&M University, USA

Bayesian Quadrature with BART for Bayesian Survey Design Tu PS
Seth Flaxman, Imperial College London, UK

Variable Selection Consistency of Gaussian Process Regression Tu PS
Sheng Jiang, Duke University, USA

A Bayesian Estimation of Panel Stochastic Frontier Models with Determinants of Persistent and Transient Inefficiencies in Both Location and Scale Parameters Tu PS
Sheng-Kai Chang, National Taiwan University, Taiwan

Towards a Bayesian nonparametric genome-wide association study Tu PS
Shijia Wang, Simon Fraser University, Canada

Bayesian cumulative shrinkage for infinite factorizations Tu PS
Sirio Legramanti, Bocconi University, Italy

Bernstein von Mises theorems for general stick-breaking process priors. Tu PS
Stefan Franssen, Leiden University, The Netherlands

Bayesian inference for multivariate extremes Tu PS
Stefano Rizzelli, EPFL, Switzerland

Challenges and proposals for Dirichlet process mixture models with Gaussian kernels Tu PS
Wei Jing, University of St Andrews, UK

Variational Nonparametric Discriminant Analysis Tu PS
Weichang Yu, University of Sydney, Australia

A Bayesian Nonparametric Spiked Process Prior for Dynamic Model Selection Tu PS
Weixuan Zhu, Xiamen University, China

EP-IS: Combining expectation propagation and importance sampling for Bayesian nonlinear inverse problems Tu PS
Willem van den Boom, National University of Singapore, Singapore

Human Behaviour Analysis Through Probabilistic Modelling of GPS Data Tu PS
Yazan Qarout, Aston University, UK

Monday 24 June

Semiparametric Bayesian estimation, with or without bias Mo KL

Aad Van der Vaart, Leiden University, Netherlands
Co-author(s): Kolyan Ray

Semiparametric models typically have multiple parameters, of which at least one is infinite dimensional. These parameters can be supplied with priors in multiple ways. Prior (in)dependence is an issue that has led to philosophical debate. Oversmoothing of a functional parameter can lead to bias in the posterior of another parameter. We review some of these issues, and relate them in particular to the matter of so-called double-robust estimation, which refers to models with two infinite-dimensional parameters. Mathematically one may be after a Bernstein-von Mises theorem for a one-dimensional parameter of interest, which roughly guarantees that the posterior is unbiased and maximally concentrated. We discuss how choices of prior for nuisance parameters may influence the validity of this approximation. We ask the open question whether a bias-variance trade-off, as possible by ad-hoc methods of estimation, is also possible within a principled Bayesian setup.

Distributed inference for Bayesian Nonparametrics Mo IS

Sinead Williamson, University of Texas at Austin/Amazon, USA
Co-author(s): M. Zhang and A. Dubey

Inference in Bayesian nonparametric models is often challenging, and can scale poorly in high dimensions. Distributed algorithms allow us to share the computational load across multiple machines, but can introduce either excessive communication costs or unwanted approximations. In this talk I discuss general-purpose approaches for distributing inference in Bayesian nonparametric models, including Dirichlet process mixture models and Indian buffet processes.

Population Random Measure Embedding Mo IS

John Paisley, Columbia University, USA
Co-author(s): A. Aonan Zhang

Many hidden structures underlying high dimensional data can be compactly expressed by a discrete random measure $\xi_n = \sum_{k\in[K]} Z_{nk}\delta_{\theta_k}$, where $(\theta_k)_{k\in[K]} \subset \Theta$ is a collection of hidden atoms shared across observations (indexed by n). Previous Bayesian nonparametric methods focus on embedding ξ_n onto alternative spaces to resolve complex atom correlations. However, these methods can be rigid and hard to learn in practice. In this paper, we temporarily ignore the atom space Θ and embed population random measures $(\xi_n)_{n\in\mathbb{N}}$ altogether as ξ' onto an infinite strip $[0,1]\times\mathbb{R}_+$, where the order of atoms is *removed* by assuming separate exchangeability. Through a "de Finetti type" result, we can represent ξ' as a coupling of a 2d Poisson process and exchangeable random functions $(f_n)_{n\in\mathbb{N}}$, where each f_n is an object-specific atom sampling function. In this way, we transform the problem from learning complex correlations with discrete random measures into learning complex functions that can be learned with deep neural networks. In practice, we introduce an efficient amortized variational

inference algorithm to learn f_n without pain; i.e., no local gradient steps are required during stochastic inference.

Scalable and Reliable Variational Inference for Dirichlet Process Clustering with Sparse Assignments Mo IS

Michael C. Hughes, Tufts University, Medford, MA, USA
Co-author(s): Joint work with E. Sudderth

We develop new optimization algorithms for training Bayesian nonparametric clustering models that improve scalability to large datasets while reliably escaping local optima. Our innovations apply to mixture models, topic models, and hidden Markov models that use the Dirichlet process prior and its hierarchical extensions. Such models are typically trained via variational optimization methods with a-priori finite truncation of the state space or via Markov chain Monte Carlo methods. Both approaches are vulnerable to slow mixing and local optima. We show that BNP clustering models admit flexible variational objective functions that can sensibly compare models with different numbers of instantiated clusters and thus coherently explore adding and removing clusters during optimization. We interleave standard coordinate ascent steps with data-driven proposal moves that can add useful new clusters or remove redundant clusters, thus escaping local optima.

We further introduce an incremental algorithm - memorized variational inference – that can exactly optimize our objective function on large datasets while processing only small batches at each step. Our approach uses cached or memoized sufficient statistics to make exact decisions for proposal acceptance or rejection based on the entire dataset. This memoized approach has the same runtime cost as previous stochastic methods but allows principled acceptance decisions for cluster proposals and avoids learning rates entirely. We demonstrate promising proposal moves for adapting the number of clusters during memoized training on millions of news articles, hundreds of motion capture sequences, and the human genome. Finally, we show in-progress work introducing an additional sparsity constraint to the variational optimization problem for local cluster assignments in mixtures and topics. This sparse assignment leads to substantial speed gains during training and evaluation without sacrificing model quality.

Bayesian Estimation of Sparse Spiked Covariance Matrices in High Dimensions Mo IS

Yanxun Xu, Johns Hopkins University, United States
Co-author(s): A. Fangzheng Xie and B. Carey Priebe and C. Joshua Cape

We propose a Bayesian methodology for estimating spiked covariance matrices with jointly sparse structure in high dimensions. The spiked covariance matrix is reparametrized in terms of the latent factor model, where the loading matrix is equipped with a novel matrix spike-and-slab LASSO prior, which is a continuous shrinkage prior for modeling jointly sparse matrices. We establish the rate-optimal posterior contraction for the covariance matrix with respect to the operator norm as well as that for the principal subspace with respect to the projection operator norm loss. We also study the posterior contraction rate of the principal subspace with respect to the two-to-infinity norm loss, a novel loss function measuring the distance between subspaces that is able to capture element-wise eigenvector perturbations. We show that the posterior contraction rate with respect to the two-to-infinity norm loss is tighter than that with respect to the routinely used projection operator norm loss under certain low-rank and bounded coherence conditions. In addition, a point estimator for the principal subspace is proposed with the rate-optimal risk bound with respect to the projection operator norm loss. These results are based on a collection of concentration and large deviation inequalities for the matrix spike-and-slab LASSO prior. The numerical performance of the proposed methodology is assessed through synthetic examples and the analysis of a real-world face data example.

A Bayesian approach for joint estimation of multiple networks Mo IS

Kshitij Khare, University of Florida, USA
Co-author(s): George Michailidis and Peyman Jalali

In this paper, we develop a novel Bayesian approach for joint estimation of multiple graphical models. This problem arises in many applications, such as understanding co-expression networks from high-dimensional Omics data obtained from different biological conditions or disease subtypes. We pursue a pseudo-likelihood based approach which provides robustness and computational efficiency. We illustrate the efficacy of our approach using simulated and real datasets.

A Bayesian Approach for Inference on Probabilistic Surveys Mo IS

Roberto Casarin, University Ca' Foscari of Venice, Italy
Co-author(s): F. Bassetti and M. Del Negro

We propose a non-parametric Bayesian approach to the estimation of forecast densities in probabilistic surveys. As a kernel, we proposed an adjusted Dirichlet density which allows for values of the random probability vector on the boundary of the simplex. The non-parametric dimension allows for a flexibility that the conventional approach to probabilistic surveys since it accommodate multi-modality, and forces the researcher to arbitrarily close the open bins, whereas the adjusted Dirichlet kernel allows for dealing with the zero-valued bin probabilities. We provide an efficient Gibbs sampler for posterior approximation and some results on the posterior consistency of the proposed inference. We use our model to study the evolution of the subjective forecast distribution for the U.S. Survey of Professional Forecasters over the past forty years. We show that the variance of aggregate forecast distribution fell substantially from the eighties to the nineties and fell again after the Fed announced its long-term inflation goal. We also show that disagreement (heterogeneity in the mean forecasts) plays a minor role, but that heterogeneity in uncertainty is very large.

Posterior contraction rates for Bayesian Functional Regression Mo CS

Giuseppe Di Benedetto, University of Oxford, United Kingdom
Co-author(s): Judith Rousseau and Angelina Roche

Functional data analysis (FDA) has received a growing interest in the last few decades thanks to the applicability of its models to different areas where data are collected over fine grids. Here we will consider a Bayesian approach to FDA and study two regression problems: functional linear regression with scalar response and functional covariate, and the functional single-index model with scalar response. In the former model we investigate the asymptotic properties of the posterior distributions for different priors on the slope parameter, providing contraction rates with respect to the prediction risk and the L^2 norm. Using these results for the linear model, and placing mixture prior on the link function, we find contraction rates for the single-index model with respect to the empirical distance.

Coverage of credible intervals for monotone regression Mo CS

Subhashis Ghoshal, North Carolina State University, USA
Co-author(s): Moumita Chakraborty

Shape restrictions such as monotonicity often naturally arise in applications. In this talk we consider a Bayesian approach to monotone nonparametric regression. We assign a prior through piecewise constant functions and impose a conjugate normal prior on the coefficient. Since the resulting functions

need not be monotone, we project samples from the posterior on the allowed parameter space to construct a "projection posterior". We obtain the limit of the coverage of a Bayesian credible interval. We observe an interesting phenomenon that the frequentist coverage is higher than the targeted credibility, the exact opposite of a phenomenon observed by Cox for smooth regression. We describe a recalibration strategy to modify the credible interval to meet the correct level of coverage.

Posterior Consistency of Tail Index for Bayesian Kernel Mixture Models Mo CS

Cheng Li, National University of Singapore, Singapore
Co-author(s): A. Lizhen Lin (University of Notre Dame) and B. David B. Dunson (Duke University)

Asymptotic theory of tail index estimation has been studied extensively in the frequentist literature on extreme values, but rarely in the Bayesian context. We investigate whether popular Bayesian kernel mixture models are able to support heavy tailed distributions and consistently estimate the tail index. We show that posterior inconsistency in tail index is surprisingly common for both parametric and nonparametric mixture models. We then present a set of sufficient conditions under which posterior consistency in tail index can be achieved, and verify these conditions for Pareto mixture models under general mixing priors.

Convergence rate results for PDE-constrained statistical inverse problems Mo CS

Sven Wang, University of Cambridge, United Kingdom
Co-author(s): R. Nickl, S. van de Geer

We consider some non-parametric statistical models where the unknown parameter f is an unknown coefficient function in the differential operator of a second order elliptic PDE, and the unique solution u_f to the boundary value problem corresponding to f is observed, under additive Gaussian white noise. Concrete examples are 1) a steady state divergence form heat equation where the heat conductivity is unknown and 2) the time-homogeneous Schrödinger equation with unknown attenuation potential. In both cases, f is modelled in Sobolev smoothness classes with a non-negativity constraint. Moreover, both maps $f \mapsto u_f$ are non-linear and induce non-linearly constrained sets of PDE solutions.

We prove convergence rates for Tikhonov-type penalised least squares estimators $\hat{f}$ for f and the associated plug-in estimators $u_{\hat{f}}$ for u_f. The penalty functionals are of squared Sobolev-norm type and thus the estimators can be interpreted as a Bayesian 'MAP'-estimators corresponding to some Gaussian process prior. We show that this approach solves the PDE-constrained regression problems in a minimax-optimal way, in prediction loss. The bounds are derived from a general convergence rate result for non-linear inverse problems whose forward map satisfies a mild modulus of continuity condition, a result which is based on arguments from M-estimation. The result, which is of independent interest, is applicable also to linear inverse problems, including for example the Radon transform.

Double power law completely random measures with application to clustering Mo CS

Fadhel Ayed, University of Oxford, UK
Co-author(s): Juho Lee and François Caron

Bayesian nonparametric approaches, in particular the Pitman-Yor process and the associated two-parameter Chinese Restaurant process, have been successfully used in applications where the data exhibit a power-law behavior. Examples include natural language processing, natural images or networks. There is also growing empirical evidence that some datasets exhibit a two-regime power-law behavior: one regime for small frequencies, and a second regime, with a different exponent, for high frequencies.

In this paper, we introduce a class of completely random measures which are doubly regularly-varying. Contrary to the Pitman-Yor process, we show that when completely random measures in this class are normalized to obtain random probability measures and associated random partitions, such partitions exhibit a double power-law behavior. We discuss in particular three models within this class: the beta prime process (Broderick et al. (2015, 2018), a novel process called generalized BFRY process, and a mixture construction. We derive efficient Markov chain Monte Carlo algorithms to estimate the parameters of these models. Finally, we show that the proposed models provide a better fit than the Pitman-Yor process on various datasets.

Evaluating Sensitivity to the Stick Breaking Prior in Bayesian Nonparametrics Mo CS

Ryan Giordano, UC Berkeley, USA
Co-author(s): Runjing Liu, UC Berkeley, USA, Michael I. Jordan, UC Berkeley, USA, Tamara Broderick, MIT, USA

A central question in many probabilistic clustering problems is how many distinct clusters are present in a particular dataset. Bayesian nonparametrics (BNP) addresses this question by placing a generative process on cluster assignment, making the number of distinct clusters present amenable to Bayesian inference. However, like all Bayesian approaches, BNP requires the specification of a prior, and this prior may favor a greater or fewer number of distinct clusters. In practice, it is important to quantitatively establish that the prior is not too informative, particularly when - as is often the case in BNP - the particular form of the prior is chosen for mathematical convenience rather than because of a considered subjective belief.

We derive local sensitivity measures for a truncated variational Bayes approximation based on the Kullback-Leibler divergence. Local sensitivity measures approximate the nonlinear dependence of a VB optimum on prior parameters using a local Taylor series approximation. Using a stick-breaking representation of a Dirichlet process, we consider perturbations both to the scalar concentration parameter and to the functional form of the stick-breaking distribution.

In the design and evaluation of our local sensitivity measures we pay special attention to our ability to accurately extrapolate to different priors, rather than treating the sensitivity as a measure of robustness per se. Extrapolation motivates the use of multiplicative perturbations to the functional form of the prior for VB, as the KL divergence is then linear in the perturbation. Additionally, we linearly approximate only the computationally intensive part of inference - the optimization of the global parameters - and retain the non-linearity of easily computed quantities.s

We apply our methods to real and simulated datasets to estimate sensitivity of the expected number of distinct clusters present to the BNP prior specification, evaluating the accuracy of our approximations by comparing to the much more expensive process of re-fitting the model.

Dynamic Shrinkage Processes Mo CS

Daniel Kowal, Rice University, USA
Co-author(s): A. David S. Matteson and B. David Ruppert

We propose a novel class of dynamic shrinkage processes for Bayesian time series and regression analysis. Building upon a global-local framework of prior construction, in which continuous scale mixtures of Gaussian distributions are employed for both desirable shrinkage properties and computational tractability, we model dependence among the local scale parameters. The resulting processes inherit the desirable shrinkage behavior of popular global-local priors, such as the horseshoe prior, but provide

additional localized adaptivity, which is important for modeling time series data or regression functions with local features. We construct a computationally efficient Gibbs sampling algorithm based on a Pólya-Gamma scale mixture representation of the proposed process. Using dynamic shrinkage processes, we develop a Bayesian trend filtering model that produces more accurate estimates and tighter posterior credible intervals than competing methods, and apply the model for irregular curve-fitting of minute-by-minute Twitter CPU usage data. In addition, we develop an adaptive time-varying parameter regression model to assess the efficacy of the Fama-French five-factor asset pricing model with momentum added as a sixth factor. Our dynamic analysis of manufacturing and healthcare industry data shows that with the exception of the market risk, no other risk factors are significant except for brief periods.

Arrival time augmentation for series representations of completely random measures Mo CS

Juho Lee, AITRICS, Republic of Korea
Co-author(s): Xenia Miscouridou and François Caron

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 (generalized) 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 framework to derive series representations and finite-dimensional approximations of CRMs. Our framework can be seen as an extension of constructions based on size-biased sampling of Poisson point process (Perman et al., 1992). It includes as special cases several known series representations as well as novel ones. In particular, we show that one can get novel series representation for the generalized gamma process. We provide some analysis of the truncation error, and show how the proposed representations can be used to derive efficient variational Bayes inference algorithms for approximate posterior inference.

Tuesday 25 June

Importance conditional sampler for nonparametric mixture models Tu CS

Riccardo Corradin, University of Milano-Bicocca, Italy
Co-author(s): A. Canale and B. Nipoti

Nonparametric mixture models based on the Pitman-Yor (PY) process are a flexible tool for density estimation and clustering. Two main classes of MCMC algorithms, namely marginal and conditional, have been considered in the literature. We propose a new sampling scheme, named Importance Conditional Sampler (ICS), which, although conditional, is reminiscent of the Polya urn marginal scheme and conveniently shares its degree of interpretability. Unlike its most popular conditional competitors, the ICS does not rely on the stick-breaking representation of the underlying PY process. The performance of the ICS is investigated and compared with commonly used competitors, by means of an extensive simulation study: the ICS turns out to be uniformly efficient across the range of values that can be taken by the discount parameter of the PY, thus making efficient posterior inference possible for models specifications where popular conditional algorithms fail. The same sampling scheme can be conveniently adopted to devise an efficient algorithm for a flexible class of dependent Dirichlet process mixture models for partially exchangeable data. We illustrate our approach by analysing a dataset (Sloan Digital Sky Survey) on the colour of 25 classes of galaxies, characterized by a complex

dependence structure across classes.

Scalable Bayesian sparse survival analysis and generalized linear models via curvature-adaptive Hamiltonian Monte Carlo for high-dimensional log-concave distributions Tu CS

Akihiko Nishimura, University of California - Los Angeles, USA
Co-author(s): Marc A. Suchard

Bayesian sparse regression based on shrinkage priors possess many desirable theoretical properties and yield posterior distributions whose conditionals mostly admit straightforward Gibbs updates. Sampling high-dimensional regression coefficients from its conditional distribution, however, presents a major scalability issue in posterior computation. The conditional distribution generally does not belong to a parametric family and the existing sampling approaches are hopelessly inefficient in high-dimensional settings. Inspired by recent advances in understanding the performance of Hamiltonian Monte Carlo (HMC) on log-concave target distributions, we develop *curvature-adaptive HMC* for scalable posterior inference under sparse regression models with log-concave likelihoods. As is well-known, HMC's performance critically depends on the integrator stepsize and mass matrix. These tuning parameters are typically adjusted over many HMC iterations by collecting statistics on the target distribution — an impractical approach when employing HMC within a Gibbs sampler since the conditional distribution changes as the other parameters are updated. Instead, we achieve on-the-fly calibration of the key HMC tuning parameters through 1) the recently developed theory of *prior-preconditioning* for sparse regression and 2) a rapid estimation of the curvature of a given log-concave target via *iterative methods* from numerical linear algebra. We demonstrate the scalability of our method on a clinically relevant large-scale observational study involving n = 1,065,745 patients and p = 8,863 predictors, designed to assess the relative efficacy of two alternative hypertension treatments.

A Bayesian non-parametric methodology for inferring grammar complexity Tu CS

Robin Ryder, Université Paris-Dauphine, France
Co-author(s): L. Murray and J. Rousseau

Based on a set of strings from a language, we wish to infer the complexity of the underlying grammar. To this end, we develop a methodology to choose between two classes of formal grammars in the Chomsky hierarchy: simple regular grammars and more complex context-free grammars. To do so, we introduce a probabilistic context-free grammar model in the form of a Hierarchical Dirichlet Process over rules expressed in Greibach Normal Form. In comparison to other representations, this has the advantage of nesting the regular class within the context-free class. We consider model comparison both by exploiting this nesting, and with Bayes' factors. The model is fit using a Sequential Monte Carlo method, implemented in the Birch probabilistic programming language. We apply this methodology to data collected from primates, for which the complexity of the grammar is a key question.

Informed proposals for local MCMC in discrete spaces Tu CS

Giacomo Zanella, Bocconi University, Italy

There is a lack of methodological results to design efficient Markov chain Monte Carlo (MCMC) algorithms for statistical models with discrete-valued high-dimensional parameters. For example, it is still unclear how to extend gradient-based MCMC (e.g. Langevin and Hamiltonian schemes) to networks or partitions spaces. This is particularly relevant when fitting Bayesian nonparametric models,

which often involve combinatorial and discrete latent parameters. Motivated by this consideration, we propose a simple framework for the design of informed MCMC proposals (i.e. Metropolis-Hastings proposal distributions that appropriately incorporate local information about the target) which is naturally applicable to discrete spaces. Using Peskun-type comparisons of Markov kernels, we explicitly characterize the class of asymptotically-optimal proposal distributions under this framework, which we refer to as locally-balanced proposals. The resulting algorithms are straightforward to implement in discrete spaces and provide orders of magnitude improvements in efficiency compared to alternative MCMC schemes, including discrete versions of Hamiltonian Monte Carlo. We discuss asymptotic analysis, applications to discrete frameworks and connections to other schemes (e.g gradient-based and multiple-try ones).

Probabilistic methods for proving error bounds in GP posterior approximation Tu CS

Mikołaj Kasprzak, University of Luxembourg, Luxembourg
Co-author(s): J. H. Huggins, T. Campbell and T. Broderick

Scalable Bayesian inference methods for infinite-dimensional problems such as Gaussian process (GP) regression and inverse problems often lack a finite-data approximation theory and tools for evaluating their accuracy. In this talk, covering parts of the joint work with Jonathan Huggins, Trevor Campbell and Tamara Broderick, I will present our novel approach to bounding posterior mean and uncertainty estimates of scalable inference algorithms. In the context of GP regression, the approach is based on a new variational objective that we call the preconditioned Fisher divergence. This objective bounds the 2-Wasserstein distance from above, which in turn provides tight bounds on the pointwise error in the mean and variance functions. Unlike the Wasserstein distance, however, the preconditioned Fisher divergence can be computed efficiently. Some of its appealing properties may be proved using probabilistic techniques involving infinite-dimensional stochastic differential equations, which I will introduce and explain. I will also discuss some other potential application areas of the preconditioned Fisher divergence and our bounding techniques.

Bernstein–von Mises Theorems for linear inverse problems Tu CS

Hanne Kekkonen, University of Cambridge, UK
Co-author(s): M. Giordano

We consider the statistical inverse problem of approximating an unknown function f from a linear measurement corrupted by additive Gaussian white noise. We employ a nonparametric Bayesian approach with standard Gaussian prior for f and prove a semi-parametric Bernstein–on Mises theorem for a large collection of functionals of f, which implies that semiparametric posterior-based inferences and uncertainty quantification are valid and optimal from a frequentist point of view.

We also demonstrate how the approach can be refined to attain a full nonparametric theorem, in an example of recovering the source function in elliptic boundary value problem, which entails the convergence of the posterior distribution to a fixed infinite-dimensional Gaussian probability measure with minimal covariance in suitable function spaces.

Bayesian high-dimensional analyses for a multivariate linear regression model and a sparse spiked-covariance model. Tu CS

Bo Ning, Yale University, USA
Co-author(s): Seonghyun Jeong and Subhashis Ghosal

In this talk, I will present two recent studies on Bayesian high-dimensional analyses. The first study is about a multivariate linear regression model. The covariance matrix of the model is unknown and group sparsity is imposed on the regression coefficients. A product of independent spike-and-slab priors is imposed on the regression coefficients. A posterior contraction rate is derived with respect to the average log-affinity. The uncertainty for the regression coefficients is quantified with frequentist validity through a Bernstein-von Mises type theorem, and the result leads to selection consistency for the Bayesian method. The second study is about a sparse spiked-covariance model. A continuous spike-and-slab prior with group sparsity is placed on the eigenvectors. A posterior contraction rate is derived and an EM algorithm is developed to compute the posterior.

Uncertainty quantification for survival analysis Tu IS

Stéphanie van der Pas, Leiden University, The Netherlands
Co-author(s): A. Ismael Castillo

The Bayesian framework offers an intuitive approach towards uncertainty quantification. We discuss our recent theoretical advances towards uncertainty quantification for several survival objects within the Bayesian paradigm.

On distributed Bayesian computation Tu IS

Botond Szabo, Leiden University, Netherlands

First, the theoretical properties of various distributed Bayesian methods are investigated on the benchmark signal-in-Gaussian-white-noise-model for known regularity parameter β. We show that some seemingly reasonable methods proposed in the literature can provide sub-optimal recovery and misleading uncertainty quantification, while others perform optimally. Then we investigate the adaptive setting, where the regularity parameter β is unknown and show that the standard methods proposed in the literature (typically) fail both for recovery and uncertainty quantification.

On multiscale properties of some Bayesian tree methods Tu IS

Ismael Castillo, Sorbonne University, Paris, France
Co-author(s): V. Rockova

Bayesian tree methods are broadly used in the practice of nonparametrics, through algorithms such as Bayesian CART or BART among others. The theoretical understanding of the empirical success of these methods is just starting to develop. In this talk, I will present some recent work in this direction, discussing multiscale properties of posterior tree-based distributions. This is joint work with Veronika Rockova (Chicago).

Recent developments in model specification, regularization, and summarization for nonparametric Bayesian models of heterogeneous treatment effects. Tu IS

Jared Murray, UT Austin, USA

We present a review of some recent developments in Bayesian approaches to estimating heterogeneous treatment effects in experiments or in observational data when all the confounders are available. There is great interest in adapting flexible general-purpose Bayesian nonparametric models for this task, but effectively doing so poses a number of challenges: Most nonparametric models provide no direct mechanism for shrinking heterogeneous treatment effects toward homogeneity without unduly

shrinking the contribution of other covariates (including confounders). Further, seemingly sensible prior distributions can actually induce significant bias – of unknown sign and magnitude – in treatment effect estimates, a phenomenon known as "regularization induced confounding" (Hahn et al, 2017). And once we have “turned the crank”, we are left with a complicated posterior distribution that does not directly deliver the inferences relevant to scientists and policymakers. We describe how we have addressed these issues in the context of Bayesian tree models, and discuss the implications for other models of heterogeneous treatment effects.

Bayesian nonparametric learning through randomized loss functions and posterior bootstraps Tu IS

Christopher Holmes, University of Oxford, UK

We introduce Bayesian nonparametric learning whereby Bayesian nonparametric models are used to train Bayesian parametric models by way of suitably randomized objective (loss) functions. The resulting posterior models exhibit provably better properties than their conventional Bayesian counterparts when the sampling distribution (likelihood function) is misspecified. For additive log-likelihoods inference is achieved through posterior sampling obtained by independent optimizations of randomly re-weighted loss-functions, as opposed to Markov chain Monte Carlo sampling. This avoids issues with MCMC such as burn-in and chain dependence, and is highly scalable on modern computer architectures allowing for samples to be drawn in parallel for the price of a single model optimization. We demonstrate the approach on a number of examples including nonparametric learning for Bayesian logistic regression, variational Bayes (VB), mixture models, and random forests. The work has its foundations in the weighted-likelihood bootstrap of Newton and Raftery (1994).

Testing for model misspecification Tu IS

Natalia Bochkina, University of Edinburgh and the Alan Turing Institute, UK

We propose a general method to test for model specification in an asymptotic setting for large data sets that applies to regular models. It is based on the Bernstein-von Mises theorem for misspecified Bayesian models. It checks for misspecification of the likelihood but can be applied to check for joint prior and likelihood misspecification by applying it to the marginal likelihood, as long as the Bernstein-von Mises theorem holds for the remaining (hyper)-parameters. For simple models (e.g Poisson likelihood) it coincides with ad hoc goodness-of-fit tests used in practice. We will illustrate how this method applies to the exponential likelihood family as well as to Student-t distribution. It can be used to check, for instance, whether an approximate likelihood provides a reasonable approximation in the context of asymptotic inference, and it applies to data with either small or large number of parameters. This method will be illustrated on real and simulated data sets.

Local Exchangeability Tu CS

Trevor Campbell, UBC, Canada
Co-author(s): Saifuddin Syed, Chiao-Yu Yang, Michael Jordan, and Tamara Broderick

Exchangeability—in which the distribution of an infinite sequence is invariant to reorderings of its elements—has powerful implications for probabilistic modeling and inference. In practice, however, this assumption is too strong an idealization; the distribution typically fails to be exactly invariant to permutations, and the useful representation theory due to de Finetti does not apply. This motivates the need for a distributional assumption that is both weak enough to hold in practice, and strong enough to have useful implications for modeling and inference. We thus introduce a relaxed notion

of local exchangeability—where swapping data associated with nearby covariates causes a bounded change in the distribution—and provide examples of popular statistical models that exhibit this property. Next, we show that locally exchangeable processes correspond to independent observations from an underlying unique smooth measure-valued stochastic process, providing justification for the Bayesian modelling approach in the spirit of de Finetti's theorem. The talk concludes with an investigation of sample continuity properties of locally exchangeable processes on the real line.

Statistically efficient coalescent inference Tu CS

Julia Palacios, Stanford University, USA
Co-author(s): A. Veber and L. Cappello

A popular model for Bayesian inference of evolutionary parameters from molecular sequence data relies on Kingman's coalescent; a model on the ancestral relationships of the samples represented by a labeled bifurcating tree (genealogy). However inference is hampered by the size of the hidden state space of labeled bifurcating trees. In order to improve computational efficiency, different lower resolution coalescent models have been proposed that while still model the correlated structure of the samples through binary tree representations, these models have smaller state space and have been successfully implemented and applied to data. In this work, we prove that statistical inference from these lower resolution coalescent models is more statistically efficient than inference from the standard Kingman coalescent model.

Nonparametric Bayesian Modeling of Correlation Functions for Global Data Tu CS

Fernando Quintana, Pontificia Universidad Católica de Chile, Chile
Co-author(s): Emilio Porcu, Pier Giovanni Bissiri and Felipe Tagle

We provide a non-parametric spectral approach to the modeling of correlation functions on spheres. The sequence of Schoenberg coefficients and their associated covariance functions are treated as random rather than assuming a parametric form. We propose a stick-breacking representation for the spectrum, and show that such a choice spans the support of the class of geodesically-isotropic covariance functions under uniform convergence. Further, we examine the first order properties of such representation, from which geometric properties can be inferred, in terms of Hölder continuity, of the associated Gaussian random field. The properties of the posterior, in terms of existence, uniqueness, and Lipschitz continuity, are then inspected. Our findings are validated with MCMC simulations and illustrated using a global data set on surface temperatures.

Bayesian inference for finite-dimensional discrete priors Tu CS

Tommaso Rigon, Bocconi University, Italy
Co-author(s): A. Lijoi and I. Prünster

Discrete random probability measures are the main ingredient for addressing Bayesian clustering. The investigation in this area has been very lively, with strong emphasis on nonparametric procedures based either on the Dirichlet process or on more flexible generalizations, such as the Pitman–Yor (PY) process or the normalized random measures with independent increments (NRMI). The literature on finite-dimensional discrete priors, beyond the classic Dirichlet-multinomial model, is much more limited. We aim at filling this gap by introducing novel classes of priors closely related to the PY process and NRMIs, which are recovered as limiting case. Prior and posterior distributional properties are extensively studied. Specifically, we identify the induced random partitions and determine explicit expressions of the associated urn schemes and of the posterior distributions. A detailed comparison

with the (infinite-dimensional) PY and NRMIs is provided. Finally, we employ our proposal for mixture modeling, and we assess its performance over existing methods in the analysis of a real dataset.

Dependent neutral-to-the-right priors for software reliability data Tu CS

Fearghal Donaghy, Trinity College, Dublin, Ireland
Co-author(s): A. Dr Bernardo Nipoti

With each update of its browser, Firefox receives reports of the time of discovery of many bugs associated with that update. We propose a model for the discovery time distribution of each release, which allows for borrowing of information across releases. To this end, we use superposed completely random measures to construct a vector of dependent neutral-to-the-right priors. The model is completed by accounting for an unobserved number of undiscovered, and thus considered right-censored, bugs per release. An explicit characterisation of the posterior distribution of the defined vector of dependent neutral-to-the-right priors is derived and, in turn, used to devise an efficient marginal Markov chain Monte Carlo sampler for posterior inference. While presented for the analysis of the Firefox data, our approach could potentially be useful across a wide range of applications of survival analysis and reliability.

Bayesian Non Parametric approaches for Survival analysis Tu CS

Sarah Filippi, Imperial College London, UK
Co-author(s): J. Rousseau and C. Holmes

In this presentation, we will explore the use of Bayesian nonparametric approaches for survival analysis. In particular, we will focus on procedures modelling the underlying distributions with Polya Tree and Dirichlet Process priors. We will discuss inference and two-sample testing for right-censored survival data as well as models taking into account the effect of covariates on survival time.

Clustering via Dirichlet Process Mixture Models for Trajectories with Fixed Effects Continuous-Time Hidden Markov Models Tu CS

Yu Luo, McGill University, Canada
Co-author(s): David Stephens and David Buckeridge

Large amounts of data that exist in the form of longitudinal health records, such as electronic health records, healthcare administrative databases and data derived from mobile health applications, are now available for dynamic monitoring of the underlying processes governing the observations. However, the underlying status, and its development across time, that is generating the observations is not observed directly and so requires inferential methods to ascertain progression. Moreover, records are observed when a subject interacts with the system, resulting in irregular observation time where the observations are not collected at equidistant time intervals with possible sparsity. These considerations suggest that the observed trajectories should be modeled as a latent continuous-time process. We develop a continuous-time hidden Markov model as the basis for a state space generalized linear model (CTHMM-GLM) to analyze the longitudinal data accounting for irregular visits, different types of observations and multiple time-dependent covariates. In addition, we extend the CTHMM-GLM for infinite mixture model-based clustering. This approach is appealing as it allows for the number of cluster to be learned as a function of the sample size. The Markov chain Monte Carlo algorithm is explored to obtain the sample from the posterior distribution, specifically restricted Gibbs sampling with split-merge proposals. The simulation study demonstrates that the proposed sampling scheme could identify the correct number of clusters from which the true data generated. We apply this

model to cluster a real cohort, and focus on the condition chronic obstructive pulmonary disease from healthcare administrative data from Montreal with the outcome being the number of drugs taken. A four-cluster model is chosen with each cluster having its own transition characteristics. Our model of the real data application demonstrated that the model could identify the meaningful clusters to help healthcare system managers measure the performance of the healthcare system over time.

Survival analysis regression Tu CS

Alan Riva-Palacio, University of Kent, United Kingdom
Co-author(s): Fabrizio Leisen and Jim Griffin

We present a Bayesian nonparametric model for regression in survival analysis. Our models builds on the classical neutral to the right model of Doksum (1974), on the Cox proportional hazards model for neutral to the right distributions as in Kim and Lee (2003) and the multiple-sample information model of Riva-Palacio and Leisen (2018). The use of vectors of dependent Bayesian non-parametric priors allows us to efficiently model the competing risks among the covariates. The model is quite flexible as it allows for the borrowing of information across covariates and does not necessarily satisfy the proportional hazards constraint. We prove theoretical results including the characterization of the posterior behaviour of the underlying dependent vector of completely random measures and the showcasing of consistency in our model. We also use a pseudo-marginal methodology to produce posterior means. Finally, applications of the models to different real datasets are illustrated.

Wednesday 26 June

Title TBA We KL

Tamara Broderick, MIT, USA

Abstract TBA.

Posterior Consistency In Large Graph Limits of Learning Algorithms We IS

Andrew Stuart, California Institute of Technology, USA

Many problems in machine learning require the classification of high dimensional data. One methodology to approach such problems is to construct a graph whose vertices are identified with data points, with edges weighted according to some measure of affinity between the data points. Algorithms such as spectral clustering, probit classification and the Bayesian level set method can all be applied in this setting. The goal of the talk is to describe these algorithms for classification, and analyze them in the limit of large data sets. Doing so leads to interesting problems in the realm of (Bayesian) posterior consistency and these problems are described.

Mixtures and inverse or semi-parametric problems We IS

Judith Rousseau, University of Oxford, UK
Co-author(s): N. Bochkina, J.B. Salomond, C. Scricciolo and Y. van der Molen

Although extremely flexible and popular in Bayesian nonparametric modeling, Bayesian nonparametric mixture models have proved very difficult to analyse and understand, theoretically. So far mostly posterior concentration rates on direct problems have been proved. In this talk I will review a few advances we have obtained in the context of nonparametric mixture models, associated to either

inverse or semi-parametric models. These will include inverse problems such as deconvolution and some semi-parametric Bernstein von Mises results. In the former case the mixing distribution has possibly infinite support while in the latter the mixing distribution has finite support and the kernels (or emission disributions) are unknown.

Nonparametric Bayesian drift estimation for multidimensional diffusions We IS

Kolyan Ray, Kings College London, United Kingdom
Co-author(s): A. Richard Nickl

We consider the problem of estimating the drift and invariant measure of a periodic multidimensional diffusion based on continuous observations, a model used for instance in molecular dynamics. Placing a high dimensional Gaussian prior on the drift, we obtain convergence rates for the posterior distribution and the corresponding maximum a posteriori (MAP) estimate. For dimension at most 3, we further obtain Bernstein-von Mises type results for the posterior distributions of both the drift and invariant measure. This provides a frequentist justification for the Bayesian approach in this model, including for uncertainty quantification.

Rates of contraction: some non-Gaussian priors and some nonlinear inverse problems We IS

Sergios Agapiou, University of Cyprus, Cyprus
Co-author(s): M. Dashti and T. Helin and P. Math

Priors with heavier than Gaussian tails are gaining popularity in a range of applications, due to their sparsity-promoting and edge-preserving properties. We study the rate of contraction of posterior distributions arising from such priors, in infinitely-informative asymptotic limits under frequentist assumptions on the data. Building on the seminal work of A. van der Vaart and J. H. van Zanten for Gaussian process priors, we will present a general posterior contraction result for a family of priors called p-exponential, which have tails ranging between Laplace and Gaussian. We will use this result to study rates of contraction in direct problems in the small noise limit (white noise model).

We will also study rates of contraction for inverse problems. Our methodology for passing from rates for direct to rates for inverse problems relies on the notion of the modulus of continuity, first used in this context in a recent work by B. T. Knapik and J-P Salomond. Our methodology applies to linear and nonlinear problems of integral type.

This is joint work with Masoumeh Dashti, Tapio Helin (arXiv:1811.12244) and Peter Math (work in progress).

Double Feature Allocation for Phenotyping with Electronic Health Records Data We IS

Yang Ni, Texas A&M University, USA
Co-author(s): A. Peter Mueller and B. Yuan Ji

In this talk, we will introduce a categorical matrix factorization method to infer latent diseases from electronic health records data in an unsupervised manner. A latent disease is defined as an unknown biological aberration that causes a set of common symptoms for a group of patients. The proposed approach is based on a novel double feature allocation model which simultaneously allocates features to the rows and the columns of a categorical matrix. Using a Bayesian approach, available prior information on known diseases greatly improves identifiability of latent diseases. This includes known

diagnoses for patients and known association of diseases with symptoms. We validate the proposed approach by simulation studies including mis-specified models and comparison with sparse latent factor models. In the application to Chinese electronic health records (EHR) data, we find interesting results, some of which agree with related clinical and medical knowledge.

Fast Search for General Bayesian Nonparametric Mixture Models We IS

George Karabatsos, University of Illinois-Chicago, U.S.A.

Bayesian nonparametric (BNP) infinite-mixture models provide flexible and accurate methods of density estimation, cluster analysis, and regression, for a wide range of scientific fields. The current Big Data era raises major computational challenges for estimating quantities from the BNP mixture model posterior distribution using standard MCMC and sequential Monte Carlo (SMC) methods. To address these challenges, we introduce a new fast-search clustering algorithm for estimating BNP mixture models, which is easy to use and can handle mixtures of general univariate or multivariate distributions. The new fast search algorithm is based on a stochastic EM algorithm, and can rapidly and accurately estimate the posterior predictive density, and the posterior mode of the cluster assignments through annealing, even for data sets containing millions of observations. If necessary, these estimates may be used to provide a quick start to a MCMC or SMC algorithm in order to estimate the full posterior distribution. The new fast-search algorithm is easily applicable to virtually the full range of BNP priors, because it only relies on the ability to generate samples from the given BNP prior distribution, using a Ferguson-Klass or other suitable sampling algorithm. They include BNP priors for the class of normalized random measures, such as the normalized generalized gamma process; the generalized Dirichlet process; stable 3-parameter beta process; the Poisson-Dirichlet process; BNP priors for other Completely Random Measures that admit an explicit series or superposition representations (resp.), and future novel BNP priors. Further, the new algorithm can be naturally extended to handle the analysis of online streaming data, while making a single pass through all the data points, which is required for such rapidly-arriving data. In contrast, previous fast search algorithms are limited to specific tractable stick-breaking priors, such as the Dirichlet process; and are unable to provide both density and cluster estimation in streaming data analysis while making a single pass through all data points. The new algorithm is illustrated through the analysis of simulated and real data sets. This includes causal analysis based on posterior inferences of conditional regressions from BNP multivariate normal mixtures.

Bayesian Nonparameteric Models for Richly Structured Data in Biomedicine We IS

Veera Baladandayuthapani, University of Michigan, USA

Modern scientific endeavors generate high-throughput, multi-modal datasets of different sizes, formats, and structures at a single subject-level. In the context of biomedicine, such data include multi-platform genomics, proteomics and imaging; and each of these distinct data types provides a different, partly independent and complementary, high-resolution view of various biological processes. Modeling and inference in such studies is challenging, not only due to high dimensionality, but also due to presence of rich structured dependencies such as serial, functional, tree, and shape-based correlations. In this talk I will cover some regression and clustering frameworks for modeling data, where the observations (statistical atoms) lie on non-standard spaces such as densities, trees and shapes. Using coherent data-based projections (basis functions and metric spaces), we will show how to build probabilistic frameworks that can extract maximal information from such data for inference. These approaches will be illustrated using several biomedical case examples especially in oncology.

Clustering data at multiple resolutions We CS

Cecilia Balocchi, University of Pennsylvania, US
Co-author(s): A. Shane T. Jensen and B. Edward I. George

In many applications, data is organized in a hierarchical structure with observations measured at different levels of resolution. We consider the problem of clustering data at multiple resolutions, when the different units are organized in a hierarchy that can be described by a tree: higher resolution entities are nested within lower resolution ones. A motivating example is the modeling of crime in urban environments at different spatial resolutions: US cities are divided into census tracts, which are divided into census block groups, which are further split into census blocks. We want to partition a city into regions with similar crime frequencies at each resolution while sharing information between partitions at higher and lower resolutions. If we knew the partition at higher levels, such as the census tract level, Hierarchical Dirichlet Processes (Teh et al. 2005) would be an appropriate model. Nested Dirichlet Processes (Rodriguez et al. 2008) instead allow to model partitions at multiple levels but would restrict block group clusters to be nested into census tract ones. Latent Nested Processes (Camerlenghi et al. 2018) extend this model to allow for more flexible partitions that do not have this constraint. We further extend Nested Dirichlet Processes by combining nested and hierarchical processes: our model incorporates the same flexibility as Latent Nested Processes while improving interpretability by allowing for block groups in different census tract clusters to take on the same values. We apply this method to crime frequencies in Philadelphia; however this is a general problem and this model can be applied in many other settings, for example in neuroimaging where one could consider functional regions in the brain at different resolutions.

Bayesian prediction in feature models We CS

Federico Camerlenghi, University of Milano - Bicocca and Collegio Carlo Alberto, Italy
Co-author(s): T. Broderick, S. Favaro and L. Masoero

The prediction of future outcomes of a random phenomenon is undoubtedly a fundamental goal of statistics. Bruno de Finetti wrote “science cannot limit itself to theorize about accomplished facts but must foresee", emphasizing the importance of prediction in statistical inference. Such a fundamental goal can be achieved if one assumes a sort of *analogy* across observations; in a Bayesian setting, natural notions of analogy are exchangeability and partial exchangeability. We define and investigate different classes of Bayesian nonparametric models suitable for exchangeable and partially exchangeable data, which are useful for prediction in feature sampling. Within these models we are able to forecast the outcome of additional samples having arbitrary size and to derive the exact posterior distribution for many statistics of interest, such as the number of hitherto unseen features that will be observed in future sampling.

The Multidimensional Partitioning Tree Process We CS

Lloyd T. Elliott, Simon Fraser University, Canada
Co-author(s): Shufei Ge, Yee Whye Teh and Liangliang Wang

The Mondrian process is a powerful model for space partitioning and it is appropriate for multidimensional data. However, the model flexibility is restricted by the axis alignment of its cuts. The Ostomachion process and the self consistent Binary Space Partitioning (BSP)-Tree process were recently introduced as approaches for space partition modeling as generalizations of the Mondrian process with non axis-aligned cuts in two dimensional space. In this work, we propose a model for the multidimensional BSP-Tree process with non-axis aligned cuts. The partition is described by a set of polytopes, and we propose a sequential Monte Carlo algorithm for inference about non-axis aligned partition structures. We present simulation studies for the three dimensional case, and demonstrate

higher accuracy over axis-aligned methods. We discuss applications to bioinformatics such as epistasis in gene expression data.

The Attraction Indian Buffet Distribution We CS

Richard Warr, Brigham Young University, USA
Co-author(s): David Dahl and Richard Warr

Latent feature models seek to uncover hidden categorical variables that explain observed data. These models often use the Indian buffet process (IBP), a distribution over a binary feature matrix with an infinite number of columns and one row per observation. The IBP assumes that the observations are exchangeable, which is not reasonable in the presence of pairwise similarity information. We propose the attraction Indian buffet distribution (aIBD), a distribution for a binary feature matrix indexed by pairwise similarity. Our formulation preserves many of the properties of the original IBP, including having the same distribution of the total number of features. Thus, much of the interpretation and intuition that one has for the IBP carries over directly to our aIBD. A temperature parameter controls the degree to which the similarity information affects feature sharing. The probability function can be written explicitly and has a tractable normalizing constant, making posterior inference on hyperparameters straight-forward using standard MCMC methods. We demonstrate the feasibility and performance of our method in examples.

Bayesian nonparametric inference of population trajectories via Tajima heterochronous n-coalescent We CS

Lorenzo Cappello, Stanford University, USA
Co-author(s): Julia A. Palacios

The observed variation in gene samples allows to infer evolutionary parameters such as past population dynamics: it is common practice to model such a variation as a mutation process superimposed on a stochastic genealogy sampled from the Kingman n-coalescent. However, the state space of Kingman's genealogies grows superexponentially; thus inference is computationally unfeasible already for small sample sizes. An alternative to Kingman coalescent has been proposed in the literature, the Tajima n-coalescent, which relies on a coarser resolution, reducing the state space substantially. Such process does not accomodate samples collected at different times, a situation that in applications is both real (e.g. ancient DNA, influenza viruses) and desirable (it reduces the variance of the estimate). In order to fill this gap, we introduce a new process, called Tajima heterochronous n-coalescent, define the exact likelihood of observed mutations given a Tajima's genealogies, and present a Bayesian nonparametric procedure to infer past population size. We propose also a new sampler to explore the space of Tajima's genealogies. We compare our procedure with state-of-art algorithms.

Computable inference for a class of non-linear state-space models We CS

Guillaume Kon Kam King, University of Turin and Collegio Carlo Alberto, Italy
Co-author(s): O. Papaspiliopoulos and M. Ruggiero

Filtering hidden Markov models, or sequential Bayesian inference on the hidden state of a signal, is analytically tractable only for a handful of models. Examples are finite-dimensional state space models and linear Gaussian systems (Baum-Welch and Kalman filters). Recently, Papaspiliopoulos et al.(2014,2016) proposed a principled approach for extending the realm of analytically tractable models, exploiting a duality relation between the hidden process and an auxiliary process. Then, the solution of the filtering problem consists in a finite mixture of distributions. We study the computational

effort required to implement this strategy for two parametric and nonparametric models: the Cox-Ingersoll-Ross process, the K-dimensional Wright-Fisher process, the Dawson-Watanabe process and the Fleming-Viot process. In all cases, the number of components involved in the filtering distributions increases rapidly with the number of observations. Although this could render the algorithm impractical for long observation sequences and undermine its practical relevance, the mathematical form of the filtering distributions suggest that the number of components which contribute most to the mixture remains small. This suggests several efficient natural approximation strategies. We assess the performance of these strategies in terms of accuracy, speed and prediction, benchmarked against the exact solution.

Beyond Whittle: Nonparametric Correction of a Parametric Likelihood with a Focus on Bayesian Time Series Analysis We CS

Renate Meyer, University of Auckland, New Zealand
Co-author(s): C. Kirch, M. Edwards, and A. Meier

Various nonparametric Bayesian approaches to time series analysis have been developed. Most notably, Carter and Kohn (1997), Gangopadhyay (1998), Choudhuri et al. (2004), and Rosen et al (2012) used Whittle's likelihood for Bayesian modeling of the spectral density as the main nonparametric characteristic of stationary time series. As shown in Contreras-Cristan et al. (2006), the loss of efficiency of the nonparametric approach using Whittle's likelihood approximation can be substantial. On the other hand, parametric methods are more powerful than nonparametric methods if the observed time series is close to the considered model class but fail if the model is misspecified. Therefore, we suggest a nonparametric correction of a parametric likelihood that takes advantage of the efficiency of parametric models while mitigating sensitivities through a nonparametric amendment. We use a nonparametric Bernstein polynomial prior on the spectral density with weights induced by a Dirichlet process. Contiguity and posterior consistency for Gaussian stationary time series have been shown in by Kirch et al (2019). Bayesian posterior computations are implemented via a MH-within-Gibbs sampler and the performance of the nonparametrically corrected likelihood is illustrated in a simulation. We illustrate this approach through applications in physiology, ecology and astrophysics: analysing heart rate variability in ECG time series, the Southern Oscillation Index, and LIGO gravitational wave data.

Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data We CS

Xenia Miscouridou, University of Oxford, UK
Co-author(s): François Caron and Yee Whye Teh

We propose a novel class of network models for temporal dyadic interaction data. Our objective is to capture important features often observed in social interactions: sparsity, degree heterogeneity, community structure and reciprocity. We use mutually-exciting Hawkes processes to model the interactions between each (directed) pair of individuals. The intensity of each process allows interactions to arise as responses to opposite interactions (reciprocity), or due to shared interests between individuals (community structure). For sparsity and degree heterogeneity, we build the non time dependent part of the intensity function on compound random measures following Todeschini et al., 2016. We conduct experiments on real-world temporal interaction data and show that the proposed model outperforms competing approaches for link prediction, and leads to interpretable parameters.

Thursday 27 June

Approximate multiple shrinkage for clustered regressions Th CS

Sameer Deshpande, CSAIL, MIT, USA
Co-author(s): A. Cecilia Balocchi and B. Edward George and C. Shane Jensen

We consider the problem of fitting multiple linear regression models to grouped data in the setting where components of the regression parameters are clustered across the groups. In particular, we allow for the possibility that the latent clustering differ across components. Our work is motivated by a case study about crime in the city of Philadelphia, in which the spatial clustering of the base level of crime and the time trends of crime within census tracts may differ. In this setting, conventional stochastic search techniques are computationally prohibitive due to the combinatorial vastness of the latent product space of partitions.

In this work, we describe an ensemble optimization procedure targeting the -tuples of partitions with largest posterior probability. At a high level, this procedure can be viewed as running several greedy searches over the posterior distribution that are made “mutually aware" through an entropic penalty that discourages search trajectories from collapsing upon one another. We use the results of our optimization procedure to approximate the full posterior means of the group-specific regression coefficients. We also discuss computational simplifications of our optimization objective that enable faster posterior exploration.

A Hierarchical Dependent Dirichlet process prior for modelling bird migration patterns in the UK Th CS

Alex Diana, University of Kent, UK
Co-author(s): E. Matechou, J. E. Griffin and A. Johnston

Environmental changes in recent years have been linked to shifts in the migration patterns of birds, which in turn are linked to the survival of species. Our work is motivated by capture-recapture data on blackcaps collected by the British Trust for Ornithology as part of the Constant Effort Sites (CES) monitoring scheme. Blackcaps overwinter abroad and migrate to the UK annually for breeding purposes. As part of the CES, data on blackcaps are collected at more than 100 sites across more than 20 years. However, CES site and year specific data are sparse and include information only on a small number of individuals. This prohibits us from estimating site and year specific migration patterns and population sizes, unless a joint modelling approach is employed. There is also considerable interest in determining the effect of environmental conditions, due to climate change, on these migration patterns.

Individual migration patterns are summarised by an arrival time and length of stay. We assume these times at each site and in each year are drawn jointly from an inhomogeneous Poisson process, whose normalised intensity is described by a Gaussian mixture model, with the Hierarchical Dependent Dirichlet process (HDDP) as the mixing measure. The HDDP combines the idea of the Hierarchical Dirichlet process (HDP) and the Dependent Dirichlet process (DDP). The first allows us to jointly model data across different sites. The latter allows us to introduce a continuous covariate, in this case the North-Atlantic Oscillation, in the means of the mixture components of arrival and length of stay. The correlation structure of the means over time is modelled using a multivariate Gaussian process, which allows us to have a bivariate output while keeping a conjugate structure. As a result, the weights of the mixture change across groups, while the cluster locations change according to the year-specific covariate.

Results show that a large part of the population tends to arrive mainly at the start of the season, and these birds can be assumed to be breeding birds, while the remaining of the birds are probably the transient birds staying only for a few weeks before moving to other sites.

The proposed modelling framework is extremely general and can be used in any context where multivariate density estimation is performed jointly across different groups and in the presence of a continuous covariate.

A Bayesian nonparametric approach to disclosure risk assessment Th CS

Stefano Favaro, University of Torino and Collegio Carlo Alberto, Italy

Protection against disclosure is a legal and ethical obligation for agencies releasing microdata files for public use. When sample records are cross-classified according to categorical identification variables (key variables), any decision about release is supported by measures of disclosure risk, the most common being the number τ_1 of sample uniques cells that are also population uniques. We first make use of tools at the interface between Bayesian nonparametrics and the theory of exchangeable random partitions to develop a methodology that makes inference on τ_1 exact, computationally efficient and of easy implementation and reproducibility. Our approach relies on: i) a generalized Poisson-Dirichlet prior for modeling the random partition induced by the cross-classification of sample records; ii) an empirical Bayes approach for estimating prior parameters in such a way to recognize a primary role to sample unique cells. These minimal model assumptions lead to an explicit, and simple, expression for the posterior distribution of τ_1, which allows to avoid the use of Markov chain Monte Carlo methods for posterior approximation. The proposed approach is tested on data from the U.S. 2000 census for the state of California and for the state of New York, revealing the same good experimental performance as recent Bayesian hierarchical semiparametric approaches that rely on modeling association among key variables at the cost of an increased computational effort.

A New Class of Time Dependent Latent Factor Models with Applications Th CS

Michael Zhang, Princeton University, USA
Co-author(s): Sinead Williamson and Paul Damien

In many applications, observed data are influenced by some combination of latent causes. For example, suppose sensors are placed inside a building to record responses such as temperature, humidity, power consumption and noise levels. These random, observed responses are typically affected by many unobserved, latent factors (or features) within the building such as the number of individuals, the turning on and off of electrical devices, power surges, etc. These latent factors are usually present for a contiguous period of time before disappearing; further, multiple factors could be present at a time.

This paper develops new probabilistic methodology and inference methods for random object generation influenced by latent features exhibiting temporal persistence. Every datum is associated with subsets of a potentially infinite number of hidden, persistent features that account for temporal dynamics in an observation. The ensuing class of dynamic models constructed by adapting the Indian Buffet Process — a probability measure on the space of random, unbounded binary matrices — finds use in a variety of applications arising in operations, signal processing, biomedicine, marketing, image analysis, etc. Illustrations using synthetic and real data are provided.

Bayesian Quantile Mixture Regression Th CS

Athanasios Kottas, University of California, Santa Cruz, USA

Co-author(s): A. Yifei Yan

Quantile regression is widely used to study the relationship between predictors and a conditional quantile of the response variable. We will present semiparametric Bayesian methodology for modeling the conditional response distribution with a weighted mixture of quantile regression components. We specify a common regression coefficient vector for all components in order to synthesize information from multiple parts of the response distribution. Different from simultaneous quantile regression, the goal is to obtain a combined estimate of the predictive effect of each covariate. For the mixture kernel density, we work with an extension of the asymmetric Laplace distribution that offers more flexible tail behavior. The mixture weights are developed through increments of a random distribution function. Model performance in prediction and variable selection will be illustrated with both synthetic and real data sets.

Flexible perturbation models for robustness to misspecification Th CS

Jeffrey Miller, Harvard T.H. Chan School of Public Health, USA

In many applications, there are natural statistical models with interpretable parameters that provide insight into questions of interest. While useful, these models are almost always wrong in the sense that they only approximate the true data generating process. In some cases, it is important to account for this model error when quantifying uncertainty in the parameters. We propose to model the distribution of the observed data as a perturbation of an idealized model of interest by using a nonparametric mixture model in which the base distribution is the idealized model. This provides robustness to small departures from the idealized model and, further, enables uncertainty quantification regarding the model error itself. Inference can easily be performed using existing methods for the idealized model in combination with standard methods for mixture models. Remarkably, inference can be even more computationally efficient than in the idealized model alone, because similar points are grouped into clusters that are treated as individual points from the idealized model. We demonstrate with simulations and an application to flow cytometry.

Projected Polya tree Th CS

Luis Nieto-Barajas, ITAM, Mexico
Co-author(s): A. Gabriel Núñez

One way of defining probability distributions for circular variables (directions in two dimensions) is to radially project probability distributions, originally defined on $\mathbb{R}^2$, to the unit circle. Projected distributions have proved to be useful in the study of circular and directional data. Although any bivariate distribution can be used to produce a projected circular model, these distribution are typically parametric. In this talk we consider a bivariate Pólya tree on $\mathbb{R}^2$ and project it to the unit circle to define a new Bayesian nonparametric model for circular data. We study the properties of the proposed model, obtain its posterior characterisation and show its performance with simulated and real datasets.

Joint Quantile Regression under Dependency Th CS

Surya Tokdar, Duke University, USA
Co-author(s): A. Xu Chen

Four decades ago, Roger Koenker and Gib Basett introduced the idea of quantile regression (QR). Today, QR is widely recognized as a fundamental statistical tool for analyzing complex predictor-response relationships, with a growing list of applications in ecology, economics, education, public

health, climatology, and so on. In QR, one replaces the standard regression equation of the mean with a similar equation for a quantile at a given quantile level of interest. But the real strength of QR lies in the possibility of analyzing any quantile level of interest, and perhaps more importantly, contrasting many such analyses against each other with fascinating consequences.

In spite of the popularity of QR, it is only recently that an analysis framework has been developed (Yang and Tokdar, JASA 2017) which transforms Koenker and Basett's four-decade old idea into a model based inference and prediction technique in its full generality. In doing so, the new joint estimation framework has opened doors to many important advancements of the QR analysis technique to address additional data complications. In this talk I will present recent such developments, specifically focusing on the issue of additional dependence between observation units. Such dependency manifests in many common situations, e.g., when one simultaneously measures multiple response variables per observation unit, when a response is measured repeatedly over time and/or space, or, when data is drawn from a network of individuals.

Bayesian Predictive Synthesis Th IS

Mike West, Duke University, USA

Bayesian Predictive Synthesis (BPS) concerns the evaluation, calibration, comparison and combination of probability distributions in inferential and predictive settings. In recent years the methodology of BPS has been established and developed with primary focuses on problems involving multiple forecast models or agents, where the probability distributions are those arising from these models or agents in specific prediction problems. Macroeconomic time series forecasting, where corresponding time series of predictive distributions from multiple competing models, and/or from multiple forecasting groups or individuals, has been and remains one of the main focus area. Theoretically, BPS includes as special cases traditional Bayesian (and other) model weighting/averaging, as well as many other "probability combination" rules, but goes beyond as it defines a foundation in subjective Bayesian analysis. From that foundation, all such methodological approaches can be understood, criticized and– if desired– modified and generalised. The essential foundations of BPS lies in "old-style" Bayesian nonparametrics (BNP) and what was an outgrowth of the "Bayesian robustness" literatures in the 1980s/90s. Unlike much of modern BNP that is based on increasing large-scale and complex models over uncertain structures, BNP is based on the purist BNP view of partial specification of priors: this addresses the question of what classes of Bayesian models are consistent with a given, limited partial specification. The extension of some early theory of "Bayesian agent opinion analysis" leads to BPS, and the recent work in this area shows the significant potential for resulting methodology. One interest at this conference is connecting with the broader BNP community with a view to potentially stimulating new research in foundational and theoretical aspects of BPS, and connecting back to earlier frameworks of Bayesian nonparametrics based on incomplete model/prior specifications.

Dualities and genealogies for time-dependent nonparametric models Th IS

Dario Spanó, University of Warwick, United Kingdom

In many statistical problems the parameter of interest depends on a covariate which we can conveniently interpret as time. From a Bayesian perspective, it is important to be able to model tractable and interpretable prior distributions on "time"-dependent parameters to capture heterogeneity in the data. I will illustrate how genealogical processes, extending popular models of population genetics models such as Kingman's coalescent and related combinatorial processes, can be used to generate continous-time-dependent variants of some known Bayesian nonparametric priors. Various notions of duality are useful to provide insight on the so-called "borrowing strength" properties of the model.

Models for related probability measures on nonstandard domains Th IS

Alejandro Jara, Pontificia Universidad Catolica de Chile, Chile

Bayesian nonparametric models for probability distributions have focussed on $\mathbb{R}^d$ or subsets of it. In some practical situations, however, the support of the response or parameter of interest can be better characterized by a non Euclidean space. An important example is Kendall's shape space, which can be viewed as the quotient of a Riemannian manifold, and arises in different application areas, including morphometry, archeology and genetics. In these contexts, to consider standard statistical procedures that do not take into account the geometrical properties of the underlying spaces can lead to wrong inferences, which explains the increasing interest in the development of statistical models for more general spaces.

To date, the development of statistical procedures for non Euclidean spaces has focussed on the problem of mean estimation, density estimation and on the regression problem for Euclidean responses based on non Euclidean predictors. In this talk, I will consider generalizations of dependent Dirichlet process models and extensions, originally defined on Euclidean spaces, to more general Polish spaces. Theoretical properties of the proposals, such as support, continuity and consistency of the posterior distribution, will be studied in detail. The support of the process and the asymptotic behaviour of the posterior distribution will be studied using generalizations of the standard topologies for spaces of probability measures.

Nonstandard Nonparametrics Th IS

Daniel Roy, University of Toronto and Vector Institute, Ontario, Canada
Co-author(s): Haosui Duanmu

Nonstandard analysis and Loeb measure theory allow one to build finite discrete random structures that correspond exactly to continuum-sized structure. This step replaces measure theory with elementary probability. I will demonstrate this approach with two examples: a new complete class theorem in statistical decision theory and some Bayes calculations with completely random measures

Bayesian aspects of preferential attachment networks Th IS

Peter Orbanz, Columbia University, USA
Co-author(s): A. Benjamin Bloem-Reddy

A considerable fraction of the network analysis literature studies variants of preferential attachment models. In statistical language, these are models of network data defined by a process that generates a graph by attaching new edges preferredly to those vertices that are already highly connected. The importance of these models is that they explain a global property observed in data (a heavy-tailed distribution of the vertex degrees, or 'power law') by a local mechanism (the way individual edges attach). We now understand that there are close connections between power law graphs and Bayesian nonparametric models (in particular the Dirichlet and Pitman-Yor process, and also Doksum's neutral-to-the-right processes). I will explain what we know about these relationships, how we can define priors for preferential attachment networks, and how stick-breaking works in such models.

A Bayesian Nonparametric View on Count-Min Sketch Th IS

Diana Cai, Princeton University, USA
Co-author(s): Michael Mitzenmacher and Ryan P. Adams

The count-min sketch is a time- and memory-efficient randomized data structure that provides a point estimate of the number of times an item has appeared in a data stream. The count-min sketch and related hash-based data structures are ubiquitous in systems that must track frequencies of data such as URLs, IP addresses, and language n-grams. We present a Bayesian view on the count-min sketch, using the same data structure, but providing a posterior distribution over the frequencies that characterizes the uncertainty arising from the hash-based approximation. In particular, we take a nonparametric approach and consider tokens generated from a Dirichlet process (DP) random measure, which allows for an unbounded number of unique tokens. Using properties of the DP, we show that we can straightforwardly compute posterior marginals of the unknown true counts and that the modes of these marginals recover the classical count-min sketch estimator, inheriting the associated probabilistic guarantees. Additionally, the posterior distribution leads to natural shrinkage estimators for the count. Using sketches constructed from simulated data and a text data stream, we investigate the properties of the inferred posterior distributions and their respective estimators. Lastly, we study a modified problem in which the data stream consists of collections of tokens (i.e., documents) arising from a stable beta process, which allows for power law scaling behavior in the number of unique tokens.

A Berry-Esseen theorem for Pitman's α-diversity Th CS

Emanuele Dolera, University of Pavia, Italy
Co-author(s): Stefano Favaro

This talk is concerned with the study of the random variable K_n denoting the number of distinct elements in a random sample $(X_1, \ldots, X_n)$ of exchangeable random variables driven by the two parameter Poisson-Dirichlet distribution, denoted by PD(α, θ). The interest in this kind of study rests on the fact that the PD(α, θ) distribution—as well as the related Ewens-Pitman sampling formula—plays a fundamental role in a variety of research areas, such as population genetics, Bayesian nonparametrics, statistical machine learning, excursion theory, combinatorics and statistical physics. There have been several studies on the large n asymptotic behaviour of K_n. For $\alpha = 0$ and $\theta > 0$ (reducing to the Dirichlet process), one has $K_n = \sum_{1 \le i \le n} Z_i$ where the Z_i's are independent random variables with $Z_i \sim Bern(\theta/(\theta + i - 1))$, for $i = 1, \ldots, n$. Hence, in [**?**] it is shown that $K_n / \log(n)$ converges almost surely to θ as $n \to +\infty$. Moreover, it follows from Lindberg's theorem that $(K_n - \theta \log(n))/\sqrt{\theta \log(n)}$ converges in distribution to a standard Gaussian random variable as $n \to +\infty$. See [**?**]. For $\alpha \in (0, 1)$, which is the case dealt with in the talk, the large n Gaussian limit for K_n no longer holds. In particular, Theorem 3.8 in Pitman (2006) exploits the martingale convergence theorem to show that

$$\frac{K_n}{n^\alpha} \xrightarrow{\text{a.s.}} S_{\alpha,\theta},$$

as $n \to +\infty$, $S_{\alpha,\theta}$ being a random variable distributed according to scaled Mittag-Leffler distribution, whose density reads $f_{S_{\alpha,\theta}}(s) = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)} s^{\frac{\theta-1}{\alpha}-1} f_\alpha(s^{-1/\alpha}) 1\{s > 0\}$ for $s > 0$, where f_α denotes the density function of the positive α-stable random variable. Then, our main result states that

$$\sup_{x \ge 0} \left| P\left[\frac{K_n}{n^\alpha} \le x\right] - P[S_{\alpha,\theta} \le x] \right| \le \frac{C(\alpha, \theta)}{n^\alpha}$$

holds with an explicit constant $C(\alpha, \theta)$. The key ingredients of the proof are a novel probabilistic representation of K_n as compound distribution and new, refined versions of certain quantitative bounds for the Poisson approximation and the compound Poisson distribution, originally due to Hwang (1999) and Adell & de la Cal (1993), respectively. Finally, we present an application of () in the context of Bayesian nonparametric inference for species sampling problems. Consider a population $(X_i)_{i \ge 1}$ of individuals belonging to an infinite number of species with unknown proportions. Given an initial (observable) random sample $(X_1, \ldots, X_n)$ from the population, a classical species sampling problem consists in the estimation of the number $K_m^{(n)}$ of hitherto unseen species that would be observed in m

additional (unobservable) samples. In the approach proposed in (Favaro et al, 2019), $(X_1, \ldots, X_n)$ is a random sample from PD(α, θ) featuring $K_n = j \leq n$ species (blocks), for which there holds

$$\frac{K_m^{(n)}}{m^\alpha} \,|\, (X_1, \ldots, X_n) \xrightarrow{\text{a.s.}} S_{\alpha,\theta}(n, j)$$

as $m \to +\infty$, where $S_{\alpha,\theta}(n, j)$ is related to $S_{\alpha,\theta+n}$. The random variable $S_{\alpha,\theta}(n, j)$ is referred to as *Pitman's posterior α-diversity*. The importance of () is motivated by the fact that the computational burden for evaluating the posterior distribution of $K_m^{(n)}$ becomes overwhelming for large m. Then Pitman's posterior α-diversity has been extensively applied to obtain large m approximated posterior inferences for $K_m^{(n)}$ via Monte Carlo sampling from $S_{\alpha,\theta}(n, j)$. In this talk, we show how to combine () with a new Berry-Esseen bound for de Finetti's theorem recently obtained to obtain a Berry-Esseen theorem for Pitman's posterior α-diversity, thus quantifying the error of approximation in replacing the posterior distribution of $K_m^{(n)}$ with Pitman's posterior α-diversity.

Density Estimation with Mixture of Spherelets Th CS

Didong Li, Duke University, USA
Co-author(s): Minerva Mukhopadhyay and David B Dunson

Data lying in a high dimensional ambient space are commonly thought to have a much lower intrinsic dimension. In particular, the data may be concentrated near a lower-dimensional subspace or manifold. There is an immense literature focused on approximating the unknown subspace and the unknown density, and in exploiting such approximations in clustering, data compression, and building of predictive models. Most of the literature relies on approximating subspaces and densities using a locally linear, and potentially multiscale, dictionary with Gaussian kernels. In this tall, we propose a simple and general alternative, which instead uses pieces of spheres, or spherelets, and the von Mises-Fisher kernel, to locally approximate the unknown subspace and density. Theory is developed showing that spherelets can produce lower covering numbers and MSEs for many manifolds, as well as the posterior consistency of the mixture model. We develop spherical principal components analysis (SPCA). Results relative to state-of-the-art competitors show gains in ability to accurately approximate the subspace and the density with fewer components and parameters. The methods are illustrated with standard toy manifold learning examples, and applications to multiple real data sets.

Adaptive Bayesian Estimation of Mixed Discrete-Continuous Distributions under Smoothness and Sparsity Th CS

Andriy Norets, Brown University, USA
Co-author(s): J. Pelenis

We consider nonparametric estimation of a mixed discrete-continuous distribution under anisotropic smoothness conditions and possibly increasing number of support points for the discrete part of the distribution. For these settings, we derive lower bounds on the estimation rates in the total variation distance. Next, we consider a nonparametric mixture of normals model that uses continuous latent variables for the discrete part of the observations. We show that the posterior in this model contracts at rates that are equal to the derived lower bounds up to a log factor. Thus, Bayesian mixture of normals models can be used for optimal adaptive estimation of mixed discrete-continuous distributions.

Asymptotic analysis of the posterior distribution in the Caron and Fox model Th CS

Zacharie Naulet, University of Toronto, Canada
Co-author(s): Judith Rousseau and François Caron

Sparse exchangeable graphs (Caron & Fox, 2017; Veitch & Roy, 2015) resolve some pathologies in traditional random graph models, notably the inability to model sparsity, while preserving some notion of exchangeability. Only a few, however, is known about the theoretical analysis of the posterior distributions. Here we consider the model of Caron & Fox (2017) and we analyse the posterior distribution in the limit of large observed graphs. The analysis is carried both under well-specificiation and misspecification of the model. In both situations we find that the posterior concentrates and we characterize its limiting shape.

Cross-study Bayesian Factor Regression Analysis in High-dimensional Biological Data Th CS

Alejandra Avalos-Pacheco, Princeton University, USA
Co-author(s): Roberta De Vito, Barbara Engelhardt and David Rossell

Analyses that integrate multiple, somewhat diverse studies, are crucial to understand and gain knowledge in high-dimensional statistical research. When considering multiple studies, some measurements reappear across them, hence common biological features are likely to be shared (Garrett-Mayer et al., 2007). However, high throughput experiments display both artifactual and biological sources of variation (Irizarry et al., 2003). The Multi-study Factor model (De Vito et al., 2018b) is able to handle multiple studies simultaneously allowing for the joint analysis of multiple high-throughput experiments, simultaneously achieving two goals: a) to capture common component(s) across studies and b) to isolate the sources of variation that are unique of each study. In this work, we generalize the Bayesian Multi-study Factor model (De Vito et al., 2018a) by adopting a latent factor regression approach (Avalos-Pacheco et al., 2018). This generalization will allow us to obtain a covariance structure that models the study/batch specific covariances in addition to the common component, keeping track of the observed variables, such as the demographic information. The method is a cross-study Bayesian regression factor analysis with four important advantages: 1. the ability to learn the latent cardinality over the share loading matrix, avoiding the pre-specification of it; 2. sparse study specific loadings due to a continuous component spike and slab prior (local and non-local); 3. a user-defined prior dispersion for the regression coefficient accounting for population structure and other demographic characteristics for the subjects; 4. a computationally efficient algorithm, based on a fast dynamic EM algorithm (Roekovaand George, 2016). We assess the operating characteristics of our method by means of simulation studies, comparison with previous methods, and we present an application to the prediction of the biological signal from seven gene expression studies on breast cancer.

Reconstructing probabilistic trees of cellular differentiation from single-cell RNA-seq data Th CS

Miriam Shiffman, MIT & Broad Institute, USA
Co-author(s): W. Stephenson, G. Schiebinger, J. Huggins, T. Campbell, A. Regev and T. Broderick

Until recently, transcriptomics was limited to bulk RNA sequencing, obscuring the underlying expression patterns of individual cells in favor of a global average. Thanks to technological advances, we can now profile gene expression across thousands or millions of individual cells in parallel. This new data regime has led to the intriguing discovery that individual cell profiles can reflect the imprint of time or dynamic processes. However, synthesizing this information to reconstruct dynamic biological phenomena — from data that are noisy, heterogenous, and sparse, and from processes that may unfold asynchronously — poses a computational and statistical challenge.

We develop a full generative model and inference for reconstructing a dynamic process (cellular differentiation) from many static snapshots (single-cell RNA-seq profiles), with calibrated uncertainties.

Specifically, we define cell state by the latent parameterization of a distribution over gene expression space, and model these latent vectors as arising from bifurcating, self-reinforcing paths along a probabilistic tree. Motivated by the biology, this approach necessitated the design of a new class of Bayesian tree models for data that arise from a latent branching spectrum. In particular, we extend the framework of the classical Dirichlet diffusion tree to simultaneously infer branch topology and latent cell states along continuous trajectories over the full tree. In tandem, we construct a novel Markov chain Monte Carlo sampler that interleaves Metropolis-Hastings and message passing (via a variable augmentation trick) to leverage model structure for efficient inference. Finally, we demonstrate that these techniques can recover latent trajectories from simulated single-cell transcriptomes. While this work is motivated by cellular differentiation, our model provides flexible densities for any data that arise from continuous evolution along a latent nonparametric tree.

Astronomical source detection and background separation via hierarchical Bayesian nonparametric mixtures Th CS

Andrea Sottosanti, University of Padova, Italy
Co-author(s): M. Bernardi, A. R. Brazzale, R. Trotta and D. van Dyk

We propose an innovative approach based on Bayesian nonparametric methods to the signal extraction of astronomical sources in gamma-ray count maps under the presence of a strong background contamination. Our model simultaneously induces clustering on the photons using their spatial information and gives an estimate of the number of sources, while separating them from the irregular signal of the background component that extends over the entire map. From a statistical perspective, the signal of the sources is modeled using a Dirichlet Process mixture, that allows to discover and locate a possible infinite number of clusters, while the background component is completely reconstructed using a new flexible Bayesian nonparametric model based on b-spline basis functions. The resultant can be then thought of as a hierarchical mixture of nonparametric mixtures for flexible clustering of highly contaminated signals. We provide also a Markov chain Monte Carlo algorithm to infer on the posterior distribution of the model parameters which does not require any tuning parameter, and a suitable post-processing algorithm to quantify the information coming from the detected clusters. Results on different datasets confirm the capacity of the model to discover and locate the sources in the analysed map, to quantify their intensities and to estimate and account for the presence of the background contamination.

Robust, Nonparametric Manifold Learning for Single Cell RNA Sequencing Th CS

Archit Verma, Princeton University, USA
Co-author(s): B. Engelhardt

Computational tools for analyzing single cell RNA-sequencing data, or count matrices of cells by genes, use dimension reduction to project data to a lower dimensional manifold capturing the complexity of gene expression patterns across 20,000 genes in a compact space. Dimension reduction techniques have been used for noise reduction, population identification, visualization, pseudotemporal ordering, and imputation in single cell RNA-seq data. Current methods, however, are poorly adapted for nonlinear, non-smooth gene expression behavior and high levels of technical and biological noise. We introduce a robust Bayesian nonparametric manifold learning model, the t-distributed Gaussian process latent variable model (tGPLVM), to effectively represent scRNA-seq count matrices in a low-dimensional space. The expression of each gene across cells is modeled as a noisy Gaussian process of latent variables, as in the Bayesian Gaussian process latent variable model (Titsias, 2010). We extend the traditional GPLVM to robustly learn latent embeddings by modifying the kernel form and noise model. The kernel covariance function is a weighted sum of Gaussian and Matern kernels to

flexibly capture the possibly non-smooth geometry of the manifold in high-dimensional space. We model residual errors with a heavy-tailed Student's t-distribution to estimate a manifold that is robust to technical and biological noise. Automatic Relevance Determination kernel structure is used to estimate the latent dimensionality of the manifold. We efficiently fit this model using black box variational inference (Ranganathan, 2014). We illustrate the applicability of our manifold embeddings to several single-cell tasks. On labeled cerebral cortex and blood cells (Pollen, 2014) we find that our latent representations demonstrate excellent separation of distinct cell populations, superior to many existing methods for clusering. On time series measurements of infected mouse T cells (Lonnberg, 2017) we find that our model accurately reconstructs the temporal structure of the data. We show that tGPLVM can be applied to data sets without data filtering or normalization by finding biologically meaningful representations of CD34+ peripheral blood mononuclear cells. We demonstrate scalability by fitting the model to data from 1 million mice neural cells (Zheng, 2017).

Friday 28 June

Bayesian Pseudo Posterior Synthesis for Data Privacy Protection Fr IS

Terrance Savitsky, U.S. Bureau of Labor Statistics, USA
Co-author(s): M. Hu and M. Williams

Statistical agencies utilize models to synthesize respondent-level data for release to the general public as an alternative to the actual data records. A Bayesian model synthesizer encodes privacy protection by employing a hierarchical prior construction that induces smoothing of the real data distribution. Synthetic respondent-level data records are often preferred to summary data tables due to the many possible uses by researchers and data analysts. Agencies balance a trade-off between utility of the synthetic data versus disclosure risks and hold a specific target threshold for disclosure risk before releasing synthetic datasets. We introduce a pseudo posterior likelihood that exponentiates each contribution by an observation record-indexed weight $\in (0, 1)$, defined to be inversely proportional to the disclosure risk for that record in the synthetic data. Our use of a vector of weights allows more precise downweighting of high risk records in a fashion that better preserves utility as compared with using a scalar weight. We illustrate our method with a simulation study and an application to the Consumer Expenditure Survey of the U.S. Bureau of Labor Statistics. We demonstrate how the frequentist consistency and uncertainty quantification are affected by the inverse risk-weighting.

Nonparametric mixture modeling on constrained spaces Fr IS

Vinayak Rao , Purdue University, USA
Co-author(s): Putu Ayu Sudyanti

We consider flexibly modeling multimodal data that lie on constrained spaces. Applications include climate or crime measurements in a geographical area, or flow-cytometry experiments, where unsuitable recordings are discarded. A simple approach to modeling such data is through the use of mixture models, with each component following an appropriate truncated distribution. Problems arise when the truncation involves complicated constraints, leading to difficulties in specifying the component distributions, and in evaluating their normalization constants. Bayesian inference over the parameters of these models results in posterior distributions that are doubly-intractable. We address this problem via an algorithm based on rejection sampling and data augmentation. We view samples from a truncated distribution as outcomes of a rejection sampling scheme, where proposals are made from a simple mixture model, and are rejected if they violate the constraints. Our scheme proceeds by imputing the rejected samples given mixture parameters, and then resampling parameters given all samples. We study two modeling approaches: mixtures of truncated components and truncated mixtures of

components. In both situations, we describe exact Markov chain Monte Carlo sampling algorithms, as well as approximations that bound the number of rejected samples, achieving computational efficiency and lower variance at the cost of asymptotic bias. We present results on simulated data and apply our algorithm to modeling crime recorded in the city of Chicago.

Bayesian nonparametric methods for analysing macroeconomic time series Fr IS

Maria Kalli, University of Kent, UK
Co-author(s): Jim Griffin, University College London (UCL) , UK

The analysis of macroeconomic time series often involves the use of a vector autoregressive (VAR) model. VAR models provide a framework for the analysis of the complex joint dynamics present between macroeconomic series, but they have been criticised for their unrealistic assumptions (linearity, homoscedasticity, Gaussianity). In this talk we are going to describe how Bayesian non-parametric methods can be used to directly model the stationary and transition densities of such a multivariate system. This approach allows for nonlinearity in the conditional mean, heteroscedasticity in the conditional variance, and non-Gaussian innovations. It can also allow for non-stationary. Our empirical applications lie within the study of monetary policy and macro financial linkages within the aggregate economy. We find that the Bayesian nonparametric VAR (BayesNP-VAR) model predictively outperforms competing models.

Posterior contraction of parameters and interpretability in Bayesian mixture modeling Fr KL

Long Nguyen, University of Michigan, USA
Co-author(s): Aritra Guha and Nhat Ho

We study posterior contraction behaviors for parameters of interest in the context of Bayesian mixture modeling, where the number of mixing components is unknown while the model itself may or may not be correctly specified. Two representative types of prior specification will be considered: one requires explicitly a prior distribution on the number of mixture components, while the other places a nonparametric prior on the space of mixing distributions. The former is shown to yield an optimal rate of posterior contraction on the model parameters under minimal conditions, while the latter can be utilized to consistently recover the unknown number of mixture components, with the help of a fast probabilistic post-processing procedure. We then turn the studyof these Bayesian procedures to the realistic settings of model misspecification. It will be shown that the modeling choice of kernel density functions plays perhaps the most impactful roles in determining the posterior contraction rates in the mis-specified situations. Drawing on concrete posterior contraction rates established in this paper we wish to highlight aspects about the interesting tradeoffs between model expressivenessand interpretability that a statistical modeler must negotiate in the rich world of mixture modeling. This work is joint with Aritra Guha and Nhat Ho.

Abstracts - Posters

Monday 24 June

Geometric Sensitivity Measures for Bayesian Nonparametric Density Estimation Models Mo PS

Abhijoy Saha, The Ohio State University, USA
Co-author(s): Abhijoy Saha and Sebastian Kurtek

We propose a geometric framework to assess global sensitivity in Bayesian nonparametric models for density estimation. We study the sensitivity of nonparametric Bayesian models for density estimation, based on Dirichlet-type priors, to perturbations of either the precision parameter or the base probability measure. To quantify the different effects of the perturbations of the parameters and hyperparameters in these models on the posterior, we define three geometrically-motivated global sensitivity measures. These measures are based on geodesic paths and distances computed under the nonparametric Fisher-Rao Riemannian metric on the space of densities, applied to posterior samples of densities: (1) the Fisher-Rao distance between density averages of posterior samples, (2) the log-ratio of Karcher variances of posterior samples, and (3) the norm of the difference of scaled cumulative eigenvalues of empirical covariance operators obtained from posterior samples. We validate our approach using multiple simulation studies and consider the problem of sensitivity analysis for Bayesian density estimation models in the context of three real datasets that have previously been studied.

Cross-study Bayesian Factor Regression Analysis in High-dimensional Biological Data Mo PS

Alejandra Avalos-Pacheco, Harvard Medical School, USA
Co-author(s): De Vito, R., Engelhardt, B., Rossell, D.

Analyses that integrate multiple, somewhat diverse studies, are crucial to understand and gain knowledge in high-dimensional statistical research. When considering multiple studies, some measurements reappear across them, hence common biological features are likely to be shared Garrett-Mayer et al., 2007). However, high throughput experiments display both artifactual and biological sources of variation (Irizarry et al., 2003). The Multi-study Factor model (De Vito et al., 2018b) is able to handle multiple studies simultaneously allowing for the joint analysis of multiple high-throughput experiments, simultaneously achieving two goals: a) to capture common component(s) across studies and b) to isolate the sources of variation that are unique of each study.

In this work, we generalize the Bayesian Multi-study Factor model (De Vito et al., 2018a) by adopting a latent factor regression approach (Avalos-Pacheco et al., 2018). This generalization will allow us to obtain a covariance structure that models the study/batch specific covariances in addition to the common component, keeping track of the observed variables, such as the demographic information. The method is a cross-study Bayesian regression factor analysis with four important advantages:

1. the ability to learn the latent cardinality over the share loading matrix, avoiding the pre-specification of it;

2. sparse study specific loadings due to a continuous component spike and slab prior (local and non-local);

3. a user-defined prior dispersion for the regression coefficient accounting for population structure and other demographic characteristics for the subjects;

4. a computationally efficient algorithm, based on a fast dynamic EM algorithm (Rockova and George, 2017).

We assess the operating characteristics of our method by means of simulation studies, comparison with previous methods, and we present an application to the prediction of the biological signal from seven gene expression studies on breast cancer.

Bayesian nonparametric modeling for large spatio-temporal data: an application to mobile networks Mo PS

Alessandra Guglielmi, Politecnico di Milano, Italy
Co-author(s): Annalisa Cadonna and Andrea Cremaschi

Spatio-temporal areal data can be seen as a collection of time series which are spatially correlated, according to a specific neighboring structure. We propose a hierarchical Bayesian model for spatio-temporal areal data, which allows for spatial model-based clustering using BNP. Then, we develop efficient MCMC algorithms based on numerical linear algebra, which exploit the sparse structure of the precision matrix of a spatio-temporal Gaussian Random Markov Field. Finally, we present an application to mobile data, with the goal to model, predict and spatially cluster population density dynamics.

Population Random Measure Embedding: A Genetic Method for Discrete Random Measures, via Distributional Symmetry Mo PS

Aonan Zhang, Columbia University, United States
Co-author(s): Aonan Zhang and John Paisley

Many hidden structures underlying high dimensional data can be compactly expressed by a discrete random measure $\xi_n = \sum_{k\in[K]} Z_{nk}\delta_{\theta_k}$, where $(\theta_k)_{k\in[K]} \subset \Theta$ is a collection of hidden atoms shared across observations (indexed by n). Previous Bayesian nonparametric methods focus on embedding ξ_n onto alternative spaces to resolve complex atom correlations. However, these methods can be rigid and hard to learn in practice. In this paper, we temporarily ignore the atom space Θ and embed population random measures $(\xi_n)_{n\in\mathbb{N}}$ altogether as ξ' onto an infinite strip $[0,1]\times\mathbb{R}_+$, where the order of atoms is *removed* by assuming separate exchangeability. Through a "de Finetti type" result, we can represent ξ' as a coupling of a 2d Poisson process and exchangeable random functions $(f_n)_{n\in\mathbb{N}}$, where each f_n is an object-specific atom sampling function. In this way, we transform the problem from learning complex correlations with discrete random measures into learning complex functions that can be learned with deep neural networks. In practice, we introduce an efficient amortized variational inference algorithm to learn f_n without pain; i.e., no local gradient steps are required during stochastic inference.

Nonparametric Bayesian Functional Regression with application to shot put data Mo PS

Alessandro Lanteri, University of Turin, Italy
Co-author(s): R. Argiento and S. Montagna

In sport analytics, there is often interest in predicting elite athlete's performance at a future sporting event given his/her competitive results tracked throughout the athlete's career and other (time-varying) covariates. Such predictions can be useful both for scouting purposes, and to build red flag indicators of unexpected increases in athlete performance for targeted anti-doping testing. We propose a predictive model for the longitudinal trajectory of athlete's performance where we characterize the curve with a sparse basis expansion allowing individual time-dependant covariates to impact the shape of the estimated trajectories. Moreover, we introduce random intercepts, distributed according to a nonparametric hierarchical process, in order to induce clustering while borrowing statistical information across curves. In particular, we assume a hierarchical normalized generalized gamma process to grants great flexibility in clustering and accuracy in prediction. We apply our model to a longitudinal study on shot put athletes, where their competitive results are tracked throughout their career.

Optimize, Learn, Sample Mo PS

Alfredo Garbuno-Inigo, Caltech, US
Co-author(s): E. Cleary, T. Schneider, and A. Stuart

The calibration of complex models to data is both a challenge and an opportunity. It can be posed as an Inverse Problem. This work focuses on the interface of Ensemble Kalman Inversion (EKI), Gaussian process emulation (GPE) and Monte Carlo-Markov Chain (MCMC) for the calibration of, and quantification of uncertainty in, parameters learned from data. The goal is to perform uncertainty quantification in predictions made from complex models, reflecting uncertainty in these parameters, with relatively few for-ward model evaluations. This is achieved by propagating approximate posterior samples obtained by judicious combination of ideas from EKI, GPE and MCMC. The strategy will be illustrated with idealized models related to climate modeling.

Investigating a Bayesian semi-parametric model for the study of synergistic interaction effects in in-vitro drug combination experiments Mo PS

Andrea Cremaschi, Universitetet i Oslo, Norway
Co-author(s): L. Rønneberg, K. Taskén, M. Zucknick.

The study of the effect of cancer therapies is often tested pre-clinically via *in-vitro* experiments, where the viability of the cell population is measured through cell counts. In this way, large libraries of compounds can be tested, comparing the efficacy of the drugs in fighting the malignancy. A similar approach is used to test combinations of drugs and study their efficacy, as well as their interactions. Drug-drug interaction studies focus on the quantification of the additional effect encountered when two drugs are combined, as opposed to using the treatments separately. With this aim, we propose a probabilistic model for the description of a drug combination experiment, and model the interaction surface using flexible Bayesian approaches. In particular, in this work we study a Bayesian model based on the use of Gaussian processes for the description of the drug-drug interaction behaviour. We apply the model to a simulated dataset, as well as to a yet unpublished ovarian cancer dataset.

On posterior contraction of parameters and interpretability in Bayesian mixture modeling Mo PS

Aritra Guha, University of Michigan, USA
Co-author(s): A. Nhat Ho and B. XuanLong Nguyen

We study posterior contraction behaviors for parameters of interest in the context of Bayesian mixture modeling, where the number of mixing components is unknown while the model itself may or may

not be correctly specified. Two representative types of prior specification will be considered: one requires explicitly a prior distribution on the number of mixture components, while the other places a nonparametric prior on the space of mixing distributions. The former is shown to yield an optimal rate of posterior contraction on the model parameters under minimal conditions, while the latter can be utilized to consistently recover the unknown number of mixture components, with the help of a fast probabilistic post-processing procedure. We then turn the study of these Bayesian procedures to the realistic settings of model misspecification. It will be shown that the modeling choice of kernel density functions plays perhaps the most impactful roles in determining the posterior contraction rates in the misspecified situations. Drawing on concrete posterior contraction rates established in this paper we wish to highlight some aspects about the interesting tradeoffs between model expressiveness and interpretability that a statistical modeler must negotiate in the rich world of mixture modeling.

Closed Form Bayesian Filtering for Multivariate Binary Time Series Mo PS

Augusto Fasano, Bocconi University, Milan, Italy
Co-author(s): G. Rebaudo, D. Durante and S. Petrone

Non-Gaussian state-space models arise routinely in several applications. Within this framework, the binary time series setting provides a source of constant interest due to its relevance in a variety of studies. However, unlike Gaussian state-space models—where the classical Kalman filter allows to sequentially update the filtering and predictive distributions—binary state-space models require either approximations or sequential Monte Carlo strategies for dynamic uncertainty quantification and prediction. This is due to the apparent absence of conjugacy between the Gaussian random states and the probit or logistic likelihood induced by the observation equation for the binary data. In this work we prove that, when the focus is on flexible Bayesian learning of dynamic probit models monitored at, possibly, infinite times, filtering and predictive distributions belong to the class of unified skew-normal variables and, moreover, the corresponding parameters can be sequentially updated online via tractable expressions. This result allows to develop an exact Kalman filter for online learning of univariate and multivariate binary time series, which provides also methods to draw independent and identically distributed samples from the exact filtering and predictive distributions of the random states, thereby improving Monte Carlo inference. As outlined in an illustrative application, the proposed method improves state-of-the-art strategies routinely used in the literature. A scalable and optimal sequential Monte Carlo, which exploits the unified skew-normal properties, is also developed and additional exact expressions for the smoothing distribution are provided.

A nonparametric Bayesian prediction approach for modelling small data Mo PS

Azizur Rahman, Charles Sturt University, Australia

In the 21st century's big data era, the small data snags still exist in many areas which causes modelling and evaluation are hard to make. Small dataset is insufficient to generate a reliable prediction especially in the small area estimation domain. This paper presents a Bayesian nonparametric prediction framework to models which are dealing with smaller data sets and/or give poor predictions. This approach uses a robust Gaussian model in weight-space notion and drive the prediction distribution of the future responses. Results revealed that the prediction distribution of a set of futures responses is conditional on a set of observed data and depends on the degree of the spline. It also provides an empirical illustration and demonstrated that the prediction outcomes depend on the realised responses only through the observations in design matrix and the sample residual sum of squares and products matrices with the Kronecker product.

Poisson Process Radial Basis Bayesian Neural Networks Mo PS

Beau Coker, Harvard University, USA
Co-author(s): Melanie Fernandez Pradier and Finale Doshi-Velez

Bayesian Neural Networks (BNNs) are flexible function priors well-suited to situations in which data are scarce and uncertainty must be quantified. Yet, BNNs only benefit partially from the Bayesian framework: the architecture is most often fixed, limiting the Bayesian benefit of model selection via integration; and common weight priors encode very little a priori functional knowledge. In this paper, we present a neural network function prior that both adapts the number of hidden units automatically and can incorporate functional properties such as (non)-stationarity. Specifically, we place a non-homogeneous Poisson process prior over location parameters of the activations of a neural network. We show that our model is consistent as the number of observations goes to infinity. We demonstrate our model's ability to infer input-dependent architectures and maintain desired properties through a variety of examples.

Models for Networks with Core-Periphery Structure Mo PS

Cian Naik, University of Oxford, United Kingdom
Co-author(s): François Caron and Judith Rousseau

We propose a statistical model for sparse networks with a core-periphery structure. This model is based on an exchangeable point process representation of a graph, using a construction that builds on vectors of completely random measures. The model uses a discrete parameter to distinguish nodes that are in the core and the periphery of a network. We develop methods for simulating this class of models, and to perform posterior inference. We explore core-periphery structure in range of simulated and real datasets.

Dependent Random Measures Indexed by a Functional Covariate Mo PS

Emmanuel Bernieri, University of Edinburgh, Scotland
Co-author(s): Emmanuel Bernieri and Miguel de Carvalho

We explore the analysis of functional data in a nonparametric Bayesian context. Specifically, we devise priors in the space of all conditional distributions, for the setting where the interest is on conditioning on a sophisticated object—such as a random function. The proposed model can be regarded as an infinite mixture of functional linear regression models. A specific version of the proposed model is explored in detail, which consists of a Dependent Dirichlet Process (DDP) whose regression functions include inner products between a functional covariate and coefficient function. We illustrate the proposed methods using simulated and real data, and evaluate the accuracy of the methods through a simulation study.

Generalised Polya urn for a class of dependent Dirichlet Processes Mo PS

Filippo Ascolani, University of Torino and Collegio Carlo Alberto, Italy
Co-author(s): A. Lijoi and M. Ruggiero

We consider the problem of learning, from indirect observations, a class of time-dependent Dirichlet processes driven by a Fleming-Viot model, which produces a diffusion whose states are atomic probability measures and whose stationary measure is a Dirichlet prior. We assume the data are random samples from the process state at discrete times, so at each collection time the sampling process is analogous to a classical Bayesian nonparametric setting. We build on previous work on the time-marginal posteriors for this family of models, which belong to the class of finite mixtures

of Dirichlet processes in the sense of Antoniak (1974), and investigate the predictive distribution of the observations given past data. We identify such distribution as a generalised Polya urn with baseline measure appropriately updated with past data, thus resulting in a convex combination of the prior baseline measure and of the empirical measures of past data and that of present data. We characterise the mixing weights, which are time-varying, and investigate their asymptotic properties, which in appropriate limits recover the classical Polya urn for the Dirichlet process or the de Finetti measure at present time. As a by product of our result, we compute the time-dependent Exchangeable Partition Probability Function for this model, and we lay out an explicit algorithm for sampling exactly from a dependent DP given past data.

Sparse Spatial Random Graphs Mo PS

Francesca Panero, University of Oxford, United Kingdom
Co-author(s): F. Caron and J. Rousseau

A number of real-world spatial networks empirically show a double power-law degree distribution: for low degrees k we have $p_k \simeq k^{-a}$, $a \in [1, 2]$, while for high degrees k $p_k \simeq k^{-b}$, $b > 2$. In this poster we present our novel work on sparse spatial random graphs built on exchangeable random measures. This approach generalizes a number of models, including hyperbolic random graphs, and is able to describe desirable networks characteristics, like double power-law degree distribution, non vanishing clustering coefficient and different levels of sparsity.

Modeling Human Microbiome Data via Latent Nested Nonparametric Priors Mo PS

Francesco Denti, University of Milan-Bicocca, Italy and Università della Svizzera italiana, Switzerland
Co-author(s): M. Guindani, F. Camerlenghi and A. Mira

The study of the human microbiome has gained substantial attention in recent years due to its relevance to the regulation of the autoimmune system. The data comprise counts of Operational Taxonomic Units (OTUs), representing microbes characterized by almost identical genome sequences. Since OTU abundances vary widely across individuals, it is of interest to characterize the diversity of the microbiome within a population and identify groups of subjects with similar microbial distributions. Here, we propose a Bayesian Nonparametric approach to model the heterogeneity of the observed OTU abundances, by regarding them as partially exchangeable across subjects. More specifically, we first express the likelihood as a rounded mixture of Gaussian kernels, to take into account both the discreteness and the sparsity of the data. The individual abundances are then clustered according to a shared nested structure: a common set of parameters (atoms at the observational level) is used to construct, at a higher level, a set of distributional atoms. The proposed "common atoms" model avoids the potential degeneracy of distributional clusters to a fully exchangeable case – a drawback of the traditional Nested Dirichlet process as outlined by Camerlenghi et al. (2018) – while retaining the flexibility and the clustering properties of the nonparametric nested framework. To perform posterior inference, we propose a novel independent slice-efficient algorithm suited for the Nested case. We evaluate the performances of the proposed common atoms nested model via simulation studies and then illustrate its application to a case-controlled study on post-diarrheal disruption in children from Bangladesh.

Joint Species Distribution Modelling: Dimension reduction using Bayesian nonparametric priors Mo PS

Giovanni Poggiato, Inria(Grenoble INP), France,
Co-author(s): Daria Bystrova, Julyan Arbel and Wilfried Thuiller

The modelling of species distribution plays an important role in both theoretical and applied ecology: given a set of species occurrence, the aim is to infer species spatial distribution over a given territory. Joint Species Distribution models (JSDM) study joint species occurrences or abundances at different locations. In the Bayesian framework, this is often done by modelling a continuous latent variable in a hierarchical generalised mixed linear model framework. The regression term models the effect of the environmental conditions on the species, to account for their habitat, that is known to be one of the main drivers of species distributions. Another important driver, the interactions between species, is modelled as the covariance structure of the residuals of the regression. However, this model suffers from the curse of dimensionality because of the estimation of the covariance matrix. Thus, there is a need for dimension reduction. In the literature, this problem is often tackled with the use of latent factors. Moreover, [Taylor-Rodriguez et al., 2017] add Dirichlet processes (DP) to cluster and further reduce the effective dimension on the rows of the matrix that represent the random effect induced by the covariance matrix. Our first extension is the use of Pitman–Yor (PY) process prior as an alternative for DP prior, which has more flexible clustering properties. We approximated this process using the ϵ-PY suggested by [Arbel et al.,2019]. Furthermore, a hierarchical layer is added by putting a prior distribution on the discount and the precision parameters of the PY process, and by fixing their related hyperparameters, to let the expected prior number of clusters match an ecological prior knowledge on the number of clusters.
References:
[Arbel et al.,2019] Arbel,J.,De Blasi,P., and Prünster, I. (2019). Stochastic approximations to the Pitman–Yor process. Bayesian Analysis.
[Taylor-Rodriguez et al., 2017] Taylor-Rodriguez, D., Kaufeld, K. A., Schliep, E. M., Clark, J. S., and Gelfand, A. E. (2017). Joint species distribution modeling: Dimension reduction using dirichlet processes. Bayesian Analysis.

Hybrid BNP priors for clustering Mo PS

Giovanni Rebaudo, Bocconi University, Italy
Co-author(s): A. Lijoi, I. Prünster

Bayesian statistics has been proven effective in borrowing information between different groups or studies. Subjects in different studies may share the same unknown distribution. A well-known BNP prior to flexibly cluster probability distributions for the partially exchangeable case is the nested Dirichlet process (NDP) which is known to degenerate to the fully exchangeable case when there are ties across samples. Generalizations of the NDP have been proposed to address this issue. We propose a novel hybrid nonparametric prior which solves the problem by combining two different discrete nonparametric random structures. We derive a closed form expression of the induced random partition distribution which allows to gain a deeper insight on the theoretical properties of the model and, further, yields a MCMC algorithm for evaluating Bayesian inference of interest. Finally a BNP test of homogeneity between different groups will be displayed and it complemented by illustrative examples.

Non-echangeable prior for feature models, a generalization of the three-parameter Indian Buffet Process Mo PS

Giuseppe Di Benedetto, University of Oxford, United Kingdom
Co-author(s): François Caron and Yee Whye Teh

Bayesian nonparametric models for sparse binary matrices have been adopted in many areas such as machine learning, networks modeling and population genetics. In feature modeling each row of the matrix represents an object with features represented by non-zero entries. Exchangeability of the observations is often assumed for tractability, however this implies that the number of objects

sharing a particular feature grows linearly with the number of observations. This asymptotic property might not be suitable for applications where the observations are ordered and features are assumed to grow slowly. We propose a non-exchangeable prior on sparse binary matrices that allows to model the sublinear rate for the growth of the number of objects sharing a feature, and capture the power-law behaviour of the number of features shared by a given number of objects.

Some developments of the generalized species sampling sequences Mo PS

Hristo Sariev, Università Commerciale Luigi Bocconi, Italy
Co-author(s): S. Fortini and S. Petrone

The species sampling sequences of Pitman (1996) form an important class of exchangeable stochastic processes with an explicit sampling scheme that encompasses some of the most well-known Bayesian nonparametric priors. In a situation of competition, selection, covariates dependence and/or other forms of non-stationary behavior, however, the assumption of exchangeability is easily violated, so these models become inadequate. To that end, Bassetti et al. (2010) have proposed a generalized version of the basic species sampling sequence by introducing further randomization into the sampling scheme. The resulting class of models satisfies the weaker assumption of conditional identity in distribution (Berti et al., 2014) and as such describes processes that are asymptotically exchangeable in the sense of Aldous (1985). In this study we show that a generalized species sampling sequence is in fact asymptotically an exchangeable species sampling sequence, so that the added randomization becomes absorbed into the directing measure. In addition, we provide results about the random partition generated by a certain subclass of generalized species sampling models in the form of distributional results about the number of distinct species.

Nonparametric temporal sequence alignment Mo PS

Ieva Kazlauskaite, University of Bath, UK
Co-author(s): A. Ivan Ustyuzhaninov and B. Carl Henrik Ek and C. Neill D. F. Campbell

We consider the problem of aligning temporally warped noisy sequences. In this set up, we encounter the following challenges: the data comes from an unknown number of distinct latent functions (or equivalently, there is an unknown number of groups of sequences), the sequences are of different lengths (or are sampled at different rates), and we make only weak assumptions on the temporal warps (such as smoothness and monotonicity). Addressing these issues with standard parametric approaches is challenging as they have limited capacity and require manual selection of the basis functions for the warps as well as for the latent functions. A different challenge appears when using non-parametric models, as we need to describe dependencies both within and between sequences. This correlation structure is complex to formulate and it poses further challenges as it is described as a joint stochastic process whose marginal properties no longer match the within nor the between sequence models. We propose a completely non-parametric model which adapts the complexity of the warps and the latent functions to the data, allowing us to use the same approach for a variety of problems. Furthermore, it enables automatic inference of the number of distinct latent functions. The proposed non-marginal procedure contains two parts: (1) we model each observed sequence independently using Gaussian processes allowing us to consider sequences of different lengths and to handle the observation noise in a principled manner, and (2) we regularize the corresponding latent functions using Bayesian non-parametric clustering to encode conditional dependencies across sequences and automatically infer the number of distinct latent functions. For part (2), we perform the clustering in one of two ways: either by considering a discrete set of latent functions using a DP prior, or as a continuous model using an additional GP. We show that our proposed approaches perform competitively on a number of benchmark data sets while also offering greater flexibility and generality in comparison to

the state-of-the-art parametric models.

This framework generalises beyond alignment of sequences. Given multiple inputs that share some structure (e.g. N misaligned sequences generated from a smaller number of latent functions), part (1) of the framework models each input in isolation using hierarchical models, and part (2) enforces that certain components of the hierarchy in these models are similar to encode our prior knowledge about the problem (e.g. warped sequences are modelled as f(g) where the true unwarped sequences f are the same for each group, and warps g are different for all of them). The exact functional forms of parts (1) and (2) depend on the problem at hand, allowing to cast a variety of problems into this setting.

Monotonic random processes Mo PS

Ivan Ustyuzhaninov, University of Tübingen, Germany
Co-author(s): A. Ieva Kazlauskaite and B. Carl Henrik Ek and C. Neill D. F. Campbell

Monotonicity is an important property of the underlying function of the observed data, therefore specifying a corresponding prior allows to reduce the space of the functions consistent with the data and thus obtain fits with lower uncertainties and more meaningful extrapolations. Moreover, monotonicity is often a desirable property at certain stages of a hierarchical model (e.g. as a warping function in a temporal alignment model). However, constructing a nonparametric probabilistic model of monotonic functions is challenging as it requires specifying a joint distribution of monotonically increasing random variables. In recent work [Andersen et al., 2018] introduced a parametric approximation to a nonparametric model of monotonic functions based on modelling a derivative of a function with a Gaussian process. We build on this work by proposing several non-parametric models of monotonic functions based on (1) a random process with non-negative independent increments on both axes [Haslett and Parnell, 2008], (2) modelling function derivatives with non-negative transformations of a Wiener process [Matsumoto and Yor, 2005], or (3) monotonic differential flows [Hedge et al., 2018]. We consider the modelling and the numerical issues related to each of these approaches, demonstrate the differences between these methods applied to regression tasks and as a first layer in hierarchical model (e.g. a deep GP), and discuss open questions and challenges related to this line of work.

Improving Inference for the Non-Stationary Contextual Bandit via Iterative Moment-Matching Algorithms Mo PS

Jack McKenzie, The University of Manchester, UK
Co-author(s): Thomas House, Neil Walton, Peter Appleby

Multi-armed bandits have long been studied to tackle the exploration-exploitation dilemma present when deciding what actions to take in an uncertain environment. Thompson sampling tackles this problem from a Bayesian perspective; it works by defining a Bayesian generative model which describes the actions and rewards received, and sequentially takes actions which maximise the expected payout based on samples generated from the posterior distribution. Chapelle et al (2011) use a Bayesian logistic regression model with Thompson sampling to decide how to display adverts to users on the world wide web. To perform inference, they resort to the well known Laplace approximation, assume stationary latent parameters and also assume independence among dimensions. In this poster, we instead use a combination of Expectation Propagation (EP) and Assumed density filtering (ADF) to sequentially approximate the non-conjugate posterior, whilst tracking any non-stationarity in the latent parameters. We then show how our method gives O(n) better accuracy that the classic Laplace approximation in the stationary enviroment and the impact this has on the regret of the algorithm. We believe our method to have tremendous practical importance due to the significance of contextual bandits in recommender systems and online advert allocation.

Posterior contraction of non-parametric Bayesian inference on non-homogeneous Poisson processes Mo PS

James Grant, Lancaster University, UK
Co-author(s): A. David Leslie

We consider the posterior contraction of non-parametric Bayesian inference on non-homogeneous Poisson processes. We consider the quality of inference on a rate function λ, given non-identically distributed realisations, whose rates are transformations of λ. Such data arises frequently in practice due, for instance, to the challenges of making observations with limited resources or the effects of weather on detectability of events. We derive contraction rates for the posterior estimates arising from the Sigmoidal Gaussian Cox Process and Quadratic Gaussian Cox Process models. These are popular models where λ is modelled as a logistic and quadratic transformation of a Gaussian Process respectively. Our work extends beyond the existing analyses by providing rates at which the posterior mass placed far from the true λ shrinks for certain finite numbers of observations.

Detection of common-variance subspace and its application to classification Mo PS

Jiae Kim, The Ohio State University, USA
Co-author(s): Steven N. MacEachern

Fisher's linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are traditional methods for classification. LDA assumes the same variance-covariance matrices for each class and results in one-dimensional linear classifier. QDA doesn't require the assumption and produces a quadratic classifier. We introduce a new classifier. We first find linear subspaces with the same variance-covariance and then detect a subspace which is the "most" efficient for classification. Resulting linear subspace can be multidimensional. We present technical details, some of which are how to transform the data, a sequential algorithm that finds the linear subspaces of the same variance-covariance and how to make a use of the efficient subspace for classification. The performance of the new classifiers is compared with state-of-art classifiers on simulated and real data.

A Generalization of Hierarchical Exchangeability on Trees to Directed Acyclic Graphs Mo PS

Jiho Lee, KAIST, South Korea
Co-author(s): Paul Jung(KAIST), Sam Staton(Oxford), Hongseok Yang(KAIST)

Motivated by problems in Bayesian nonparametrics and probabilistic programming discussed in Staton et al. (2018), we present a new kind of partial exchangeability for random arrays which we call DAG-exchangeability. In our setting, a given random array is indexed by certain subgraphs of a directed acyclic graph (DAG) of finite depth, where each nonterminal vertex has infinitely many outgoing edges. We prove a representation theorem for such arrays which generalizes the Aldous-Hoover representation theorem.

In the case that the DAGs are finite collections of certain rooted trees, our arrays are hierarchically exchangeable in the sense of Austin and Panchenko (2014), and we recover the representation theorem proved by them. Additionally, our representation is fine-grained in the sense that representations at lower levels of the hierarchy are also available. This latter feature is important in applications to probabilistic programming, thus offering an improvement over the Austin-Panchenko representation even for hierarchical exchangeability.

Bayesian semi-parametric density estimation for nonregular models Mo PS

Johan Van Der Molen Moris, University of Edinburgh, UK
Co-author(s): N. Bochkina, J. Rousseau and J. B. Salomond

We consider a Bayesian semi-parametric model for estimating a nonregular parameter which is the end point of the support of an unknown density with a discontinuity. For this purpose we consider different priors including a non-homogeneous Completely Random Measure mixture and find appropriate conditions on hyper prior distributions so that the corresponding marginal posterior distribution of the end point satisfies a BvM-type theorem when the true unknown density is monotone non-increasing, and illustrate it using simulated and real data. This is joint work with Natalia Bochkina (University of Edinburgh), Judith Rousseau (University of Oxford) and J. B. Salomond (Université Paris-Est Créteil).

Bayesian non-parametric methods for malware classification Mo PS

Jose Antonio Perusquia Cortes, University of Kent, UK
Co-author(s): Jim Griffin and Cristiano Villa

A malware is a software which is specifically designed to disrupt, damage or gain access to a computer system. Cyber-security is concerned not only with the fast and accurate detection of new malwares but also how to fix them. Therefore, correct classification of new malwares into known families has a key role for reverse engineering and hence for cyber-security. Assuming that each executable malware and hence each family is completely characterised through the presence or absence of features provides a natural justification for the use of factorial models. In this case and given the hexadecimal representation of the binary code, we can actually create a set of features known as n-grams that completely characterise each malware. In a parametric setting, the number of n-grams would remain fixed and finite, which is a strong assumption due to the exponential growth of information. Therefore, a Bayesian non-parametric setting seems the natural way to address this problem. In the following, we present classification methodologies based on the Hierarchical Beta Process and Compound Random Measures. These methodologies exploit the fact the number of different families is fixed and known; therefore, for a new malware we can obtain the probability that it actually belongs to each family.

Bayesian nonparametric priors for hidden Markov random fields: Application to image segmentation Mo PS

Julyan Arbel, Inria Grenoble Rhône-Alpes, France
Co-author(s): Hongliang Lu and Florence Forbes

One of the central issues in statistics and machine learning is how to select an adequate model that can automatically adapt its complexity to the observed data. Bayesian nonparametric methods are thought of as promising candidates that are capable of handling such tasks. Based on an infinite-dimensional parameter space, Bayesian nonparametric models are highly flexible and thus can be employed for parameterizing or learning about complex data sets. In the present paper, we address the problem of determining the structure of clustered data, without any prior knowledge of the number of clusters or any other information about their composition. The required guess on the number of clusters is avoided by considering models with an infinite number of components as suggested in Bayesian nonparametrics. More concretely, we propose to combine a Markov random field model with Bayesian nonparametric priors and illustrate such a combination by means of a Potts model together with a Dirichlet process and a Pitman-Yor process, respectively. To perform inference, the variational Bayesian method is adopted due to its lower computational cost with respect to its Markov chain

Monte Carlo counterpart. Finally, the proposed framework is applied to unsupervised segmentation of natural images.

Bayesian Nonparametric Unsupervised Concept Drift Detection for Data Stream Mining Mo PS

Junyu Xuan, University of Technology Sydney, Australia

Online data stream mining is of great significance in practice because of its ubiquity in many real-world scenarios, especially in the big data era. Traditional data mining algorithms cannot be directly applied to data streams due to 1) the possible change of underlying data distribution over time (i.e., concept drift) and 2) delayed, short, or even no labels for streaming data in practice. A new research area, named unsupervised concept drift detection, has emerged to tackle this difficulty mainly based on two-sample hypothesis tests, such as the Kolmogorov-Smirnov test. However, it is surprising that none of the existing methods in this area exploit the Bayesian nonparametric hypothesis test which has clear interpretability and straightforward prior knowledge encoding ability and no strict or unrealistic requirement of prefixing the form for the underlying data distribution. In this paper, we present a Bayesian nonparametric unsupervised concept drift detection method based on the Polya tree hypothesis test. The basic idea is to decompose the underlying data distribution into a multi-resolution representation which transforms the whole distribution hypothesis test into recursive and simple binomial tests. Also, an incremental mechanism is especially designed to improve its efficiency in the stream setting. The method does not only effectively detect drifts, and it also locates where a drift happens and the posteriors of hypotheses. The experiments on synthetic data verify the desired properties of the proposed method, and the experiments on real-world data show the better performance of the method for data stream mining compared with its frequentist counterpart in the literature.

Post-Processed Posteriors for Band-Structured Covariances Mo PS

Kwangmin Lee, Seoul National University, South Korea
Co-author(s): Kyoungjae Lee and Jaeyong Lee

We consider two classes of band-structured covariances, classes of banded and bandable covariances. Due to the difficulty of constructing priors with computational efficiency and theoretical optimality, the Bayesian inference for band structured covariances remains elusive. In this paper, we propose post-processed posteriors for the banded and bandable covariances. The post-processed posterior is obtained by post-processing the inverse-Wishart posterior, the conjugate posterior for the covariance without any structural restriction. The structural restriction of the posterior is satisfied by the post-processing. We show that the proposed post-processed posteriors have optimal minimax rate for bandable covariances and nearly optimal minimax rate for banded covariances. The advantage of the post-processed posterior is demonstrated by a simulation study and a real data analysis.

Consistent reconstruction of electrical impedance tomography images Mo PS

Kweku Abraham, University of Cambridge, United Kingdom
Co-author(s): A. Richard Nickl

Electrical impedance tomography is a non-invasive medical imaging technique, in which one measures voltages corresponding to different current profiles to infer conductivity; since the conductivity of different medical tissues drastically varies, from the conductivity profile one can build a 3d image of the internal structure. However, the measurement model (surface only measurements for the Dirichlet-to-

Neumann map) leads to a severely ill-posed, somewhat intractable inverse problem. Bayesian methods allow practitioners to bypass the inverse nature of the problem, sampling from the posterior with only calls to the forward operator. We aim to show that such methods have good asymptotic properties, recovering the true conductivity profile in the vanishing noise limit.

Tuesday 25 June

Policymaking and Statistical Estimates: A Bayesian Decision-Analytic Model for a Binary Outcome Tu PS

Akisato Suzuki, University College Dublin, Ireland

How should we assess the policy implications of quantitative research? There are two problems in the conventional practice. First, solely relying on statistical significance misses the fact that uncertainty is a continuous scale. Second, the criterion of so-called substantive significance is rarely explained and formally justified. To overcome these problems, I propose a Bayesian statistical decision-analytic model for a binary outcome. The posterior distribution of a causal effect estimated by Bayesian logistic regression is incorporated into a loss function over the cost of realizing the effect through a policy and the cost of not doing so. The model implies the optimal choice depends not only on the size of the causal effect but also on the ratio of the cost of implementing the policy to the cost of keeping the status quo. I exemplify my model through a replication study. While using a simple parametric Bayesian model as a basis, the decision-analytic perspective of the article suggests that nonparametric Bayesian models will also be able to benefit from a decision-analytic model such as mine, thereby helping better policymaking.

Multiple kernel learning with structured Gaussian processes: an application to drug interaction prediction Tu PS

Leiv Rønneberg, University of Oslo, Norway
Co-author(s): M. Zucknick

Cancer drugs have the potential to be more effective when given in combination. When two or more drugs are combined, they may reinforce one another, producing a combined effect larger than could be expected had the drugs acted independently. Community efforts to develop predictive models for these combination effects, utilizing data from large-scale ex-vivo screening experiments, have highlighted the complex nature of drug response, in which nonlinear modelling and efficient use of prior biological knowledge are crucial for predictive performance. We propose a multiple kernel learning (MKL) approach, that allows the flexible integration of different data sources, learning convex combinations of kernels encoding different notions of similarity. In addition, kernels for drugs and cancer cell lines are combined using a Kronecker structure which, coupled with structured Gaussian processes (SGP), allow fast parameter learning and inference. In combination, this approach allows full uncertainty quantification of predictions, the inclusion of relevant biological knowledge, and opens the door for sequential optimisation in the design of drug combination experiments.

A Bayesian nonparametric testing procedure for paired samples Tu PS

Luis Gutierrez, Pontificia Universidad Catolica de Chile, Chile
Co-author(s): L. A. Pereira and D. Taylor-Rodriguez

We propose a Bayesian hypothesis testing procedure for comparing the distributions of paired samples. The procedure is based on a flexible model for the joint distribution of both samples. The flexibility is given by a mixture of Dirichlet processes. Our proposal uses a spike-slab prior specification for the base measure of the Dirichlet process and a particular parametrization for the kernel of the mixture in order to facilitate comparisons and posterior inference. The joint model allows us to derive the marginal distributions and test whether they differ or not. The procedure exploits the correlation between samples, relaxes the parametric assumptions and detects possible differences throughout

the entire distributions. A Monte Carlo simulation study comparing the performance of this strategy to other traditional alternatives is provided. Finally, we apply the proposed approach to spirometry data collected in the U.S. to investigate changes in pulmonary function in children and adolescents in response to air polluting factors.

Bayesian neural network priors at the level of units Tu PS

Mariia Vladimirova, Inria Grenoble Rhone-Alpes, France
Co-author(s): Jakob Verbeek, Pablo Mesejo, Julyan Arbel

We investigate deep Bayesian neural networks with Gaussian priors on the weights and ReLU-like nonlinearities, shedding light on novel sparsity-inducing mechanisms at the level of the units of the network, both pre- and post-nonlinearities. The main thrust of the work is to establish that the units prior distribution becomes increasingly heavy-tailed with depth. We show that first layer units are Gaussian, second layer units are sub-Exponential, and we introduce sub-Weibull distributions to characterise the deeper layers units. Bayesian neural networks with Gaussian priors are well known to induce the weight decay penalty on the weights. In contrast, our result indicates a more elaborate regularisation scheme at the level of the units. This result provides new theoretical insight on deep Bayesian neural networks, underpinning their natural shrinkage properties and practical potential.

Bayesian Nonparametric Vector Auto Regressive models via a logit stick-breaking prior Tu PS

Mario Beraha, Politecnico di Milano and Universita degli Studi di Bologna, Italy
Co-author(s): Alessandra Guglielmi and Fernando Quintana

Vector Autoregressive (VAR) models may provide a flexible and powerful representation of longitudinal data; moreover extensions such as ARX models account for exogenous covariates. Bayesian nonparametrics has been successfully applied to VAR models in recent years but limited to purely autoregressive cases. We propose a semiparametric VAR model with exogenous covariates for non-stationary multidimensional longitudinal data, where a dependent stick-breaking prior is assumed for the autoregressive component through logit stick-breaking. We develop an efficient Gibbs sampling algorithm for MCMC posterior simulation, leveraging recent results on logit stick-breaking priors. We also illustrate the approach through simulations and medical data.

Measuring the sensitivity to prior specification for time-to-event data through the Wasserstein distance Tu PS

Marta Catalano, Bocconi University, Italy
Co-author(s): A. Lijoi and I. Prünster

A substantial portion of the literature on Bayesian nonparametric analysis of time-to-event data builds upon the notion of random measure. The paper focuses on priors for the hazard rate function, a popular choice being the kernel mixture with respect to a gamma random measure that depends itself on hyperparameters. The main goal we pursue is a quantification of the effects of changes in the specification of these hyperparameters through the Wasserstein distance. Though easy to simulate, the Wasserstein distance is generally difficult to evaluate, strengthening the need for tractable and informative bounds. Here we accomplish this task on the wider class of completely random measures and specialize our results to the gamma random measure and the related kernel mixtures. The techniques that we introduce yield upper and lower bounds for the Wasserstein distance between hazard rates, cumulative hazard rates and survival functions, both *a priori* and conditionally on the

data.

Turing.jl: Probabilistic programming with discrete random probability measures. Tu PS

Martin Trapp, Graz University of Technology, Austria
Co-author(s): Emile Mathieu, Maria Lomeli, Hong Ge

Probabilistic modelling is a core component of a scientists' toolbox for incorporating uncertainties about the model parameters and noise in the data. Statistical models with a Bayesian nonparametric component are difficult to handle due to the infinite dimensionality which prevents the straightforward use of standard inference methods. Probabilistic programming languages – such as Turing.jl (Ge et al (2018)) – enable the rapid development of new probabilistic models while simultaneously automating statistical inference. This is made possible by separating the model definition from the inference scheme and by using generic inference algorithms. Previously, Bloem-Reddy et al (2017) proposed an efficient way of representing the class of infinite mixture models as a probabilistic program for the Pitman-Yor and normalised inverse-gamma processes.

In this work we present the integration of Bayesian nonparametric priors into Turing.jl which allows non-experts to use a variety of Bayesian nonparametric mixture models using a generic modelling framework. In addition, we discuss the associated challenges for a statistics audience.

Bloem-Reddy, B., Mathieu, E., Foster, A., Rainforth, T., Ge, H., Lomeli, M., Ghahramani, Z., and Teh, Y.W., 2017, "Sampling and inference for discrete random probability measures in probabilistic programs", Approximate Inference workshop, NIPS.

Ge, H., Xu, K., and Ghahramani, Z., 2018, "Turing: Composable inference for probabilistic programming," In proceedings of AISTATS.

Bayesian nonparametric graphical models for time-varying parameters VAR Tu PS

Matteo Iacopini, Cao Foscari University of Venice & Scuola Normale Superiore of Pisa, Italy
Co-author(s): L. Rossini

Over the last decade, big data have poured into econometrics, demanding new statistical methods for analysing high-dimensional data and complex nonlinear relationships. A common approach for addressing dimensionality issues relies on the use of static graphical structures for extracting the most significant dependence interrelationships between the variables of interest. Recently, Bayesian nonparametric techniques have become popular for modelling complex phenomena in a flexible and efficient manner, but only few attempts have been made in econometrics. In this paper, we provide an innovative Bayesian nonparametric time-varying graphical framework for making inference in high-dimensional time series. We propose a novel Bayesian nonparametric graph-based approach to estimate and forecast vector autoregressive (VAR) models where the parameters are allowed to vary over time. We include a Bayesian nonparametric dependent prior specification on the matrix of coefficients and the covariance matrix by mean of a Time-Series DPP as in Nieto-Barajas et al. (2012). Following Billio et al. (2019), our hierarchical prior overcomes over-parametrization and overfitting issues by clustering the VAR coefficients into groups and by shrinking the coefficients of each group toward a common location. Thus this hierarchical prior allows to contemporaneously estimate the (potentially) sparse time-varying causal network structure and to cluster the corresponding coefficients. In our TVP-BNP-VAR model, time-varying coefficients allow to (i) estimate the temporal networks of contemporaneous and causal structures, (ii) identify different sources of time variation, from the size of shocks and/or the propagation mechanism, and (iii) accommodate for potential non-linearities.

A Data-driven Posterior for Uncertainty Exploration in Bayesian Variable Selection via Hopfield Networks Tu PS

Matteo Vestrucci, University of Texas at Austin, USA
Co-author(s): Stephen Walker

Variable Selection can be seen as a subproblem of Model Selection: given a model, a set of covariates, and a dependent variable, we are interested in selecting the optimal subset of covariates. Already for a small number of variables, the number of candidate models becomes very large, and searching the solution space with Bayesian techniques usually involves using slow MCMC methods that only partially explore the posterior distribution. In this poster we use Hopfield Networks to quickly find the modes of the posterior distribution across the whole solution space. To analyze their performance we construct a Bayesian linear model and analytically calculate the joint posterior distribution of inclusion for each variable in the model, applying it to different scenarios. After having identified the modes, we propose a strategy to explore the uncertainty around those estimates, and proceed to describe a data-driven posterior, centered on the modes, based on a Random Walk on the graph drawn by the optimization pathways of the aforementioned Hopfield Networks.

Fast Bayesian Hazard Regression Under General Censoring via Monotone P-Splines Tu PS

Matthias Kaeding, RWI - Leibniz Institute for Economic Research, Germany

The baseline hazard is the major building block of the Cox model, giving the instantaneous rate of failure, conditional on survival up to time t and covariate values of zero. Most Bayesian nonparametric approaches model the log-baseline hazard, causing the need for numerical integration for likelihood evaluation. We propose to model the integrated baseline hazard of the Cox model via monotone penalized B-splines instead; giving an analytically available likelihood, speeding up inference and eliminating approximation error. Left, right and interval censoring can be accounted for. The advantages of Bayesian P-splines carry over to the monotone case: (1) Fully automatic smoothness parameter estimation, (2) sparseness of involved design matrices. The closed form expression for the cumulative baseline hazard is used to extend inference beyond marginal effects on the hazard rate; allowing fast computation of the conditional mean and survival function and the simulation of random deviates. Inference is carried out using MCMC and posterior mode estimation. The proposed approach is tested using a Monte Carlo simulation and applied on a large data set of times until change of gas price.

Updating Variational Bayes for Online Inference of a Dirichlet Process Mixture Tu PS

Nathaniel Tomasetti, Monash University, Australia
Co-author(s): A. Catherine Forbes and B. Anastasios Panagiotelis

Variational Bayesian (VB) inference allows Bayesian analysis to be applied in a time frame significantly smaller than traditional Markov Chain Monte Carlo approaches. Although the VB posterior is an approximation, by using a rich class of approximating distributions the VB pos-terior has been shown to produce good parameter estimates and predicted values. In this paper we propose Updating VB (UVB), a recursive algorithm to update a sequence of VB posterior approximations in an online setting, with the computation of each posterior update requiring only the data observed since the previous update. An extension to the proposed algorithm, named UVB-IS, reduces the computational burden of repeated updates, by allowing the user to trade accuracy for significant increases in computational speed through the use of importance sampling. The two methods and their properties are detailed

in two separate simulation studies. An empirical applicatoin of the UVB method predicts the future behaviour of 500 vehicles on a stretch of the US Highway 101 are predicted using a Dirichlet Process Mixture model.

Bayesian Nonparametrics for Circular Statistics and Density Estimation on Compact Metric Spaces Tu PS

Olivier Binette, University of Quebec at Montreal, Canada
Co-author(s): S. Guillotte

We introduce a density basis of the trigonometric polynomials that is suitable to mixture modelling. Statistical and geometric properties are derived, suggesting it as a circular analogue to the Bernstein polynomial densities. Nonparametric priors are constructed using this basis and a simulation study shows that the use of the resulting Bayes estimator may provide gains over comparable circular density estimators previously suggested in the literature.

From a theoretical point of view, we propose a general prior specification framework for density estimation on compact metric space using sieve priors. This is tailored to density bases such as the one considered herein and may also be used to exploit their particular shape-preserving properties. Furthermore, strong posterior consistency is shown to hold under notably weak regularity assumptions and adaptative convergence rates are obtained in terms of the approximation properties of positive linear operators generating our models.

Efficient Bayesian shape–restricted function estimation with constrained Gaussian process prior Tu PS

Pallavi Ray, Texas A&M University, USA
Co-author(s): A. Debdeep Pati and B. Anirban Bhattacharya

This article revisits the problem of Bayesian shape–restricted inference in the light of a recently developed approximate Gaussian process that admits an equivalent formulation of the shape constraints in terms of the basis coefficients. We propose a strategy to efficiently sample from the resulting constrained posterior by absorbing a smooth relaxation of the constraint in the likelihood and using circulant embedding techniques to sample from the unconstrained modified prior. We additionally pay careful attention to mitigate the computational complexity arising from updating hyperparameters within the covariance kernel of the Gaussian process. The developed algorithm is shown to be accurate and highly efficient in simulated and real data examples.

Whittle approximation for locally stationary time series Tu PS

Patricio Maturana-Russel, University of Auckland, New Zealand
Co-author(s): A. Renate Meyer and B. Claudia Kirch

In many situations, real time series behaviour can locally, i.e. in a small environment around each time point, be described as approximately stationary, where the dependence structure slowly varies with time so that a global stationarity assumption does not hold true (not even approximately). This is the case for real Advanced LIGO noise in astrophysics, ENSO data in ecology and EEG data in medicine. They display a gradual change in dependence structure. Various Bayesian nonparametric approaches to analysing locally stationary time series have been developed in the recent literature. Most of these are based on (adaptively) partitioning the time series into non-overlapping and stationary segments and using the Whittle likelihood approximation for the stationary parts. Consequently, these

approaches model the time series as piecewise stationary with (few) abrupt change points, whereas often it is more realistic to assume a slowly changing structure. Here we suggest a novel Whittle-type likelihood approximation based on a moving Fourier transform explicitly developed to take the slowly changing nature into account. We explain the motivation, the potential use for Bayesian nonparametric inference, and some preliminary results.

Modeling data in simplexes Tu PS

Rayleigh Lei, University of Michigan, USA
Co-author(s): Long Nguyen

Modeling data distributions and transformations in simplexes and performing Bayesian inference can be challenging due to simplicial constraints and the behavior near boundaries. However, as demonstrated by the popularity of topic modeling, being able to do so can prove useful in a vast array of applications. In this work we propose several approaches to accomplish these goals. The first is to apply a parsimonious, simplex-preserving transformation and to induce different noise distributions on the transformed data. The second is organize the simplicial elements into a hierarchy of subsets, an approach inspired by the nested Chinese Restaurant Process. We proceed to show how to perform inference and apply it to simulations and real world data sets.

Generalized modes in Bayesian inverse problems Tu PS

Remo Kretschmann, Universität Duisburg-Essen, Germany
Co-author(s): Christian Clason, Tapio Helin and Petteri Piiroinen

We examined generalised modes in nonparametric Bayesian inverse problems. We gave examples for measures where the common definition of a (strong) mode excludes points that we would intuitively consider a mode, whereas the generalised definition covers these points. We showed that under certain conditions, which are fulfilled for a number of widely used measures, strong and generalised modes coincide. Then we considered inverse problems with a uniform prior on a compact set, characterised the generalised posterior modes as minimisers of a canonical objective functional and showed consistency of the generalised MAP estimator in the presence of Gaussian noise. 1

Hierarchical Species Sampling Models Tu PS

Roberto Casarin, University Ca' Foscari of Venice, Italy
Co-author(s): F. Bassetti and L. Rossini

This paper introduces a general class of hierarchical nonparametric prior distributions which includes new hierarchical mixture priors such as the hierarchical Gnedin measures, and other well-known prior distributions such as the hierarchical Pitman-Yor and the hierarchical normalized random measures. The random probability measures are constructed by a hierarchy of generalized species sampling processes with possibly non-diffuse base measures. The proposed framework provides a probabilistic foundation for hierarchical random measures, and allows for studying their properties under the alternative assumptions of diffuse, atomic and mixed base measure. We show that hierarchical species sampling models have a Chinese Restaurants Franchise representation and can be used as prior distributions to undertake Bayesian nonparametric inference. We provide a general sampling method for posterior approximation which easily accounts for non-diffuse base measures such as spike-and-slab.

Bayesian Non-Parametric Inference for Stochastic Epidemic Models Tu PS

Rowland Seymour, University of Nottingham, UK
Co-author(s): T. Kypraios and P. D. O'Neill

Simulating from and making inference for stochastic epidemic models are key strategies for understanding and controlling the spread of infectious diseases. Despite the enormous attention given to methods for parameter estimation, there has been relatively little activity in the area of non-parametric inference. That is, drawing inference for the infection rate without making specific modelling assumptions about its functional form. We develop novel Bayesian non-parametric methodology to fit heterogeneously mixing models in which the infection rate between two individuals is a function, $f(\cdot)$, of their characteristics, for example location or type. Making non-parametric inference in this context is very challenging because the likelihood function of the observed data is intractable. We adopt a fully Bayesian approach by assigning a Gaussian Process (GP) prior to $f(\cdot)$ and then develop an efficient data augmentation Markov Chain Monte Carlo methodology to estimate $f(\cdot)$, the GP hyperparameters and the unobserved infection times. We then extend this method by using multi-output GP prior distributions to infer infection rates which depend on both continuous and discrete characteristics and covariates. We illustrate our methodology using simulated data and by analysing a data set on Avian Influenza from the Netherlands.

Evaluating Sensitivity to the Stick Breaking Prior in Bayesian Nonparametrics Tu PS

Ryan Giordano, UC Berkeley, USA
Co-author(s): Runjing Liu, UC Berkeley, USA, Michael I. Jordan, UC Berkeley, USA, Tamara Broderick, MIT, USA

A central question in many probabilistic clustering problems is how many distinct clusters are present in a particular dataset. Bayesian nonparametrics (BNP) addresses this question by placing a generative process on cluster assignment, making the number of distinct clusters present amenable to Bayesian inference. However, like all Bayesian approaches, BNP requires the specification of a prior, and this prior may favor a greater or fewer number of distinct clusters. In practice, it is important to quantitatively establish that the prior is not too informative, particularly when—as is often the case in BNP—the particular form of the prior is chosen for mathematical convenience rather than because of a considered subjective belief.

We derive local sensitivity measures for a truncated variational Bayes approximation based on the Kullback-Leibler divergence. Local sensitivity measures approximate the nonlinear dependence of a VB optimum on prior parameters using a local Taylor series approximation. Using a stick-breaking representation of a Dirichlet process, we consider perturbations both to the scalar concentration parameter and to the functional form of the stick- breaking distribution.

In the design and evaluation of our local sensitivity measures we pay special attention to our ability to accurately extrapolate to different priors, rather than treating the sensitivity as a measure of robustness per se. Extrapolation motivates the use of multiplicative perturbations to the functional form of the prior for VB, as the KL divergence is then linear in the perturbation. Additionally, we linearly approximate only the computationally intensive part of inference—the optimization of the global parameters—and retain the non-linearity of easily computed quantities.s

We apply our methods to real and simulated datasets to estimate sensitivity of the expected number of distinct clusters present to the BNP prior specification, evaluating the accuracy of our approximations by comparing to the much more expensive process of re-fitting the model.

Bayesian Varying Coefficients Models Based on Gaussian Process Priors Tu PS

Sanvesh Srivastava, The University of Iowa, USA
Co-author(s): A. Rajarshi Guhaniyogi and B. Cheng Li and C. Terrance Savitsky

Bayesian varying coefficients models based on Gaussian processes are popular in many disciplines because they balance flexibility and interpretability. Markov chain Monte Carlo methods are available to fit these models, but they are inefficient even for moderately large data. Motivated by the task of flexible modeling of massive spatiotemporal data, we develop a three-step divide-and-conquer method for fitting space-time varying coefficients models based on Gaussian processes. Our method randomly partitions the space-time tuples into a large number of overlapping subsets, obtains Markov chain Monte Carlo (MCMC) samples of parameters and predictions in parallel across the subsets, and combines MCMC samples from all the subsets into parameter samples and predictions from a posterior distribution that conditions on the entire data. We provide guidance for choosing the number of subsets depending on the analytic properties of the Gaussian processes and develop a new data augmentation scheme for sampling from the posterior distribution parameters on the subests. Our method has better coverage, shorter credible intervals, and smaller mean square error than its main competitors across diverse simulations, where sampling from the full data posterior distribution is inefficient. Our method retains its superior performance in the analysis of the temperature and precipitation data over 100 years in the U.S.A. While developed in the context of spatiotemporal applications, our algorithms is applicable to any Bayesian varying coefficients model based on Gaussian processes.

Bayesian Hierarchical Modeling on Covariance Valued Data Tu PS

Satwik Acharyya, Texas A&M University, USA
Co-author(s): A. Zhengwu Zhang, B. Anirban Bhattacharya and C. Debdeep Pati

Analysis of structural and functional connectivity (FC) of human brains is of pivotal importance for diagnosis of cognitive ability. The Human Connectome Project (HCP) provides an excellent source of neural data across different regions of interest (ROIs) of the living human brain. Individual specific data were available from an existing analysis (Dai et al., 2017) in the form of time varying covariance matrices representing the brain activity as the subjects perform a specific task. As a preliminary objective of studying the heterogeneity of brain connectomics across the population, we develop a probabilistic model for a sample of covariance matrices using a scaled Wishart distribution. We stress here that our data units are available in the form of covariance matrices, and we use the Wishart distribution to create our likelihood function rather than its more common usage as a prior on covariance matrices. Based on empirical explorations suggesting the data matrices to have low effective rank, we further model the center of the Wishart distribution using an orthogonal factor model type decomposition. We encourage shrinkage towards a low rank structure through a novel shrinkage prior and discuss strategies to sample from the posterior distribution using a combination of Gibbs and slice sampling. We extend our modeling framework to a dynamic setting to detect change points. The efficacy of the approach is explored in various simulation settings and exemplified on several case studies including our motivating HCP data.

Bayesian Quadrature with BART for Bayesian Survey Design Tu PS

Seth Flaxman, Imperial College London, UK
Co-author(s): R. Kang, X. Liu, Z. Shen, H. Zhu, S. Flaxman

Bayesian Quadrature (BQ) is an important tool for solving statistical and scientific problems with an integral at their heart. We propose a new approach to BQ based on Bayesian Additive Regression Trees (BART). BART is easy to tune, automatically handles a mixture of discrete and continuous variables,

and has attractive theoretical results. We show how BART lends itself to an elegant formulation of BQ, with a simple but effective sequential sampling approach. We explore the use of BART in Bayesian survey design, providing an effective alternative to both a simple random sample (Monte Carlo) and a more sophisticated block (stratified) random sampling design.

Variable Selection Consistency of Gaussian Process Regression Tu PS

Sheng Jiang, Duke University, USA
Co-author(s): Surya Tokdar

Sparse, rescaled Gaussian process regression achieves adaptation with near optimal convergence rates up to a logarithmic factor. However, estimation consistency does not guarantee variable selection consistency. It remains unclear if sparse, rescaled Gaussian process regression is able to exactly identify the set of true regressors. This paper shows variable selection consistency of Gaussian process regression with sparse, rescaled Squared Exponential kernel Gaussian process priors. The proof follows the prior mass and existence of tests framework of Schwartz theory and shows the posterior probability of false positive models vanishes to 0 in probability. The key technical development is to provide sharp upper and lower bounds for various small ball probabilities in L_2 distance at all rescaling levels of the Gaussian process prior.

A Bayesian Estimation of Panel Stochastic Frontier Models with Determinants of Persistent and Transient Inefficiencies in Both Location and Scale Parameters Tu PS

Sheng-Kai Chang, National Taiwan University, Taiwan
Co-author(s): Ruei-Chi Lee

By incorporating factors to explain both persistent and transient technical inefficiencies, we estimate the four-random-component stochastic frontier model by means of the Bayesian approach. The impacts of those persistent and transient technical inefficiency factors in terms of both location and scale parameters of inefficiency distributions are considered and marginal effects of the determinants for both persistent and transient technical inefficiencies are computed in the paper. We also apply the proposed Bayesian estimators to study Taiwanese banks over the period 2008 to 2016. It is found that bank capitalization has a positive effect in the short run and total assets have a negative effect in the long run. Moreover, in terms of the Deviance Information Criterion (DIC), it is shown that the model with determinants of inefficiencies in both location and scale parameters performs better than the models with determinants of inefficiencies in only location or scale parameters.

Towards a Bayesian nonparametric genome-wide association study Tu PS

Shijia Wang, Simon Fraser University, Canada
Co-author(s): Caroline Colijn, Liangliang Wang and Lloyd T. Elliott

In genome wide-association studies, linear mixed models (LMMs) are powerful methods for controlling confounding caused by population structure. Work based on the LMM assumes that the genetic similarity matrix (used as a fixed effect) is known. However, uncertainty about the phylogeny may degrade the quality of LMM results for studies in which the number of samples is small, or the samples are poorly genotyped, as is the case with bacteria studies. In this work, we develop a Bayesian hierarchical model to jointly estimate the LMM parameters and the genetic similarity matrix using genetic sequences and phenotypes. We link the genetic similarity matrix to the phenotype through a phylogenetic tree with a Bayesian nonparametric prior. We propose a sequential Monte Carlo method to jointly approximate the posterior distributions of the LMM and the phylogeny. The proposed

method is easy to parallelize and provides a consistent representation of uncertainty in the phylogeny. In order to conduct efficient inference in this model, we must ensure that the sampled trees approach the posterior. Mismatch in the dimensionality of the likelihood and the prior can lead to poor mixing. We examine three ways of annealing the joint likelihood function based on conditional effective sample size and demonstrate these methods using simulation studies.

Bayesian cumulative shrinkage for infinite factorizations Tu PS

Sirio Legramanti, Bocconi University, Italy
Co-author(s): D. Durante and D. B. Dunson

There are a variety of Bayesian models relying on representations in which the dimension of the parameter space is, itself, unknown. For example, in factor analysis the number of latent variables is, in general, not known and has to be inferred from the data. Although classical shrinkage priors are useful in these situations, incorporating cumulative shrinkage can provide a more effective option which progressively penalizes more complex expansions. A successful proposal within this setting is the multiplicative gamma process. However, such a process is limited in scope, and has some drawbacks in terms of shrinkage properties and interpretability. We overcome these issues via a novel class of convex mixtures of spike and slab distributions assigning increasing mass to the spike through an adaptive function which grows with model complexity. This prior has broader applicability, simple interpretation, parsimonious representation, and induces adaptive cumulative shrinkage of the terms associated with redundant, and potentially infinite, dimensions. Performance gains are illustrated in simulation studies.

Bernstein von Mises theorems for general stick-breaking process priors. Tu PS

Stefan Franssen, Leiden University, The Netherlands

We introduce a class of species sampling process priors with independent relative stick-breaking weights, including the Dirichlet process prior and more generally the Pitman-Yor process prior, and study their theoretical performance. We show that the corresponding posteriors are consistent under the assumption that the relative stick-breaking weights are identically distributed, with densities satisfying certain smoothness assumptions. Furthermore, we prove Bernstein-von Mises theorems in two cases. First for the new class of priors for atomless true distribution P_0 and then for the specific Pitman-Yor process prior for general P_0.

Bayesian inference for multivariate extremes Tu PS

Stefano Rizzelli, EPFL, Switzerland
Co-author(s): S. Padoan

Multivariate extreme value theory provides the probabilistic framework for modelling the extremal behavior of a set of random variables. Different asymptotic characterizations of multivariate extreme events are available. We focus on max-stable distributions. In its general formulation, this is a multivariate semi-parametric class of distributions, which makes the Bayesian approach particularly appealing for simultaneous inference about the marginal parameters and the extremal dependence structure. The latter can be equivalently represented through Pickands dependence functions (A) and angular probability measures (H). An elegant way of modelling both representations is via a nonparametric Bayesian approach based on Bernstein polynomials, as shown by Marcon, Padoan and Antoniano [Electron. J. Stat. 10 (2016) 3310-3337] in the bivariate case. We expand their approach to higher dimensional cases and extend prior specification to include the parameters of the

univariate marginal distributions. We investigate the asymptotic properties of our procedure, e.g. the contraction of the full posterior distribution at the true marginal parameters and dependence functions. We conclude by extending our asymptotic results to the more realistic case of data drawn from a distribution which is in the domain of attraction of a max-stable one.

Challenges and proposals for Dirichlet process mixture models with Gaussian kernels Tu PS

Wei Jing, University of St Andrews, UK
Co-author(s): M. Papathomas and S. Liverani

We consider the Dirichlet process mixture model (DPMM) in the context of clustering for continuous data, when the conditional likelihood is set to be the multivariate normal distribution. Our simulation studies show that the DPMM may struggle to uncover the true clusters when the data contain even just a handful of variables, even when the normality assumption is correct. In this poster, we first give potential reasons of why it can be difficult for the DPMM to identify true clusters. Specifically, this may be because of the difference between the overall covariance matrix for the variables (calculated from pooling the data of all the clusters) and the within-cluster covariance matrices. For standard MCMC algorithms, this difference impedes the sampler from moving towards the target cluster allocation. Another possible reason is that the Inverse Wishart distribution is not flexible enough to serve as the prior distribution for the within-cluster covariance matrix. To investigate the effect of the prior specification we implemented different prior distributions for the within-cluster covariance matrix, and compared their performance for datasets of different size and structure. We also propose how to initialize the MCMC sampler to effect considerable improvement in the clustering results. Finally, we discuss limitations of our current proposals and future work.

Variational Nonparametric Discriminant Analysis Tu PS

Weichang Yu, University of Sydney, Australia
Co-author(s): Lamiae Azizi and John T. Ormerod

Variable selection and classification methods are common objectives in the analysis of high-dimensional data. Most of these methods make distributional assumptions that may not be compatible with the diverse families of distributions in the data. In this paper, we propose a novel Bayesian nonparametric discriminant analysis model that performs both variable selection and classification in a seamless framework. Pólya tree priors are assigned to the unknown group-conditional distributions to account for their uncertainty and allow prior beliefs about the distributions to be incorporated simply as hyperparameters. The computational cost of the algorithm is kept within an acceptable range by collapsed variational Bayes inference and a chain of functional approximations. The resultant decision rules carry heuristic interpretations and are related to the existing two-sample Bayesian nonparametric hypothesis test. By an application to some simulated and publicly available real datasets, we showed that our method performs well in comparison to current state-of-art approaches.

A Bayesian Nonparametric Spiked Process Prior for Dynamic Model Selection Tu PS

Weixuan Zhu, Xiamen University, China
Co-author(s): A. Cassese, M. Guindani and M. Vannucci

In many applications, investigators monitor processes that vary in space and time, with the goal of identifying temporally persistent and spatially localized departures from a baseline or normal behavior. In this manuscript, we consider the monitoring of pneumonia and influenza (P&I) mortality, to detect

influenza outbreaks in the continental United States, and propose a Bayesian non-parametric model selection approach to take into account the spatio-temporal dependence of outbreaks. More specifically, we introduce a zero-inflated conditionally identically distributed species sampling prior which allows borrowing information across time and to assign data to clusters associated to either a null or an alternate process. Spatial dependences are accounted for by means of a Markov random field prior, which allows to inform the selection based on inferences conducted at nearby locations. We show how the proposed modeling framework performs in an application to the P&I mortality data and in a simulation study, and compare with common threshold methods for detecting outbreaks over time, with more recent Markov switching based models, and with spike-and-slab Bayesian nonparametric priors that do not take into account spatio-temporal dependence.

EP-IS: Combining expectation propagation and importance sampling for Bayesian non-linear inverse problems Tu PS

Willem van den Boom, National University of Singapore, Singapore
Co-author(s): Alexandre H. Thiery

Bayesian analysis of inverse problems provides principled measures of uncertainty quantification. However, Bayesian computation for inverse problems is often challenging due to the high-dimensional or functional nature of the parameter space. We consider a setting with a Gaussian process prior in which exact computation of nonlinear inverse problems is too expensive while linear problems are readily solved. Motivated by the latter, we iteratively linearize nonlinear inverse problems. Doing this for data subsets separately yields an expectation propagation (EP) algorithm. The EP cavity distributions provide proposal distributions for an importance sampler that refines the posterior approximation beyond what linearization yields. The result is a hybrid between fast, linearization-based approaches, and sampling-based methods, which are more accurate but usually too slow.

Human Behaviour Analysis Through Probabilistic Modelling of GPS Data Tu PS

Yazan Qarout, Aston University, UK
Co-author(s): A. Yordan Raykov and B. Max Little

In urban city planning, it is becoming increasingly difficult to improve and maintain the inhabitants' quality of life and security due to the quickly increasing populous. Human behavioural characteristics and movement understanding can be an important tool to help assure improvement in the realm of urban planning. However, particularly when examining automated geolocated data (for example GPS data), studies in the field of human movement analysis are uncommon and often require labels that are difficult to find in real world applications. This is possibly due to the challenges associated with mining and analysing the dynamic, highly-irregularly sampled, noisy and sparse data. Nevertheless, the lower sensing modality of GPS data has less ethical and privacy concerns to other behaviour monitoring sensors such as CCTV cameras, yet can hold very rich information on behavioural characteristics making it an attractive data source to study.

Previous research in extracting behavioural information from GPS data utilised time series similarity measures and parametric modelling given fixed pre-set assumptions on the data trends. However, different trajectories often have varying number of observations and multi-modular dynamics; causing inaccuracies or raising difficulties in using these techniques. In order to understand the data, and bypass the previously mentioned challenges we propose a generative, multi-modal probabilistic model for GPS data summarising the high dimensional time series information with the models? parameters and states. Using only the trajectory sequence $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ where $\mathbf{x}_t \in \mathbb{R}^3 = (time, longitude, latitude)$, the feature sequence Y is calculated to describe the trajectory path where $Y = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$,

$\mathbf{y}_t \in \mathbb{R}^4 = (longitude, latitude, h, v)$, h is the hour of day and v is the velocity calculated using the Haversine equation. Human movement patterns are complex making them difficult to describe with a single set of parameters. Therefore, the generative model structure may better resemble the mechanics of a switching model with parameters optimised for each state in the data sequence corresponding to different behaviour patterns. We propose a novel Input infinite Switching Vector Autoregressive Model (IiSVARM). Building on the theory of the Vector Autoregressive infinite Hidden Markov (VAR-iHMM) structure, we introduce a new discrete input variable $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_T)$ where $\tau_t = i$. This acts as a parent to the state indicator variable $\mathbf{z}$ in the Probabilistic Graphical Model (PGM) to represent environmental factors that influence human movement behaviour. $\boldsymbol{\tau}$ assumes possibly different state transition dynamics for different environmental conditions, while still sharing the same state parameters across all transition structures. For example, if $\tau_t \in \{1, 2\}$ representing off-peak, and peak times respectively, the model can better identify the differences in behaviour between these two time regions while still allowing for similarities and overlaps in patterns should they occur. The final formalization of the model can be summarised with the following equations

$$
\begin{aligned}
Q|\psi, H &\sim DP(\psi, H), \\
G_0^{(i)}|\gamma, H, \boldsymbol{\tau} &\sim DP(\gamma, Q), \\
G_k^{(i)}|\alpha, G_0^{(i)} &\sim DP(\alpha, G_0^{(i)}),
\end{aligned}
$$

where Q is the base probability measure, H is the master probability measure associated with the values of $\boldsymbol{\tau}$, $G_o^{(i)}$ is the global probability measure for $\tau = i$, and $G_j^{(i)}$ is the random probability measure for state j when $\tau = i$ and ψ, γ and α are the concentration parameters for each respective Dirichlet Process (DP).

Experiments have demonstrated that different transition matrices are learned for different values of τ_t; matching the expected state transition dynamics of the contextual information it represents, with unique and overlapping state transitions according to what behaviour the state represents. Consequently, the identified states segment the data into highly informative clusters based on the information that the input features convey, and the results outperform alternative techniques in literature.

Useful Information

Orientation

All the talks and poster sessions will be held on the mezzanine floor (-1 level) of the Mathematical Institute (MI), Andrew Wiles Building, Woodstock Road, Oxford.
The **registration desk** will be located on the mezzanine floor (-1 level) of the Mathematical Institute. It will be open from 8:30am on Monday 24 June, and after that during the conference hours. Please collect your badge there when you arrive. Badges should be worn at all time to have access to the conference sessions.
Keynote talks, invited sessions and the **Bayesian Nonparametrics Section Meeting** will be held in the lecture theatre 2 (L2). **Contributed sessions** will be held in the lecture theatres 2 (L2) and 3 (L3). **Tea and Coffee breaks** will be offered on the mezzanine floor outside the lecture theatres.
Poster sessions will be held on Monday and Tuesday evening on the mezzanine floor. Drinks and nibbles will be served during the poster session.

Social events

The **welcome reception** on Sunday evening will be held at the Pitt Rivers Museum. The access is via the Pitt Rivers Museum's South Entrance in Robinson Close, off South Parks Road.
The **conference dinner** on Thursday evening will be held at Lady Margaret Hall, (Norham Gardens, Oxford OX2 6QA) one of Oxford's colleges. Please bring your dinner voucher to the conference dinner.
Two **walking tours** of Oxford are organised on Wednesday evening, starting at 7:30pm or 8pm. Bookings can be made at the registration desk. They are free of charge but the number of places is limited and will be allocated on a first come, first served basis.
A **Junior ISBA reception** will be organised on Wednesday evening from 7:15pm to 9pm in the Department of Statistics, 24-29 St-Giles, Oxford. The reception is intended for students or researchers within 5 years of having completed their degree. Non ISBA members are very welcome to attend.

Lunches and dinner

Lunches and dinners (except the conference dinner on Thursday evening for attendants who took that option) are not included in the conference fees. The Café is a cafeteria situated on the Mezzanine floor of the MI, open from 8:30 – 16:15 Monday to Friday. Close to the venue, you will find a number of sandwich places on Woodstock Road next to the junction with Little Clarendon Street, and various restaurants in Little Clarendon Street, Walton street or North parade Avenue. Some restaurants will be offering discounts to BNP12 delegates on a la carte menus, when shown your conference badge (Café Rouge 40%, Browns 20%, Carluccios 20%).

Wifi

Eduroam is available at the Mathematical Institute and all over the University of Oxford.
`https://www.eduroam.org/`
Attendees who do not have access to Eduroam can access the internet using The Cloud network. 1/ Select "_The Cloud" from the available network list. 2/ Open your internet browser, the venue landing page will appear. 3/ If this is your first time using The Cloud Wifi network, follow the simple one-time registration process.

Information for speakers and poster presenters

Talks. Please bring your presentation on a USB stick and upload it on the lecture room's computer at least 15 min before the start of the session. Each invited talk has a 30 min slot, including floor discussion. Each contributed talk has a 22 min slot, including floor discussion.

Posters. The poster format is A0 portrait. Please hang your poster before the poster session (you are scheduled either on Monday or Tuesday), and remove it by 10am the next day. There will be poster prizes for the best posters in the categories "Theory and Methods" and "Applications", awarded by the scientific committee.

Childcare

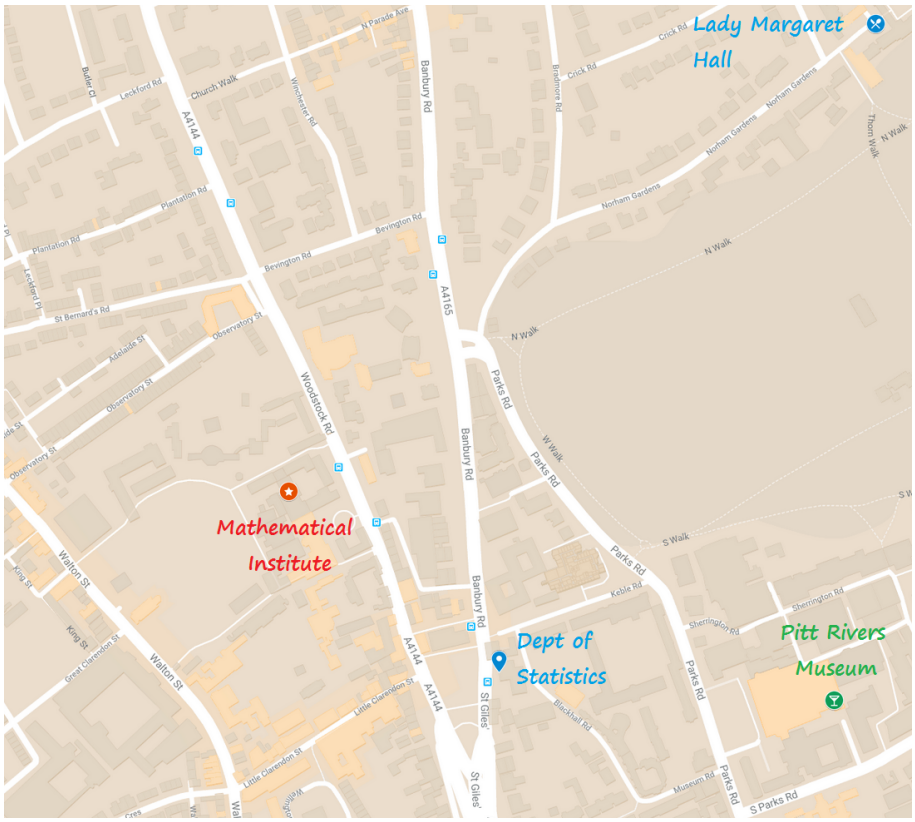
A childcare service is provided by Little Hens Childcare during the conference, and located at the Mathematical Institute. The service requires prior booking, but places may still be available.

Mobility

To reach the lecture theatres (for those who cannot manage stairs): pass through the large glass doors nearest reception and access the lifts to go to the mezzanine (-1) level. The reception staff can release the door lock. The doors are not powered.
To reach the lower levels of the lecture theatres there are individual platform lifts located down corridors alongside the theatre - as these lifts are 'behind the scenes' you will need to be escorted to them. Please arrange access via the conference organisers or otherwise ask MI reception staff.

Quiet room

There is a quiet room that can be used for those wishing to breastfeed if they would like. Please ask the conference desk (mezzanine floor), or the MI reception (ground floor).



List of presenters