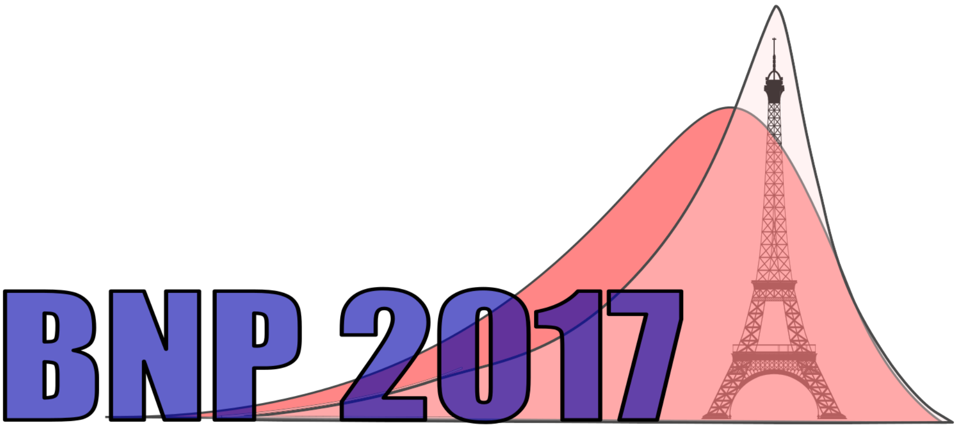

11th Meeting on Bayesian Non-Parametrics

26th to 30th of June 2017
Paris, France

Book of Abstracts

	Monday	Tuesday	Wednesday	Thursday	Friday
9:00 am			**Feature allocations in theory and practice (N. Foti)** S. Williamson, T. Campbell, V. Rockova	**Multiscale and multiresolution (D. Dunson)** L. Ma,T. Canale, S. Ghoshal	
9:30 am	**Plenary speaker 1** Ryan Adams Chair: B.Engelhardt	**Plenary speaker 2** Barbara Engelhardt Chair : P. Mueller			**Plenary speaker 3** Peter Orbanz Chair: S. Petrone
10:00 am					
10:30 am	**Coffee Break**	**Coffee Break**	**Coffee Break**	**Coffee Break**	**Coffee Break**
11:00 am	**BNP & Big Data (T. Campbell)** J. Johndrow, N. Foti, R. Ranganath	**Recent Contributions to BNP arising from Biostatistics (F. Quintana)** C. Holmes, A. Jara, L. Trippa	**BNP, Empirical Likelihood, and Support-Based Nonparametric Modeling (C. Robert)** N. Hjort, A. Owen, R. Solgi	**Model misspecification (C. Holmes)** J. Miller, R. Martin, S. Walker	**Bayesian nonparametric modeling of networks and relational data (F. Caron)** V. Veitch, D. Durante, I. Antoniano-Villalobos
11:30 am					
12:00 am					
12:30 am	**Lunch**	**Lunch**	**Lunch**	**Lunch**	
1:00 pm					
1:30 pm					
2:00 pm	**Model selection in high-dimensional and nonparametric problems (A. Bhattacharya)** D. Rossell,S. Tokdar,D. Pati	**Recent developments in asymptotics Posterior contraction and credible set (I. Castillo)** A. van der Vaart, R. Nickl, B. Szabo	**Recent advances in dependent random measures and multivariate processes (F. Caron)** F. Leisen,A. Lijoi, L. James	**Bayesian Nonparametrics for Improved Decision Making (W. Johnson)** M. Guindani, A. Guglielmi, P. Mueller	
2:30 pm					
3:00 pm					
3:30 pm	**Coffee Break**	**Coffee Break**	**Coffee Break**	**Coffee Break**	
4:00 pm	**(ENS) Scalable Bayesian nonparametric inference** Barrientos A., Zhang M., Raykov Y. **(IHP) Completely random measures and DDPs** Camerlengh F., Filippi S., Huggins J.	**(ENS) Applications of Bayesian nonparametrics methods** Wu A. ; DeVito R. ; Xu J. **(IHP) Asymptotia** : Yun Y. ; Yoo W. ; Bissiri P-G.	**(ENS) Machine learning** Trapp M. ; Li D. ; Saha A. **(IHP) Partition models and clustering**: Battiston M. ; Benedetto G. ; Stephenson B.	**(ENS) Graphs** Miscouridou X., Bloem-Reddy B.; Cai D. **(IHP) BNP & EFab** : M. Jensen ; L. Liu ; J. Griffin	
4:30 pm					
5:00 pm					
5:30 pm					
6:00 pm					
6:30 pm	**Poster Session & Drinks at Université Paris 6**	**Poster Session & Drinks at Université Paris 6**			
7:00 pm					
7:30 pm				**Conference Diner** **“Chez Françoise”**	
8:00 pm					
8:30 pm					
9:00 pm					

Detailed Program

Monday

9:30 - 10:30 – Plenary Session **Chair**: B. Engelhardt
Ryan Adams Building Probabilistic Structure into Massively Parameterized Models

10:30 - 11:00 – Coffee Break

11:00 - 12:30 – Invited Session : BNP & Big Data **Chair**: T. Campbell
J. Johndrow MCMC for large scale problems
N. Foti Variational Boosting: Iteratively Refining Posterior Approximations
R. Ranganath Variational Inference: Modern Methods

12:30 - 14:00 – Lunch

14:00 - 15:30 – Invited Session : Model selection in high-dimensional and nonparametric problems
Chair: A. Bhattacharya
D. Rossell Choice of mixture components via non-local priors
S. Tokdar airGP: Bayesian Smoothing in High Dimension
D. Pati Community detection and goodness of fit tests in random graph models: a probabilistic approach

15:30 - 16:00 – Coffee Break

16:00 - 17:30 – Contributed (ENS): Scalable Bayesian nonparametric inference **Chair**: V. Rockova
A. Barrientos Bayesian Bootstraps for Massive Data
M. Zhang Embarrassingly parallel inference for Gaussian processes
Y. Raykov Efficient block inference for BNP models using Beta distributed slices

16:00 - 17:30 – Contributed (IHP) : Completely random measures and DDPs **Chair**: R. Mena
F. Camerlenghi Dependent nonparametric priors for partially exchangeable data
S. Filippi Bayesian Learning of Kernel Embeddings
J. Huggins Truncated Random Measures

18:30 - 21:00 – Poster Session I

Joseph Antonelli
Julyan Arbel
Andrea Arfè
Raffaele Argiento
Douglas Armstrong
Morgane Austern
Fadhel Ayed
Cecilia Balocchi
Sayantan Banerjee
Brenda Betancourt
Anirban Bhattacharya
Saranjeet Kaur Bhogal
Ilaria Bianchini
Enrico Bibbona
Olivier Binette
Sudip Bose
Lorenzo Cappello
Minwoo Chae
Antik Chakraborty
Noirrit Kiran Chandra

Sheng-Kai Chang
Ya-Ting Chang
Zhen Chen
Will Cipolli
Andrea Cremaschi
Miguel de Carvalho
Kumaresh Dhara
Stefanos Dimitrakopoulos
Haosui Duanmu
Bianca Dumitrascu
Ehtsham Elahi
Clément Elvira
Melanie F. Pradier
Stefano Favaro
Tamara Fernandez
Jairo Alberto Fuquene Patiño
Bastian Galasso
Clara Grazian
Rajarshi Guhaniyogi
Luis Gutierrez
Jarno Hartog
Boris Hejblum
Katherine Heller
Gudmund Hermansen
Reyhaneh Hosseini
Colman Humphrey
Matteo Iacopini
Vanda Inacio de Carvalho
Moumita Karmakar
Milad Kharratzadeh
Samer Kharroubi
Alisa Kirichenko

Tuesday

9:30 - 10:30 – Plenary Session **Chair**: P. Mueller
Barbara Engelhardt Bayesian nonparametrics in the wild: Opportunities and challenges in genomic analyses

10:30 - 11:00 – Coffee Break

11:00 - 12:30 – Invited Session : Recent Contributions to BNP arising from Biostatistics **Chair**: F. Quintana
C. Holmes Nonparametric tests
A. Jara Marginal Bayesian Semiparametric Modelling of Mismeasured Multivariate Interval-Censored Data
L. Trippa Bayesian Nonparametric Ordination for the Analysis of Microbial Communities

12:30 - 14:00 – Lunch

14:00 - 15:30 – Invited Session : Recent developments in asymptotics Posterior contraction and credible set **Chair**: I. Castillo
A. van der Vaart Credible sets for sparse models
R. Nickl Bernstein von Mises theorems for non-linear inverse problems
B. Szabo Frequentist coverage properties of Bayesian credible sets for sieve priors in general settings

15:30 - 16:00 – Coffee Break

16:00 - 17:30 – Contributed (ENS): Applications of Bayesian nonparametrics methods **Chair**: J. Lee
A. Wu Brain Kernel: A covariance function for fMRI data using a large-scale Gaussian process latent variable model
R. DeVito Bayesian Multi-study Factor Analysis in High-dimensional Biological Data
J. Xu General Mode Decompositions and Spectral Analysis Using a Mechanistic Hierarchical Based Deep Gaussian process

16:00 - 17:30 – Contributed (IHP) : Asymptotia **Chair**: J-B.Salomond
Y. Yang Bayesian model selection consistency and oracle inequality with intractable marginal likelihood
W. Yoo Bayesian inference in infinite-dimensions: Adaptive procedures and their implications
P-G. Bissiri

18:30 - 21:00 – Poster Session II

Ilja Klebanov
Guillaume Kon Kam King
Suprateek Kundu
Minjung Kyung
Juho Lee
Frank Liu
Shuhei Mano
Nicolo' Margaritella
Asael Fabian Martínez
Gianluca Mastrantonio
Alexander Meier
Anuj Mishra
Masahiro Nakano
Zacharie Naulet
Yang Ni
Bernardo Nipoti

Akihiko Nishimura
Andriy Norets
Nurzhan Nurushev
Garritt Page
Annalisa Cerquetti
Freddy Palma
Sayan Patra
Iliana Peneva
Valerio Perrone
Valerie Poynor
Anthony Quinn
Fernando Quintana
Kolyan Ray
Alan Riva Palacio
Jenovah Rodrigues
Smit Rohan
Luca Rossini
Massimiliano Russo
S. Yaser Samadi
Hristo Sariev
Terrance Savitsky
Ingmar Schuster
Paulo Serra
Siva Sivaganesan
Sanvesh Srivastava
Rebecca Steorts
Justin Strait
Laura Turbatu
Willem van den Boom
Johan Van Der Molen Moris
Jan van Waaij
Elodie Vernet
Junyu Xuan
Juan Yan
Mingan Yang
Mahmoud Zarepour
Mingyuan Zhou
Shuang Zhou

Wednesday

9:00 - 10:30 – Invited Session : Feature allocations in theory and practice **Chair**: N. Foti
S. Williamson Restricted Indian buffet processes
T. Campbell Exchangeable Trait Allocations
V. Rockova Particle EM for Variable Selection

10:30 - 11:00 – Coffee Break

11:00 - 12:30 – Invited Session : BNP, Empirical Likelihood, and Support-Based Nonparametric Modeling **Chair**: C. Robert
N. Hjort Empirical likelihood and Bayesian nonparametrics
A. Owen A survey of Bayesian empirical likelihood
R. Solgi Nonparametric hierarchical Bayesian quantiles

12:30 - 14:00 – Lunch

14:00 - 15:30 – Invited Session : Recent advances in dependent random measures and multivariate processes **Chair**: F. Caron
F. Leisen Some recent constructions of vectors of dependent completely random measures
A. Lijoi Posterior representations of dependent random probabilities
L. James q-Hierarchical Latent Feature Models

15:30 - 16:00 – Coffee Break

16:00 - 17:30 – Contributed (ENS): Machine Learning **Chair**: D. Rossel
M. Trapp Infinite Sum-Product Networks
D. Li Bayesian manifold learning using locally curved basis functions
A. Saha A Geometric Variational Approach to Bayesian Inference

16:00 - 17:30 – Contributed (IHP) : Partition models and clustering **Chair**: A.Guglielmi
M. Battiston A characterization of product-form exchangeable feature probability functions
G. Benedetto Non-exchangeable random partition model for microclustering
B. Stephenson Robust model-based clustering for multivariate and grouped data via local deviation processes

Thursday

9:00 - 10:30 – Invited Session : Multiscale and multiresolution **Chair**: D. Dunson
L. Ma Three techniques for multi-scale nonparametric modeling
T. Canale Multiscale mixtures of kernels
S. Ghoshal Fast Translation-Invariant Cycle-Spinning for Multi-Scale Image Denoising Methods

10:30 - 11:00 – Coffee Break

11:00 - 12:30 – Invited Session :Model misspecification **Chair**: C. Holmes
J. Miller Several interpretations of the power posterior
R. Martin Model misspecification on purpose
S. Walker Good and Bad Misspecified Models

12:30 - 14:00 – Lunch

14:00 - 15:30 – Invited Session : Bayesian Nonparametrics for Improved Decision Making **Chair**: W. Johnson
M. Guindani A Spike-and-Slab prior process for Dynamic model selection
A. Guglielmi Statistical analysis of recurrent events for improved health care
P. Mueller A Bayesian Nonparametric Utility-Based Design for Comparing Treatments to Resolve Air Leaks After Lung Surgery

15:30 - 16:00 – Coffee Break

16:00 - 17:30 – Contributed (ENS): Graphs **Chair**: S. Williamson
X. Miscouridou Exchangeable Random Measures for Sparse and Modular Graphs with Overlapping Communities
B. Bloem-Reddy Nested urn models of random partitions and graphs
D. Cai Edge-exchangeable graphs, sparsity, and power laws

16:00 - 17:30 – Contributed (IHP) : BNP & EFab **Chair**: A. Norets
M. Jensen Predicting Mortage Defaults when Data or Defaults are Sparse
L. Liu Density forecasts in panel data models: a semiparametric Bayesian perspective
J. Griffin Bayesian time-varying nonparametric vector autoregression modelling

19:30 - 23:00 – Conference Diner

Friday

9:30 - 10:30 – Plenary Session **Chair**: S. Petrone
Peter Orbanz Network models and symmetry

10:30 - 11:00 – Coffee Break

11:00 - 12:30 – Invited Session : Bayesian nonparametric modeling of networks and relational data **Chair**: F. Caron
V. Veitch Sampling and Sparse Exchangeable Graphs
D. Durante Nonparametric Bayes Modeling of Populations of Networks
I. Antoniano-Villalobos A Nonparametric Latent Space Model for Dynamic Relational Networks

Invited Sessions

Ryan Adams, (Harvard)

Building Probabilistic Structure into Massively Parameterized Models
Scientific applications of statistics and machine learning often require the identification of interpretable structure from high-dimensional observations. Bayesian nonparametric models provide an elegant and flexible way to capture such structure and let the data determine the balance between simplicity and complexity. It is often challenging, however, to couple these appealing models for latent structure to the massively parameterized likelihoods that are required for high-dimensional data such as video streams. I view this challenge through the lense of semiparametric modeling, in which a massively-parameterized function approximator is coupled to a compact and interpretable probabilistic model that may itself be nonparametric. Of particular interest in this vein is the merging of deep neural networks with graphical models containing latent variables, which enables each of these powerful components to play to its strengths. I will discuss several different classes of such models, and various applications.

Isadora Antoniano-Villalobos, (Bocconi University)

A Nonparametric Latent Space Model for Dynamic Relational Networks
Network and relational data arise in a variety of fields. They constitute the source of extensive literature and, specially in recent years, an active area of research among statisticians. One of the most popular approaches for the study of this type of data is based on a latent variable representation, justified under the assumption of row-column exchangeability for the error terms in a model which, in the general case, incorporates additional information through covariates. We present a model for Bayesian inference on dynamic network data which extends the static latent variable representation by introducing an infinite hidden Markov process driving the temporal evolution within the latent space. The model can be interpreted in terms of a global time-varying phenomena affecting the dynamics of the network while maintaining, at each point in time, the desirable properties provided by latent distance representations for static networks. The evolution can be monitored via the Network's edge density and other structural quantities of interest. The resulting model has a complex infinite-dimensional structure which makes it intractable. We therefore propose a computational algorithm that enables MCMC posterior inference.

Trevor Campbell, (CSAIL, MIT)

Exchangeable Trait Allocations
Clustering requires placing data into mutually exclusive groups, while feature allocations allow each datum to exhibit binary membership in multiple groups. But often binary membership is overly restrictive, as data points can exhibit different levels of belonging in different groups. We refer to the resulting combinatorial structure as a "trait allocation." In this talk, we will examine representations and characterizations of exchangeable trait allocations, where the

probability of any particular allocation is invariant to reordering the data sequence. We will develop trait allocation analogues of the Kingman paintbox and exchangeable partition probability function (EPPF) from clustering, which have been invaluable tools for analysis and the development of practical inference algorithms; we refer to our new constructions as the "trait paintbox" and the "exchangeable trait probability function" (ETPF). We show the first direct connection between probability functions for clustering, feature allocations, and trait allocations—along with the first full characterization of allocations involving single-occurrence "dust" groups. As an application of our general theory, we characterize the distribution of all edge-exchangeable graphs, a recently-developed model that captures realistic sparse graph sequences.

Antonio Canale, (University of Padua)

Multiscale mixtures of kernels
There is a considerable interest in multiscale nonparametric density estimation. The Bayesian nonparametric literature is however dominated by single scale methods, with the exception of Polya trees, which favour overly-spiky densities even when the truth is smooth. In this talk I describe a multiscale mixture model which produce smooth realizations that do not rely on hard partitioning of the support. The basic tool is a multiscale generalization of the well-known stick-breaking process. The latter generates stochastically decreasing set weights that are placed at each node of an infinitely-deep binary tree. The model is then completed by placing, at each level of the binary tree, a fixed dictionary of densities; Motivated by problems of computational scalability to high-dimensional cases, I introduce a method to learn the set of dictionary densities on a single scan of the data. Using this procedure along with the multiscale stick-breaking leads to a hybrid frequentist-Bayesian approach which scale well to high dimensional problems and provides uncertainty characterization.

Daniele Durante, (Department of Statistical Sciences, University of Padova)

Nonparametric Bayes Modeling of Populations of Networks
Many fields of research provide increasingly complex data along with novel motivating applications and new methodological questions. In approaching these data sets it is fundamental to rely on parsimonious representations which make the problem tractable and provide interpretable inference procedures to draw meaningful conclusions. However, in reducing complexity, it is important to avoid restrictive models that lead to inadequate characterization of relevant patterns underlying the observed data. Within this framework, network data representing relationship structures among a set of nodes are a relevant example. Although there has been abundant focus on models for a single network, there is a lack of methods for replicated network-valued data collected from a common population distribution. These data open new avenues for studying underlying connectivity patterns, how they are distributed in the population and if this distribution changes with predictors of interest. Motivated by neuroscience applications, this talk will discuss some issues associated with the available statistical models and will outline a possible solution to cover some of the current gaps via Bayesian nonparametric models leveraging latent space representations.

Barbara Engelhardt, (Princeton University)

Bayesian nonparametrics in the wild: Opportunities and challenges in genomic analyses
As technologies for generating genomic data proliferate, statistical approaches to analyze these data need to adapt to keep pace. Bayesian nonparametric models provide a framework for building these adaptive analytic tools because of their ability to grow as the data grow, their hierarchy that allows arbitrary structure, their flexibility of functional and distributional forms of the observations, and their explicit characterization of uncertainty. I will discuss examples

of genomic data generated from recent technologies where we have used nonparametric models to address analytic challenges. For each approach, robust inference, crisp interpretation, and mechanistic validation of the results are critical to the utility of the model in these applications.

Nicholas Foti, (University of Washington)

Variational Boosting: Iteratively Refining Posterior Approximations
The central problem of a Bayesian analysis is determining the posterior distribution of the model parameters. Variational Bayes formulates this problem as finding the best approximation to the posterior from amongst a family of simpler distributions. This optimization perspective has the advantage over Monte Carlo methods of knowing when the best approximation has been found. However, many of the approximating families used in practice are too simple, resulting in overly certain marginal posteriors and inaccurate posterior correlations between parameters. Furthermore, the true posterior distribution is rarely contained in the approximating family. We propose a scalable black-box variational inference method to approximate intractable posterior distributions with an increasingly rich approximating class. The method, termed variational boosting, iteratively refines an existing variational approximation by solving a sequence of optimization problems, allowing practitioners to trade computation time for more accurate approximations. We show how to sequentially expand the approximating class of distributions by incorporating additional covariance structure and by introducing new components to form a mixture. We apply variational boosting to a variety of Bayesian models on synthetic and real data, demonstrating that the recovered posterior distributions are as accurate as those produced by existing methods.

Subhashis Ghoshal, (North Carolina State University)

Fast Translation-Invariant Cycle-Spinning for Multi-Scale Image Denoising Methods
Image denoising methods commonly exhibit staircase-like artifacts due to the discrete nature of pixeleted images. A simple remedy is to cycle-spin to overcome these effects, where the image is shifted, denoised and then shifted back to the original position. The final image is computed by averaging over all such cycle-spins and is the called translation-invariant (TI) version of the denoising method. If n is the number of pixels, a naive approach to computing TI version requires $O(n^2)$ computations, which is prohibitive. In Poisson and Gaussian noised images, the likelihood admits a multi-scale decomposition. We show that for any denoising method based on the multi-scale decomposition of the likelihood such as Bayesian methods, the TI version can be computed in just $O(n \log(n))$ time, which is only slightly higher than the $O(n)$ time taken by the method without any cycle-spinning. We also introduce the notion of k-TI, where only the k finest levels in the multi- scale decomposition are cycle-spinned. By extensive simulations, we show that cycle-spinning greatly enhances the quality of image processing and that even for moderate values of k, the k-TI version performs nearly as well as the TI-version but can be computed at a fraction of the cost the full TI version.

Jim Griffin, (University of Kent)

Bayesian time-varying nonparametric vector autoregression modelling
Vector autoregressions (VARs) are the main method used by macroeconomists for modelling multivariate, economic time series data. The models are simple to interpret and fit but their assumptions of stationarity, linearity and normality are often too restrictive. Some of these shortcomings can be addressed using time-varying VARs which allow the parameters of the VAR to change over time. However, these models allow for non-linearity, non-normality and non-stationarity using a single process which may be inappropriate. In this talk, a time-varying Bayesian nonparametric VAR will be developed which combines a nonparametric stationary VAR model with a time-varying Dirichlet process

prior. This allows us to separate local non-linearity and non-normality (which are modelled through the nonparametric stationary VAR model) from non-stationarity (which is modelled through the time-varying nonparametric prior). The model will be illustrated and compared to competing models on UK and US economic data.

Alessandra Guglielmi, (Politecnico di Milano)

Statistical analysis of recurrent events for improved health care
We consider recurrent event data in the health care context, where often a terminating event such as death can occur during the follow-up period precluding further occurrence of the recurrent events. Specifically, we propose a BNP joint model for gap times between two consecutive hospitalizations and survival time for heart failure (HF) patients. HF is one of the main causes of morbidity, hospitalization and death in the western world and the economic burden associated with HF management is relevant and expected to increase in the future. Here the main clinical outcome of interest is time to death and research focus is on investigating how recurrent hospitalizations or potential explanatory factors affect the time to event. This is an important feature of our approach as several recurrences (in our case re-hospitalizations) are often related to risk of death and are likely to affect it. The strength of the association between recurrences and terminal event may then be interpreted in terms of patients' risk profiling, and a better understanding of how recurrences affect survival may lead to a more effective planning of healthcare resources. Our model also accounts for patient heterogeneity in recurrent event trajectories; covariates in the survival and/or recurrence process may be included through the specification of appropriate regression terms. Data on hospitalizations for HF were extracted from the administrative data warehouse of the regional healthcare system of Lombardia, the most populated Italian Region.

Michele Guindani, (University of California, Irvine)

A Spike-and-Slab prior process for Dynamic model selection
We describe a novel Bayesian nonparametric approach for the analysis of spatio-temporal data. We assume the observed measurements to possibly arise from models that are allowed to change over space and time. We employ flexible species sampling priors which, through proper definition of the weights, enable spatio-temporal model selection. In addition, our modeling framework allows for clustering of the observations over time. Posterior inference is carried out through MCMC methods. We assess the performance of our approach by means of a simulation study. We also illustrate an application to a disease surveillance dataset.

Nils Hjort, (University of Oslo)

Empirical likelihood and Bayesian nonparametrics
Suppose F is an unknown distribution and $\theta = \theta(F)$ a focus parameter, an aspect of particular interest. One method for reaching nonparametric Bayesian inference for θ, based on a sample from F, is to start with a prior for F, derive its posterior, and then compute the posterior for θ, perhaps via simulation. Another strategy is via the empirical likelihood for θ and a prior for θ, bypassing the direct need to operate with a prior for the full F. I will explore these two avenues and prove that they often will lead to the very same inference, for large samples. This will also lead to two Bernstein–von Mises theorems.

Chris Holmes, (University of Oxford)

Nonparametric tests

Hypothesis testing, such as two-sample tests and tests for independence, are important components of biomedical data analysis. We discuss recent advances in the use of Bayesian nonparametric (BNP) models for this task. One advantage of Bayesian nonparametric methods is that they provides an explicit model based probability measure for the null hypothesis $Pr(H_0|data)$, as well as for the alternative, $Pr(H_1|data) = 1 - Pr(H_0|data)$, without making strong assumptions on the form of the likelihood function. This allows one to incorporate prior information and control for multiple testing, while naturally extending to tests for contrasts where we wish to detect for an effect under one condition but not under another. This latter task is problematic for non-Bayesian methods as one of the tests has a higher dimensional null than the alternative, but is readily accommodated in the Bayesian approach. We present some examples from biomedical genomics including recent results on survival analysis under right-censored data. This is joint work with Sarah Filippi and Judith Rousseau.

Lancelot James, (Hong Kong University of Science and Technology)

q-Hierarchical Latent Feature Models
In this talk we describe the notion of q-Hierarchical Feature Models which can be viewed as generalizations of the Indian Buffet Process. In the special case of $q = 1$, this includes the class of Hierarchical Indian Buffet processes. When $q = 2$, models exhibiting behavior similar to the model of Caron and Fox arise. When $q \geq 2$ these models correspond to as of yet unexplored feature models.We point out some of the obstacles inherent in efficiently sampling these models due to the use of completely random measures. We describe modifications which considerably simplify this.

Alejandro Jara, (Department of Statistics, Pontificia Universidad Católica de Chile)

Marginal Bayesian Semiparametric Modelling of Mismeasured Multivariate Interval-Censored Data
Motivated by data gathered in an oral health study, we propose a Bayesian nonparametric approach for population-average modeling of correlated time-to-event data, when the responses can only be determined to lie in an interval obtained from a sequence of examination times and the determination of the occurrence of the event is subject to misclassification. The joint model for the true, unobserved time-to-event data is defined semiparametrically; proportional hazards, proportional odds, and accelerated failure time (proportional quantiles) are all fit and compared. The baseline distribution is modeled as a flexible tailfree prior. The joint model is complete by considering a parametric copula function. A general misclassification model is discussed in detail, considering the possibility that different examiners were involved in the assessment of the occurrence of the events for a given subject across time. We provide empirical evidence that the model can be used to estimate the underlying time-to-event distribution and the misclassification parameters without any external information about the latter parameters.

Mark Jensen, (Federal Reserve Bank of Atlanta)

Predicting Mortage Defaults when Data or Defaults are Sparse
Some of the bank holding companies (BHC) subject to the Federal Reserve's annual Comprehensive Capital Analysis and Review (CCAR) have portfolios that have either never experienced any defaults or is comprised of newer loans that have yet to experience enough actual defaults to accurately predict the probability a loan will default in the future. We propose a quantitative model that leverages the default histories found in other BHC's loan portfolios to estimate how probability of default is distributed over the entire population of BHC. This population distribution of the probability of defaulting is the best quantitative prediction of a loan defaulting when it is held by a bank that has

not experienced any defaults. In essence we borrower from the default histories found at other banks to predict the default probability at banks that are similar but that have a sparse history of defaults.

James Johndrow, (Stanford University)

MCMC for large scale problems

An important motivation for Bayesian analysis in large scale (big p and/or n) settings is uncertainty quantification via the posterior distribution. MCMC provides a default computational algorithm, but it can be unacceptably slow when n or p are large. Computational complexity of MCMC is a function of two factors: the cost of one step from the Markov kernel and the rate at which the spectral gap converges to zero in n and p. This talk considers strategies for reducing the computational complexity of MCMC by addressing both of these factors, focusing on theoretical results. Commonly used MCMC algorithms for nonparametric models, including huge contingency tables and nonparametric regression, provide important motivating examples. New algorithms are proposed that have superior performance in high dimensional settings where existing MCMC performs poorly.

Fabrizio Leisen, (University of Kent)

Some recent constructions of vectors of dependent completely random measures

In this talk, I will illustrate some recent models for dependent completely random measures I introduced with A. Lijoi, J. Griffin, D. Spano and W.Zhu. In particular, I will focus on Levy copula constructions, Compound Random Measures (CoRM) and their relations. Applications to Survival Analysis will be discussed.

Antonio Lijoi, (Bocconi University)

Posterior representations of dependent random probabilities

Dependent random probability measures are appealing and effective tools for modeling non-exchangeable data. In this talk we consider different constructions that yield dependent models able to cope with density estimation and survival analysis problems, in the presence of partially exchangeable observations. We display some novel conditional representations that are the key for devising simulation algorithms for a full Bayesian inference. The comparison with standard algorithms based on suitable extensions of the Polya urn scheme will be illustrated through examples with real and simulated datasets.

Laura Liu, (University of Pennsylvania)

Density forecasts in panel data models: a semiparametric Bayesian perspective

This paper constructs unit-specific density forecasts for a panel of firms or households using a dynamic linear model with common and heterogeneous parameters and cross-sectional heteroskedasticity. The distribution of the heterogeneous coefficients is modeled nonparametrically, allowing for correlation between heterogeneous coefficients and initial conditions as well as unit-specific regressors. A benchmark for the density forecasts is the (infeasible) oracle forecast which is defined as the posterior predictive distribution for the unit-specific outcomes under the assumption that the common parameters and the distribution of the heterogeneous coefficients are known. I develop a simulation-based posterior sampling algorithm specifically addressing the nonparametric density estimation of unobserved heterogeneous coefficients. I prove that both the estimated common parameters and the estimated distribution of the heterogeneous coefficients achieve posterior consistency, and that the density forecasts asymptotically converge to the oracle forecast. Monte

Carlo simulations demonstrate improvements in density forecasts relative to parametric approaches. An application to young firm dynamics also shows that the proposed predictor provides more accurate density predictions

Li Ma, (Duke University)

Three techniques for multi-scale nonparametric modeling

Multi-scale models, such as the Polya tree and its friends, form a powerful toolbox for carrying out nonparametric inference in a computationally efficient manner while achieving analytic simplicity and theoretical guarantees. Inference under such models often enjoys a computational complexity that scales approximately linearly with the sample size and is easily parallelizable, making these models particularly promising in the face of big data. However, in practice several factors have substantially limited the popularity of such models in competition with mixture-based methods. First, the performance of traditional multi-scale models—in terms of prediction and estimation accuracy as well as sensitivity and power in testing—often lags behind mixture-based models. Second, existing development of multi-scale models in more complex modeling settings such as modeling conditional distribution (or density regression) sacrifices the most desirable aspect of such models—their computational efficiency. Third, hierarchical extensions to multi-scale models for characterizing related samples are generally lacking, to some extent due to the lack of scalable algorithms for inference. In this talk, I describe three recently developed techniques for multi-scale model construction and inference that address these difficulties: (i) adaptive stochastic shrinkage/smoothing, (ii) covariate-response nested partitioning, and (iii) INLA-based inference for hierarchical multi-scale models. They can be used in conjunction with each other. The computational and theoretical properties of traditional multi-scale models are preserved when they are employed.

Yasaman Maleki, (Alzahra University)

A Nonparametric Estimator of Covariance Function for Parameterized Family of Locally Self-Similar Processes

The optimal covariance function estimate, in the sense of mean-square of errors (MSE), for the class of discrete-time locally self-similar processes is accomplished from one observed realization by weighting observations with a kernel function. This paper investigates the MSE optimal kernel function for any parameterized family of locally self-similar processes by solving a system of linear equations. Furthermore, it is shown that the optimal kernel is close to optimal for all members of the family. We consider situations in which the scale invariant process, belongs to a parameterized family of stochastic processes with a priori parameter distribution. Say, some generalizations of scale invariant processes, such as locally self-similar processes, where the Hurst parameter may change in time. We show that the MSE optimal kernel, for such a parameterized family of scale invariant processes with a priori parameter distribution, is computed by solving a system of linear equations.

Ryan Martin, (North Carolina State University)

Model misspecification on purpose

Model misspecification is usually considered a bad thing. However, for various reasons (computational burden, robustness concerns, specification of priors, etc), it may be advantageous in some cases to purposely work with a misspecified model. The challenge is that not just any misspecified model will do. In this talk, I will present two specific examples, both nonparametric in their own way, where a purposely misspecified model leads to a posterior distribution having certain asymptotic optimality properties, along with some added benefits.

Jeff Miller, (Harvard School of Public Health)

Several interpretations of the power posterior

It is generally acknowledged that "all models are wrong", and even nonparametric models often involve parametric assumptions that can lead to misleading inferences when they are incorrect. Recently, there has been increased interest in using power posteriors to improve robustness to model misspecification. A power posterior is defined as proportional to the likelihood raised to a fractional power (between zero and one) times the prior. This surprisingly simple device has a number of attractive properties. In this talk, we describe several different ways to interpret and justify the power posterior from the coarsened posterior (c-posterior) perspective, providing intuition for why the method makes sense and why it works. We compare and contrast with other perspectives. Additionally, we present a new theoretical result on the approximation of c-posteriors by power posteriors.

Peter Mueller, (UT Austin)

A Bayesian Nonparametric Utility-Based Design for Comparing Treatments to Resolve Air Leaks After Lung Surgery

A Bayesian Nonparametric Utility-Based Design for Comparing Treatments to Resolve Air Leaks After Lung Surgery We propose a Bayesian nonparametric utility-based group sequential design for a randomized clinical trial to compare a gel sealant to standard care for resolving air leaks after pulmonary resection. Clinically, resolving air leaks in the days soon after surgery is highly important, since longer resolution time produces undesirable complications that require extended hospitalization. The problem of comparing treatments is complicated by the fact that the resolution time distributions are skewed and multi-modal, so using means is misleading. We address these challenges by assuming Bayesian nonparametric probability models for the resolution time distributions and basing the comparative test on weighted means. The weights are elicited as clinical utilities of the resolution times. The proposed design uses posterior expected utilities as group sequential test criteria. The procedure's frequentist properties are studied by computer simulation. If time permits we will briefly discuss another application of BNP to comparing treatments in the presence of semi-competing risks.

Richard Nickl, (University of Cambridge)

Bernstein von Mises theorems for non-linear inverse problems

We discuss recent results on exact semi- and non-parametric frequentist asymptotics for Bayesian posterior distributions arising in certain non-linear inverse problems. Examples include discretely sampled stochastic processes (such as Lévy processes and diffusions) as well as basic elliptic PDE problems in Gaussian white noise. The main insight is that in those applications, Bayes solutions give optimal asymptotically optimal and efficient inference procedures, and that for many interesting problems, associated Bayesian credible sets are actually valid frequentist confidence sets.

Peter Orbanz, (Columbia University)

Network models and symmetry

The fundamental theorem of Bayesian statistics is arguably de Finetti's theorem. It concerns exchangeability, a symmetry property. Other symmetry properties exist, but until recently, only exchangeability seemed particularly relevant to statistics. The picture has changed since interest has broadened from data represented as a sequence to graphs, networks, and other discrete structures. In network models, there are various symmetries; some are closely related to exchangeability, others are not. I will review what such properties mean for statistical modeling, and highlight facts we have learned in the last few years: Various forms of symmetry in network models; how they explain inference from a single, growing graph; and what symmetry implies for convergence rates.

Art Owen, (Stanford University)

A survey of Bayesian empirical likelihood
In Bayesian inference we specify a prior and a likelihood and then use whatever computational tools we can to extract scientific insight from the posterior distribution so formed. Both choices, prior and likelihood, can be challenging to choose. There may be a clear choice, but more often than that, we have only qualitative information. TheSlog-iSlogi convenient choices for these factors may imply some unwanted consequences. For instance, the data may be overdispersed compared to the likelihood we choose. We can reduce some of these unwanted consequences by choosing noninformative, or weakly informative priors.The analogous choice on the likelihood side is the empirical likelihood. This talk will survey work using the empirical likelihood in Bayesian applications, beginning with work by Lazar (2003) and including the Bayesian exponentially tilted empirical likelihood of Schennach (2005). The motivation for revisiting this problem is the study of RNA-seq data. This data is overdispersed relative to the Poisson distribution and so the negative binomial distribution is widely used. But the data need not be negative binomial either.

Debdeep Pati, (Florida State University)

Community detection and goodness of fit tests in random graph models: a probabilistic approach
In this talk, I focus on two key problems in random graph models. The first part of the talk deals with clustering the nodes of a network into groups which share a similar connectivity pattern. Existing algorithms for community detection assume the knowledge of the number of clusters or estimate it a priori using various selection criteria and subsequently estimate the community structure. Ignoring the uncertainty in the first stage may lead to erroneous clustering, particularly when the community structure is vague. I instead propose a coherent probabilistic framework for simultaneous estimation of the number of communities and the community structure. The methodology is shown to outperform recently developed community detection algorithms in a variety of synthetic data examples and in recovering activation regions from brain connectivity matrices. In addition, I derive minimax optimal bounds for the Bayes risk, which is novel in the Bayesian context to best of my knowledge. In the second part of the talk, I focus on deriving goodness of fit tests when the number of communities is unknown. The test relies on algebraic geometric techniques, which can potentially be generalized to construct goodness of fit tests in latent class models.

Rajesh Ranganath, (Princeton University)

Variational Inference: Modern Methods
Variational inference is undergoing a renaissance. Classically, deploying variational inference requires pages of model-specific analysis. This barrier limits variational inference's ability to help quickly explore models for the data being studied. To address this, I will describe recent advances that simplify the use of variational inference. I will begin with black box variational inference (BBVI). BBVI is based on stochastic optimization of the KL divergence and can be applied to new models with little model specific work. In addition to making variational inference easier to use, BBVI has opened the door to richer families of variational approximations. In the second part of the talk, I will describe a class of variational approximations called hierarchical variational models (HVM). Hierarchical variational models place priors on the parameters of existing variational approximations, such that it captures complex structures for both discrete and continuous latent variables. I will describe two HVMs: one based on normalizing flows and the other on Gaussian processes. Along the way, I will describe uses and open questions for variational inference in Bayesian nonparameteric models.

Boyu Ren, (Harvard University)

Fast Bayesian nonparametric method for detecting alternations in species distribution
In the context of authorship determination, Thisted and Efron (1987) proposed a nonparametric empirical Bayes model to estimate observed recurrent frequencies of rare words in a new article and test whether those are consistent with the habitual vocabulary of the author. However, their estimator requires complicated smoothing when the words are not extremely rare in the existing corpus. We propose a new Bayesian nonparametric method based on species sampling model with Gibb-type priors to overcome this difficulty. This method yields closed-form expressions for the moments as well as the probability mass functions of the posterior predictive distributions of words' recurrent frequencies, which avoid the necessity of posterior simulations. The point estimates given by the means of these predictive distributions are well defined and can be treated as counterparts of Thisted's empirical Bayes estimator, without the need of ad hoc smoothing. By using the full posterior predictive distributions, we are able to replace the Poisson regression model in their work with a new testing procedure for authorship determination. We show with simulation studies that this testing procedure has better performance especially when the Poisson model implemented in their work is not valid. This method is further applied on a microbiome dataset to determine the change points of microbial community profile of rare species in an unsupervised manner. The model correctly identifies the time points which corresponds to the major changes of diet and antibiotic usage.

Veronika Rockova, (University of Chicago (Booth School of Business))

Particle EM for Variable Selection
Despite its long history of success, the EM algorithm has been vulnerable to local entrapment when the posterior/likelihood is multi-modal. This is particularly pronounced in spike-and-slab posterior distributions for Bayesian variable selection. The main thrust of this article is to introduce the Particle EM algorithm, a new population-based optimization strategy that harvests multiple modes in search spaces that present many local maxima. Motivated by non-parametric variational Bayes strategies, Particle EM achieves this goal by deploying an ensemble of interactive repulsive particles. These particles are geared towards uncharted areas of the posterior, providing a more comprehensive summary of its topography than simple parallel EM deployments. A sequential Monte Carlo variant of Particle EM is also proposed that explores a sequence of annealed posteriors by sampling from a set of mutually avoiding particles. Particle EM outputs a deterministic reconstruction of the posterior distribution for approximate fully Bayes inference by capturing its essential modes and mode weights. This reconstruction reflects model selection uncertainty and is supported by asymptotic considerations, which indicate that the requisite number of particles need not be large in the presence of sparsity (when $p \geq n$).

David Rossell, (Universitat Pompeu Fabra)

Choice of mixture components via non-local priors
We consider the problem of choosing the number of components in a mixture model from a Bayesian testing point of view. Our proposal is built on a non-local prior (NLP) specification that induces a separation between probability models. This is very natural in mixtures. Conditional on a model saying that there are k components, the components cannot have zero weights and must be minimally separated from each other, otherwise the model would correspond to e.g. k-1 components. We provide some theory showing that NLPs lead to stronger parsimony to discard spurious components, and a construction that is convenient from a computational point of view. Our empirical findings agree with recent theoretical developments in the algebraic statistics literature, namely that default model choice criteria such as the BIC can present a marked lack of sensitivity to detect components. We also show how, by ensuring that the chosen model has well-separated components with non-negligible weight, the solution tends to be more interpretable

Reza Solgi, (Harvard University, Statistics Department)

Nonparametric hierarchical Bayesian quantiles

We develop a method for performing nonparametric Bayesian inference on quantiles. Relying on geometric measure theory and employing a Hausdorff base measure, we are able to specify meaningful priors for the quantile while treating the distribution of the data otherwise nonparametrically. We further extend the method to a hierarchical model for quantiles of subpopulations, linking subgroups together solely through their quantiles. Our approach is computationally straightforward, allowing for censored and noisy data. We demonstrate the proposed methodology on simulated data and an applied problem from sports statistics, where it is observed to stabilize and improve inference and prediction.

Botond Szabo, (Leiden University)

Frequentist coverage properties of Bayesian credible sets for sieve priors in general settings

One of the main advantages of Bayesian methods is that they readily provide an uncertainty statement of the statistical procedure. This uncertainty is visualized by plotting the (typically) 95% credible sets, i.e. sets which accumulate 95% of the posterior mass. The question naturally arises whether these sets can also be used for uncertainty quantification from a frequentist point of view or by doing so one provides haphazard, missleading uncertainty statements. In this work we investigate the coverage properties of credible sets in general settings derived from both the hierarchical and empirical Bayes methods using sieve type of priors. We show that under mild assumptions (in the spirit of Ghoshal and van der Vaart (2007)) and the so called "general polished tail" condition sieve priors will produce adaptive and honest confidence sets. We apply this abstract results to various specific examples, including the nonparametric regression model, density estimation using exponential families of priors, density estimation using histogram priors and nonparametric classification model.

Matt Taddy, (University of Chicago and Microsoft)

Counterfactual Prediction with Deep Instrumental Variables Networks

We are in the middle of a remarkable rise in the use and capability of artificial intelligence. Much of this growth has been fueled by the success of deep learning architectures: models that map from observables to outputs via multiple layers of latent representations. These deep learning algorithms are effective tools for unstructured prediction, and they can be combined in AI systems to solve complex automated reasoning problems. This paper provides a recipe for combining ML algorithms to solve for causal effects in the presence of instrumental variables – sources of treatment randomization that are conditionally independent from the response. We show that a flexible IV specification resolves into two prediction tasks that can be solved with deep neural nets: a first-stage network for treatment prediction and a second-stage network whose loss function involves integration over the conditional treatment distribution. This Deep IV framework imposes some specific structure on the stochastic gradient descent routine used for training, but it is general enough that we can take advantage of off-the-shelf ML capabilities and avoid extensive algorithm customization. We outline how to obtain out-of-sample causal validation in order to avoid over-fit. We also introduce schemes for both Bayesian and frequentist inference: the former via a novel adaptation of dropout training, and the latter via a data splitting routine.

Surya Tokdar, (Duke University)

airGP: Bayesian Smoothing in High Dimension

Regression smoothing in high and ultra high dimensions presents formidable theoretical and computational challenges. The theory of additive-interactive regression offers a practicable framework on which we build a Bayesian estimation technique by using hierarchical Gaussian process priors. Designing efficient, reproducible and scalable algorithms for posterior computation remains a challenge for this model. The main difficulty is designing a Markov chain sampler on the space of an ensemble of high-dimensional binary inclusion vectors. I will present recent advancements on this by the use of various adaptation techniques and introduce our upcoming R package 'airGP'.

Lorenzo Trippa, (Harvard, Dana-Farber Cancer Institute)

Bayesian Nonparametric Ordination for the Analysis of Microbial Communities
Human microbiome studies use sequencing technologies to measure the abundance of bacterial species or Operational Taxonomic Units (OTUs) in samples of biological material. Typically the data are organized in contingency tables with OTU counts across heterogeneous biological samples. In the microbial ecology community, ordination methods are frequently used to investigate latent factors or clusters that capture and describe variations of OTU counts across biological samples. It remains important to evaluate how uncertainty in estimates of each biological sample's microbial distribution propagates to ordination analyses, including visualization of clusters and projections of biological samples on low dimensional spaces. We propose a Bayesian analysis for dependent distributions to endow frequently used ordinations with estimates of uncertainty. A Bayesian nonparametric prior for dependent normalized random measures is constructed, which is marginally equivalent to the normalized generalized Gamma process, a well-known prior for nonparametric analyses. In our prior the dependence and similarity between microbial distributions is represented by latent factors that concentrate in a low dimensional space. We use a shrinkage prior to tune the dimensionality of the latent factors. The resulting posterior samples of model parameters can be used to evaluate uncertainty in analyses routinely applied in microbiome studies. Specifically, by combining them with multivariate data analysis techniques we can visualize credible regions in ecological ordination plots. The characteristics of the proposed model are illustrated through a simulation study and applications in two microbiome datasets.

Aad van der Vaart, (Leiden University)

Credible sets for sparse models
We consider uncertainty quantification by a posterior distribution in the situation that there are many parameters, most of which are thought to be zero. The prior distribution will model the presumed sparsity and the corresponding posterior distribution will then shrink many parameters to zero. We consider both marginal credible intervals for single parameters and a credible ball for the full parameter vector. The most interesting case is that the sparsity level of the prior, and the resulting shrinkage of the posterior, is set by either equipping it with a hyper prior or by an empirical Bayes method. General theory implies that credible sets, of either type, cannot give a correct uncertainty quantification for every parameter in this case. We discuss for which parameters the procedures are reliable, with perhaps the most interesting conclusion that marginal credible intervals that are bounded away from zero are essentially correct in indicating a nonzero effect.

Victor Veitch, (University of Toronto)

Sampling and Sparse Exchangeable Graphs
Recent work has introduced sparse exchangeable graphs and the associated graphex framework, as a generalisation of dense exchangeable graphs and the associated graphon framework. The sparse models retain the essential structure that makes graphon models practically appealing, while also allowing us to model a much larger range of network

behaviour. However, the interpretation and applicability of the new model class is difficult. This is because these models are defined by a somewhat cryptic construction that represents random graphs as point processes on the real plane, and then formalizes the models of the generalized framework as those corresponding to point processes that are exchangeable. We identify the sampling scheme on networks induced by this notion of exchangeability—a modification of independent vertex sampling that we call p-sampling. We show that invariance under this sampling scheme is equivalent to exchangeability, thereby giving an easily interpretable characterization of the models. We further make use this sampling notion to identify a general non-parametric estimator for the graphex parameter underlying the models; this is a network version of the empirical measure. (this covers work with Daniel Roy, and also with Christian Borgs, Jennifer Chayes, and Henry Cohn)

Stephen Walker, (University of Texas at Austin)

Good and Bad Misspecified Models

Misspecified models are the norm and yet difficult to understand by their very nature. They will behave quite differently depending on where observations are actually coming from. In this talk I will show they can lead to both failure and success. For the former one can be moving further away from from the true density after every Bayesian update and for the latter a purposefully designed misspecified model actually leads to the right answer.

Sinead Williamson, (University of Texas at Austin)

Restricted Indian buffet processes

Latent feature models are a powerful tool for modeling data with globally-shared features. Nonparametric distributions over exchangeable sets of features, such as the Indian Buffet Process, offer modeling flexibility by letting the number of latent features be unbounded. However, current models impose implicit distributions over the number of latent features per data point, and these implicit distributions may not match our knowledge about the data. In this work, we demonstrate how the Restricted Indian Buffet Process circumvents this restriction, allowing arbitrary distributions over the number of features in an observation. We discuss several alternative constructions of the model and apply the insights to develop Markov Chain Monte Carlo and variational methods for simulation and posterior inference.

Dan Zhu, (Department of Econometrics and Business Statistics, Monash University)

Automated Sensitivity Computations for MCMC Gibbs Output

MCMC algorithms enable Bayesian researchers to evaluate the posterior distribution of the parameters which depends on a potentially large set of prior inputs in the complex manner. Each MCMC algorithm is a mapping based on the posterior distribution, which converts the information from the data via the model likelihood and a set of parameters of the prior distribution and the starting values into an estimate of the posterior distribution. In this paper we introduce the first general numerical approach to compute the sensitivity of the statistics computed based on the output of an MCMC algorithm with respect to all prior inputs (including prior means, variances and starting values) via embedded sensitivity algorithms based on automatic differentiation methods. Our approach presents the first general numerical method, and first methods to compute sensitivities with respect to all prior inputs. The methods can be applied even when the bumping method does not work (i.e. gamma update) and also and when the analytical differentiation approach (Mueller) for the computation of sensitivities of posterior means with respect to prior mans fails, i.e. outside Exponential family.

Contributed and Posters

Joseph Antonelli, (Harvard T.H. Chan School of Public Health)

A flexible tensor regression approach to estimating the effect of chemical mixtures on birth weights
Humans are exposed to a complex mixture of chemicals and pollutants, and estimating the health effects of these mixtures is of crucial importance in environmental epidemiology. It is expected that the different pollutants we are exposed to interact with each other in a complex and nonlinear manner. To estimate the health effects of complex mixtures we use tensor regression, which uses a low-rank tensor decomposition to estimate a complex, nonlinear function using simple, local functions. We also induce sparsity using spike and slab priors on the local functions that make up the tensor factorization. Using the indian buffet process, we let the number of rank-one tensor products required to estimate the health effects of multiple pollutants be unknown. On top of it's flexibility, the proposed approach is interpretable as we can use the posterior samples to identify pollutants that interact with each other, which is of great importance in environmental research. We illustrate our approach's ability to estimate complex functions using simulated data, and apply our method to a study of metal pollutants and their subsequent effect on birth weight.

Julyan Arbel, (Inria)

Sequential Quasi Monte Carlo for Dirichlet Process Mixture Models
In mixture models, latent variables known as allocation variables play an essential role by indicating, at each iteration, to which component of the mixture observations are linked. In sequential algorithms, these latent variables take on the interpretation of particles. We investigate the use of quasi Monte Carlo within sequential Monte Carlo methods (a technique known as sequential quasi Monte Carlo) in nonparametric mixtures for density estimation. We compare them to sequential and non sequential Monte Carlo algorithms. We highlight a critical distinction of the allocation variables exploration of the latent space under each of the three sampling approaches.

Andrea Arfè, (Bocconi University)

Reinforced urns and the subdistribution beta-Stacy process prior for competing risks analysis
In clinical prognostic research with a time-to-event end-point, the occurrence of one of several competing risks often precludes the occurrence of another event of interest. Classical approaches to prediction in presence of competing risks focus on the subdistribution function, which represents the probability that a specific event occurs within a given time period. In this work, we introduce the subdistribution beta-Stacy process, a novel Bayesian nonparametric process prior for subdistribution functions useful for the analysis of discrete-time competing risks data. In particular, we i) characterize this process from a predictive perspective by means of an urn model with reinforcement, ii) show that it is conjugate with respect to right-censored data, and iii) highlight its relations with other prior processes for competing

risks data. Additionally, we consider the subdistribution beta-Stacy process prior in a nonparametric regression model for discrete-time competing risks data, with application to the prognosis of HIV-infected men from the Amsterdam Cohort Studies.

Raffaele Argiento, (University of Torino)

Normalized almost sure finite point processes for mixture models.
Modelling via finite mixtures is one of the most fruitful Bayesian approach, particularly useful when there is unobserved heterogeneity in the data. The most popular algorithm under this approach is the reversible jump MCMC, that can be nontrivial to design, especially in high-dimensional spaces. In this work we will show how nonparametric methods can be transferred into the parametric framework. We first introduce a class of almost sure finite discrete random probability measures obtained by normalization of finite point processes. Then, we use the new class as mixing measure of a mixture model and derive its posterior characterization. The resulting class encompasses the popular finite Dirichlet mixture model. In order to compute posterior statistics, we propose an alternative to the reversible jump: borrowing notation from the nonparametric Bayesian literature, we set up a conditional MCMC algorithm based on the posterior characterization of the unnormalized point process. To discuss the performance of our algorithm and the flexibility of the model, we illustrate some examples on simulated and real data.

Douglas Armstrong, (South Dakota State University)

Nonparametric methods for model selection problems arising in forensic evidence interpretation
The inference of source for complex, high-dimensional trace evidence in forensic science is a difficult but necessary task. Many evidence forms, such as fingerprints, very small particles, and toolmarks, are difficult to handle and often impossible to assign a probability distribution in its natural feature space. One common approach to circumvent this problem is to reduce the data dimension by assigning pairwise similarity scores between pairs of objects that compose the evidence. These methods are typically used to assign probative values from a univariate distribution of the scores and are now well-known as ad-hoc approaches to forensic source identification. In this presentation we will summarize recent advances made on kernel-based models for forensic source identification. This recent approach has moved to look at the entire kernel space of observations in an analogous manner to support vector machines. The development of this model is based on capturing the dependencies between pairwise scores from a hierarchical sample and modelling them in the kernel space using a linear model. The linear model has a covariance matrix that captures the dependency structure inherent to the pairwise scores and therefore allows the calculation of likelihoods for trace evidence given what we know about all the control sources and samples. Ultimately, this allows for the use of a parametric model for use in a non-parametric data to calculate a likelihood ratio for a prosecution (specificity) model against a defense (typicality) model to assign the probative value of evidence.

Morgane Austern, (Columbia University)

Asymptotic normality and rates for exchangeable structures
Exchangeability is a cornerstone of Bayesian statistics, and two of its statistical consequences are well known: Representation theorems (of de Finetti, Kingman, Aldous-Hoover, etc) and laws of large numbers. A representation theorem explains the data as drawn from a randomly chosen distribution P; the associated law of large numbers then guarantees that suitable sample averages of functions of the data converge to the expectation of that function under P. We show that stronger properties, namely asymptotic normality of such averages and bounds on the rate of convergence, also follow in a similarly generic fashion: Laws of large numbers are known to hold for a function f of the data if

the expectation of f exists. If the second moment also exists, we give conditions under which a central limit theorem holds. Previously known special cases of this result include Hans Buehlmann's central limit theorem for exchangeable sequences, and the asymptotic normality of certain random graph functionals established by Bickel, Chen and Levina [Ann Statist 39(2011)38-59]. Additional conditions on the third and fourth moments yield bounds on the rate of convergence that resemble the Berry-Esseen theorem. From exchangeability, the results generalize to other forms of invariance (that is, invariance under actions of other groups than the symmetric group). In these general results, a mixing condition reminiscent of mixing in time series arises naturally.

Fadhel Ayed, (Department of Statistics, University of Oxford)

Bayesian Nonparametric Poisson factorization with completely random measures
Recommender systems aim at predicting which items a user may like based on a history of ratings or purchases. In this work, we study Bayesian Nonparametric Poisson factor models for recommendation systems. We assume that each user/item has an infinite set of latent features, which are modeled using completely random measures. The model is flexible and allows the number of active features to grow unboundedly with the number of user/items. We show how the properties of the CRM relate to the growth of the number of active features. We also derive a Markov chain Monte Carlo algorithm for posterior inference for this class of models.

Cecilia Balocchi, (Department of Statistics, University of Pennsylvania)

Non-parametric Spatial Modeling of Changes in Crime over Time
Understanding the relationship between change in crime over time and the geography of urban areas is an important problem for urban planning. Accurate estimation of changing crime rates throughout a city would aid law enforcement as well as enable studies of the association between crime and the built environment. Bayesian hierarchical modeling is a promising direction since areal data, such as crime counts in neighborhoods over time, require principled sharing of information to address spatial autocorrelation between proximal locations. Available methods range from fully parametric models such as the Conditional Autoregressive (CAR) model of Besag [1974] to non-parametric models that cluster regions based on proximity and similar behaviors such as the MDP/MRF model of Orbanz & Buhmann [2008]. We build a hybrid model that nests a parametric CAR approach within regions that are found by a locally-exchangeable non-parametric clustering model, while also building spatially-varying economic and demographic predictors into our estimation of changes in crime over time. This semi-parametric approach allows for spatial discontinuities between regions that are the result of natural or human barriers, while addressing spatial sharing between neighborhoods within regions. We also evaluate how heterogeneity in the degree of connectedness of neighborhoods affects the local shrinkage around those neighborhoods. We apply our developed methodology to estimate changes in crime throughout Philadelphia over the 2006-15 period.

Sayantan Banerjee, (Indian Institute of Management, Indore)

Bayesian Nonparametric Graph Clustering with applications to Proteomic data
We present clustering methods for multivariate data exploiting the underlying geometry of the graphical structure between variables. As opposed to standard approaches that assume known graph structures, we first estimate the edge structure of the unknown graph using Bayesian neighborhood selection approaches, wherein we account for the uncertainty of graphical structure learning through model-averaged estimates of the suitable parameters. Subsequently, we develop a nonparametric graph clustering model on the lower dimensional projections of the graph based on Laplacian embeddings using Dirichlet process mixture models. In contrast to standard algorithmic pproaches, this fully

probabilistic approach allows incorporation of uncertainty in estimation and inference for both graph structure learning and clustering. More importantly, we formalize the arguments for Laplacian embeddings as suitable projections for graph clustering by providing theoretical support for the consistency of the eigenspace of the estimated graph Laplacians. We develop fast computational algorithms that allow our methods to scale to large number of nodes. Through extensive simulations we compare our clustering performance with standard clustering methods. Our methods are motivated by and applied to a novel pan-cancer proteomic data set, where we evaluate protein networks and clusters across multiple cancer types, to get biological insights about proteomic activities by developing an accurate understanding of the compositional and topological structures of different signaling networks.

Andrés Barrientos, (Duke University)

Bayesian Bootstraps for Massive Data

The Bayesian bootstrap (BB) was proposed in Rubin (1981) as a Bayesian analogue to the bootstrap (Efron, 1979). BB can be implemented through a direct sampling scheme; however, even simple, closed-form updates can be challenging when the size of the datasets is massive. Recently, two scalable adaptations of the bootstrap have been proposed: the bag of little bootstraps (Kleiner et al., 2014) and the subsampled double bootstrap (Sengupta et al., 2016). In this paper, we show that these methods can be modified for the Bayesian bootstrap while retaining similar theoretical and computational properties. In fact, for certain classes of functionals, lossless inference for the BB can be performed at no additional computational cost. Moreover, we propose extensions of the algorithms that approximate the posterior distribution of functionals of the Dirichlet Process. Finally, we illustrate the performance of our methods in simulated and real datasets.

Marco Battiston, (University of Oxford)

A characterization of product-form exchangeable feature probability functions

We characterize the class of exchangeable feature allocations assigning probability $V_{n,k}\prod_{l=1}^{k} W_{m_l}U_{n-m_l}$ to a feature allocation of n individuals, displaying k features with counts $(m_1,\ldots,m_k)$ for these features. Each element of this class is parametrized by a countable matrix V and two sequences U and W of non-negative weights. Moreover, a consistency condition is imposed to guarantee that the distribution for feature allocations of $n-1$ individuals is recovered from that of n individuals, when the last individual is integrated out. In Theorem **??**, we prove that the only members of this class satisfying the consistency condition are mixtures of the Indian Buffet Process over its mass parameter γ and mixtures of the Beta–Bernoulli model over its dimensionality parameter N. Hence, we provide a characterization of these two models as the only, up to randomization of the parameters, consistent exchangeable feature allocations having the required product form.

Brenda Betancourt, (Duke University)

Record linkage via Dirichlet Process Mixtures

The task of merging databases in the absence of a unique identifier can be a challenging problem. We introduce a flexible approach for merging two deduplicated databases by using a Dirichlet Process prior on the linkage structure, as opposed to a uniform prior as presented in Steorts, Hall and Fienberg (2015). By assuming a DP prior over the partitions, we address the main weakness of the original model, namely having to choose the latent population size in advance. In contrast, the DP mixture automatically determines the number of latent entities (or clusters) dependent on the concentration parameter α. We exploit the bipartite graph representation of the links between records across databases to sample α and therefore provide a fully automatic approach that learns the linkage structure from the data. We explore the performance of the model using simulations and real data from the Italian Survey on Household and Wealth.

Anirban Bhattacharya, (Texas A&M University)

Bayesian fractional posteriors

We consider the fractional posterior distribution that is obtained by updating a prior distribution via Bayes theorem with a fractional likelihood function, a usual likelihood function raised to a fractional power. First, we analyze the contraction property of the fractional posterior in a general misspecified framework. Our contraction results only require a prior mass condition on certain Kullback-Leibler (KL) neighborhood of the true parameter (or the KL divergence minimizer in the misspecified case), and obviate constructions of test functions and sieves commonly used in the literature for analyzing the contraction property of a regular posterior. We show through a counterexample that some condition controlling the complexity of the parameter space is necessary for the regular posterior to contract, rendering additional flexibility on the choice of the prior for the fractional posterior. Second, we derive a novel Bayesian oracle inequality based on a PAC-Bayes inequality in misspecified models. Our derivation reveals several advantages of averaging based Bayesian procedures over optimization based frequentist procedures. As an application of the Bayesian oracle inequality, we derive a sharp oracle inequality in the convex regression problem under an arbitrary dimension. We also illustrate the theory in Gaussian process regression and density estimation problems.

Apurva Bhingare, (Florida State University)

Multivariate Skewed Responses: New Semiparametric Regression Model and a Bayesian Recourse

For many real-life studies with skewed multivariate responses, the size of the covariate effects as well as the degrees of skewness are important to evaluate the covariate effects on the response and the predictive distributions. We propose a new class of semiparametric multivariate skewed response models with associated Bayesian method implementable via convenient computational tool. Compared to existing models, our proposed model enjoys several desirable properties, including a flexible multivariate association, meaningful physical interpretations of marginal covariate effects and skewness, and assurance of consistent Bayesian estimates of parameters and nonparametric error density under a set of practical assumptions on the priors. We examine the finite sample performance and robustness of our proposed method through simulation studies. We also illustrate the practical advantages of our methods over existing parametric methods by analyzing a periodontal disease study.

Saranjeet Kaur Bhogal, (University of Pune)

Bayesian Semi-Parametric Modelling of Durations

In this paper, we propose a two-component semi-parametric model for durations in high-frequency data. The first component is the intra-day seasonality factor and the second component corresponds to the intra-day dynamics explained by Autoregressive Conditional Duration (ACD) type of model with functional coefficients. Bayesian P-Splines are used for the estimation of all the unknown parameters of the model. In order to ensure the theoretical properties of the model, there are some constraints on the parameters which further decrease the already low acceptance rate in the Metropolis-Hastings (MH) algorithm. To solve this problem we use an extension of Hamiltonian Monte Carlo (HMC) called the No-U-Turn Sampler (NUTS) which, along with computational efficiency, also helps in the better exploration of the parameter space. Unlike MH and HMC algorithms, the NUTS needs no hand-tuning of user specified parameters. The model is tested under a simulation study and is illustrated with the analysis of the high-frequency data for the year 2015 from National Stock Exchange (NSE), India. A comparison of NUTS with the standard MH based MCMC is also made.

Ilaria Bianchini, (Politecnico di Milano)

Constructing stationary time series of completely random measures via Bayesian conjugacy

One flexible approach to building stationary time-dependent processes exploits the mathematical concept of conjugacy in a Bayesian framework. Under this approach, the transition law $\mathcal{L}\left(X_t|X_{t-1}\right)$ of the process $\{X_t\}$ is defined as the predictive distribution of an underlying Bayesian model (see e.g. Pitt and Walker, JASA, 2005). Then, if the model is conjugate, the transition kernel can be analytically derived, making the approach particularly appealing. We aim at achieving such a convenient mathematical tractability in the context of completely random measures (CRMs), i.e. when the variables exhibiting a time dependence are CRMs. In order to take advantage of the conjugacy, here we consider the large class of exponential family of completely random measures (see Broderick et al., Bernoulli, 2017). This leads to a simple description of the process which has an AR(1)-type structure and offers a framework for generalizations to more complicated forms of time-dependence. The proposed process can be straightforwardly employed to extend CRM-based Bayesian nonparametric models such as feature allocation models to time-dependent data. These processes can be applied to problems from modern real life applications in very different fields, from computer science to biology. In particular, we develop a dependent latent feature model for the identification of features in images and a dynamic Poisson factor analysis for topic modelling, which are fitted to synthetic and real data.

Enrico Bibbona, (Politecnico Di Torino)

A new way to inducing dependence in iid Dirichlet processes

In the late years, a lot of works proposed ways to induce dependence among iid replications of Dirichlet processes. What we are going to present here are some theoretical results, based on the copulae approach, that show how to induce dependence letting invariant the marginal distributions of the processes. We also propose a MCMC algorithm that can be used to obtain posterior samples in a very general setting. Our approach can be generalized and used with iid replication of a Hierarchical Dirichlet process. The results are shown on simulated and a real data example. In the real data we use our approach to define a mixture model where mixture probabilities of the same class are temporal dependent.

Olivier Binette, (Université du Québec à Montréal)

Constrained semiparametric modelling for directional statistics

We investigate a trigonometric polynomial analogue of the Bernstein polynomial densities and its use for circular density modeling. In particular, we obtain sufficient conditions on the parameter space to enforce shape constraints such as total variation bounds and periodic unimodality. Our models, as is common, arise as the image of a linear approximation operator having shape preserving properties. We exploit this relationship to propose a general framework of prior elicitation on shape constrained density spaces, ensuring strong posterior consistency at all bounded density. The result partly relies on well-known properties of mixtures of priors, while the operator correspondence allows us to obtain Kullback-Leibler support on the set of all bounded densities of the constrained space. This extends known results by removing an usual continuity hypothesis on the densities at which consistency is achieved. We apply this framework to the specification of density priors on the circle and the sphere, combining our trigonometric model to the Bernstein polynomial densities. Strong posterior consistency is then immediately obtained from approximation-theoric properties of the considered operators. Applications in protein bioinformatics, for the analysis of phi-psi angles distributions, are discussed.

Pier Giovanni Bissiri Bissiri, (University of Milano Bicocca)

Bayesian analysis of the Gini-Simpson index
A very important topic in ecology is measuring biological diversity, that is the variety and abundance of species in a population. The Gini-Simpson index is the probability to sample without replacement from the population two units belonging to two different species or categories. Such index is one of the most important and well known diversity indexes. If the number of species K is unknown and a Bayesian model is to be implemented, then a prior on a potentially infinite sequence of frequencies needs to be assessed. Three different models are been compared: two of them exhibit a finite (random) K with conditional distribution of the frequencies given K symmetric Dirichlet; in the third model K is infinite and the frequencies are given a prior that makes them distributed as weights of a Poisson Dirichlet Process. An extensive analysis is presented of the prior first two moments together with exact formulas for the posterior estimates of the Gini-Simpson index implied by the three models. This enables to provide a clear interpretation of model parameters, thereby helping their elicitation. It also allows a in-depth comparison of the three models, highlitening advantages and limitations. Furthermore, for each model, the posterior distribution of the index is shown to be asymptotically normal with convergence rate $1/n^{1/2}$ in case of non-equiprobability of the species and asymptotically chi square with convergence rate $1/n$ in case of equiprobability of the species.

Benjamin Bloem-Reddy, (Columbia University (as of May, I will be at the University of Oxford))

Nested urn models of random partitions and graphs
We propose a class of multi-color urn models and the resulting random partition processes, *nested Pólya partitions*, that condition on the arrival times of new colors. Such models are neutral-to-the-left, exhibit a constrained form of exchangeability that gives rise to a nested paint-box representation, and their asymptotic properties are governed by the asymptotic behavior of the increments. By randomizing the arrival times, a number of well-known models are obtained as special cases, the most notable being Exchangeable Gibbs Partitions and the (non-exchangeable) Yule-Simon process. Both models give rise to constructive representations of the (conditional) scaled limiting block sizes in terms of Mittag-Leffler and Beta random variables. We make connections to preferential attachment-type graphs, yielding insight to the connection between sparsity and power law degree distributions.

Sudip Bose, (Department of Statistics, The George Washington University)

Bayesian robustness in Bayesian nonparametrics
We consider Bayesian robustness in the context of Bayesian Nonparametrics, and specifically analyses with the Dirichlet Process prior. To specify the prior distribution of even a single real parameter, one has to specify exactly, the prior probability of (countably) infinitely many sets. It seems unrealistic to assume that one can do this precisely in finite time. The robust Bayesian viewpoint replaces a single prior by a class of priors that represents one's prior uncertainty. Berger (JSPI, 1990) points out the central role of the range of posterior expectations (as the prior varies over a neighborhood class) in assessing sensitivity to the prior. For example, the posterior expectation of a parameter is its Bayes estimator under squared error loss. Posterior probabilities play a central role in hypothesis testing and credible set estimation. If the range of posterior expectations is small, one has robustness. We consider the density ratio class and the epsilon-contamination class, two widely used classes and show how to determine the range of posterior expectations and posterior probabilities. One may also be interested in finding robust procedures. If one uses the wrong prior, one can measure the opportunity cost of not using the true prior by the posterior regret. A minimax posterior regret procedure thus provides, in a sense, the most protection against prior misspecification. We show how to find the C-minimax posterior regret (CMPR) for classes of priors C, and provide illustrative examples.

Diana Cai, (University of Chicago)

Edge-exchangeable graphs, sparsity, and power laws

Many popular network models rely on the assumption of (vertex) exchangeability, in which the distribution of the graph is invariant to relabelings of the vertices. However, the Aldous-Hoover theorem guarantees that these graphs are dense or empty with probability one, whereas many real-world graphs are sparse. We present an alternative notion of exchangeability for random graphs, which we call edge exchangeability, in which the distribution of a graph sequence is invariant to the order of the edges. We demonstrate that edge-exchangeable models, unlike models that are traditionally vertex exchangeable, can exhibit sparsity. To do so, we outline a general framework for graph generative models; by contrast to the pioneering work of Caron and Fox (2014), models within our framework are stationary across steps of the graph sequence. In particular, our model grows the graph by instantiating more latent atoms of a single random measure as the dataset size increases, rather than adding new atoms to the measure.

Federico Camerlenghi, (University of Bologna, BIDSA and Collegio Carlo Alberto)

Dependent nonparametric priors for partially exchangeable data

In Bayesian nonparametrics a large amount of literature has been developed for exchangeable data. However in a large variety of applications exchangeability is a quite restrictive assumption. A typical example is when data are generated by different experiments: even if these experiments may be related, they have some specific features that induce heterogeneity across observations. Hence, one needs to resort to more general dependence structures. In such a situation data are usually divided into different groups, and a more appropriate assumption is partial exchangeability. In particular the construction of dependent random probability measures to deal with partially exchangeable observations has recently attracted great attention in Bayesian nonparametric literature. Here we define and investigate general classes of nonparametric priors based on transformations of completely random measures, that have a nested and a hierarchical structure. We discuss some of their distributional properties which are relevant to carry out posterior inference. These theoretical results are employed for devising MCMC algorithms to face prediction in species sampling problems and survival analysis.

Lorenzo Cappello, (Università Bocconi)

Recursive non-parametric predictive for a discrete regression model

We propose a recursive algorithm to estimate a finite set of conditional distributions. The procedure is fully nonparametric and has a Bayesian interpretation. Indeed the recursive updates follow a Bayesian update to a certain extent and exploit a recent characterisation of Bayesian predictive models through a sequence of bivariate copulas. We prove weak convergence of the distribution estimates, and demonstrate numerical accuracy via simulations. The estimate is very fast and requires limited computing power; being also parallelizable. We show that it is competitive with both frequentist and Bayesian non-parametric models.

Annalisa Cerquetti, (None)

Bayesian nonparametric testing of Hubbell's neutral hypothesis: expected heterozygosity decomposition under mainland-island community structure.

The neutral hypothesis originates in mathematical population genetics (Kimura, 1968) as opposite to selective Darwinian theory. The theory asserts that a large fraction of observed genetic variation is nonselective but occurs purely by chance. In a community ecology transposition abundances within a neutral population fluctate and diversity arises as a balance between the immigration of new species and local extinction. In 2001 Hubbell extends this theory to multiple sites populations using a mainland-island community structure: local communities governed by neutral dynamics are connected through migration to a neutral metacommunity where diversity is generated through speciation. Interest in testing Hubbell's neutral hypothesis is huge, being suspected to play a role in a wide set of interacting living systems, spanning from tropical rain forests to human gut microbiome. The current approach is by Montecarlo significance test based on Etienne's (2007) sampling formula. Just recently it has been shown (Harris et al. 2015) that a large class of neutral models with mainland-island structure converges, in the large population limit, to the Hierarchical Dirichlet process (Teh et al. 2006). Moving from these findings, in a Bayesian nonparametric perspective, we investigate an alternative approach to neutrality testing based on Jost's (2007) diversity differentiation. Relying on properties of structural distributions of infinite random discrete distributions, we focus on prior and posterior expectations of alpha and beta heterozygosities under Hierarchical Dirichlet priors and study their possible role both in an hybrid Bayesian-frequentist approach and in the construction of a proper Bayesian solution.

Minwoo Chae, (Department of Mathematics, The University of Texas at Austin)

A novel approach to Bayesian consistency
It is well-known that the Kullback-Leibler (KL) positivity condition implies the posterior consistency in the weak topology, but not sufficient for consistency in total variation. Barron et al. [*Ann. Statist.* (1999)] provided a counterexample showing that the KL positivity is insufficient for the strong consistency. On the other hand, based on a prior satisfying the KL positivity condition, Walker and Hjort [*J. R. Statist. Soc. B* (2001)] constructed a sequence of data dependent priors yielding strongly consistent posteriors. We construct a sequence of priors with the same property for the posterior, but depending only on the sample size. Given a prior with the KL positivity property, new priors are constructed by penalizing the smoothness of densities. It depends on the sample size through the convergence rate in Levy-Prokhorov metric. Based on this result, we provide improved conditions for the posterior consistency in examples.

Antik Chakraborty, (Texas A&M University)

Bayesian sparse multiple regression for simultaneous rank reduction and variable selection
We develop a Bayesian methodology aimed at estimating low rank and row sparse matrices in a high dimensional multivariate response linear regression model. Starting with a full rank matrix and thus avoiding any prior specification on the rank, we let our estimate shrink towards the space of low rank matrices using continuous shrinkage priors. For selecting rows we propose a one step post processing scheme derived from putting group lasso penalties on the rows of the coefficient matrix with default choice of tuning parameters. We then provide an adaptive posterior estimate of the rank using a novel optimization function achieving dimension reduction in the covariate space. We exhibit the performance of the proposed methodology in an extensive simulation study and a real data example.

Noirrit Kiran Chandra, (Indian Statistical Institute, Kolkata)

Detecting miRNA expression level differences for Gingivo buccal squamous cell carcinoma patients
Gingivo buccal squamous cell carcinoma (GBSCC) is one of the most prevalent cancers in oral cavity, especially among the tobacco users in India. MicroRNAs (miRNAs) are small (typically 22 nucleotides) non-coding, endogenous, single-stranded RNAs. The advantages and utility of miRNAs as diagnostic and prognostic cancer markers is at

the vanguard in recent years. Several gene studies revealed that miRNAs are also involved in human tumorigenesis and could act either as tumorigenic/oncogenic or anti-tumorigenic molecules. Expressions of 522 miRNAs among 18 unrelated patients were checked for both case and control cells. The paired observations are statistically tested to detect significant difference in their expression levels. We proposed to consider a spatial Gaussian process prior over the miRNAs in this case. In this data, sample size is much smaller compared to its dimension and shrinkage estimation assuming sparsity is quite commonplace. In our model such assumptions are not needed. For simultaneous testing of paired expression values multiple testing methods are generally used. In human genome it is expected that miRNA expression values to have some kind of dependence among themselves. However, in most of the methods the decision rules mainly focus on the validity of the test procedure rather than actually utilizing the dependency to increase efficiency. The decisions rules of different hypotheses are not directly related and depends only on marginal distribution of corresponding test statistics. We have proposed a novel Bayesian multiple testing method which takes into account the joint behavior of different hypotheses.

Sheng-Kai Chang, (National Taiwan University)

A Bayesian estimator for stochastic frontier models with errors in variables

A Bayesian estimator is proposed for a stochastic frontier model with errors in variables. The model assumes a truncated-normal distribution for the inefficiency and accommodates exogenous determinants of inefficiency. An empirical example of Tobin's Q investment model is provided, in which the Q variable is known to suffer from measurement error. Results show that correcting for measurement error in the Q variable has an important effect on the estimation results.

Ya-Ting Chang, (University of Edinburgh)

Evidence evaluation and functional data analysis

A mathematical formulation is developed for a likelihood ratio as a measure of support for one proposition H_p over another H_d where the data are functions. Control data $\mathbf{Y_c}$, whose source is known, and recovered data $\mathbf{Y_r}$, whose source is not known, are to be compared. In the context of forensic science, H_p represents the proposition that the recovered and control data have the same source and H_d represents the proposition that the recovered data have a different source from, and independent of, the control data. A nonparametric regression model $Y_j = f(j/m) + r_j,\ j = 1, \ldots, m$ is assumed where the regression curve f is modelled by a B-spline: $f(x) = \sum_{b=1}^{B} \theta_b \phi_b(x)$ and various models for the error term $\mathbf{r}$ including constant variance, autocorrelation between neighbouring variances, functional variance and functional principal component scores are considered. The control data and recovered data are hierarchical. They are parameterised with expectation Θ and covariance Σ that have hyperparameters (η, T) and (U, ℓ), respectively. The likelihood ratio is given by

$$\frac{\int_\Sigma \int_\Theta f(\mathbf{Y}_c \mid \Theta, \Sigma) f(\mathbf{Y}_r \mid \Theta, \Sigma) f(\Theta \mid \eta, \Sigma, T) d\Theta\ f(\Sigma \mid U, \ell) d\Sigma}{\prod_{s=c,r} \int_\Sigma \int_\Theta f(\mathbf{Y}_s \mid \Theta, \Sigma) f(\Theta \mid \eta, \Sigma, T) d\Theta\ f(\Sigma \mid U, \ell) d\Sigma}$$

where the hyperparameters are to be estimated from training data. Functional data are available on microspectrophotometry measurements in the visual (VIS) range made for pen inks and for which the methods are tested for correct assignations of common or different sources using cross-validation of the training set.

Zhen Chen, (National Institutes of Health)

Estimating constrained and correlated ROC surfaces: A Bayesian semiparametric approach

In application of diagnostic accuracy, it is possible that a priori information may exist regarding the test score distributions, either between different disease populations for a single test or between multiple correlated tests. Motivated by a study on diagnosing endometriosis, we propose an approach to estimating diagnostic accuracy measures that can incorporate different stochastic order constraints on the test scores when an ordinal true disease status is in consideration. We show that the Dirichlet process mixture provides a convenient framework to both flexibly model the test score distributions and embed the a priori ordering constraints. We also utilize the Dirichlet process mixture to model the correlation between multiple tests. In taking a Bayesian perspective to inference, we develop an efficient Markov chain Monte Carlo algorithm to sample from the posterior distribution and provide posterior estimates of the receiver operating characteristic surfaces and associated summary measures. The proposed approach is evaluated with extensive simulation studies, and is demonstrated with an application to the endometriosis study.

Will Cipolli, (Colgate University)

Supervised Learning via Polya Trees
The goal in any classification scheme is to design a system that classifies new observations into their true class as often as possible. The nonparametric approach of the multivariate Polya tree proposed realizes impressive results in simulations and real data analyses, performing similarly to, or better than, current approaches in many cases. The flexibility gained from eliminating certain distributional assumptions from the model can greatly improve the ability to correctly classify new observations; even minor deviations from parametric distributional assumptions could lead to missing an important feature in any one class's density. Data analyses and simulations cover classification in various settings covering a variety of dimensions, sample sizes and number of classes were successful. The performance of classification was evaluated using ten-fold cross validation which was kept low during simulation and data analysis. A nice aspect of the methodology is that it is computationally efficient and can handle a large number of classes. It has proven quite successful in correctly classifying observations in a variety of scenarios without any user-chosen specification making it very simple to implement as there are no kernel tricks or initialization steps for the user.

Andrea Cremaschi, (Universitetet i Oslo / NCMM)

Adaptive truncation of a Bayesian nonparametric multivariate regression model for a study of fertility and partnership patterns of Colombian women
We propose a flexible Bayesian nonparametric multivariate regression model, which can capture nonlinear regression functions and the presence of non-normal errors, such as heavy tails or multi-modality. The infinite mixture model has interpretable covariate-dependent weights constructed through normalization, allowing for combinations of both discrete and continuous covariates, and extends the model developed in (Antoniano - Villalobos et al., 2014) for a multivariate response. The infinite number of components and intractable normalizing constant pose computational difficulties, which are overcome through an adaptive truncation algorithm (Griffin, 2014). The algorithm combines adaptive Metropolis-Hastings with sequential Monte Carlo to create a sequence of truncated posteriors and automatically determines the level of truncation. The model and algorithm are applied to a lifestyle study on Colombian women, which aims to understand the relationship between some focal life events (e.g. age at first sexual intercourse, relationship, child, exit from education, presence in the labour market) and various baseline factors, such as year of birth, region of birth, and ethnicity. Regression function and conditional density estimates are presented, along with an analysis of the implied covariate-dependent clustering.

Miguel de Carvalho, (University of Edinburgh)

Affinity-Based Measures of Diagnostic Test Accuracy

We propose new summary measures of diagnostic test accuracy which can be used as companions, or possibly as alternatives, to existing diagnostic accuracy measures. Conceptually, our summary measures are tantamount to the so-called Hellinger affinity, and a covariate-adjusted version is developed to assess the discrimination performance of a medical diagnostic test, conditionally on the value of a predictor. Nonparametric Bayes estimators for the proposed indexes are devised, theoretical properties of the corresponding priors are derived, and the performance of our methods is assessed through a simulation study. Data from a prostate cancer diagnosis study are used to illustrate our methods.

Roberta De Vito, (Princeton University)

Bayesian Multi-study Factor Analysis in High-dimensional Biological Data

High-dimensional data are the norm in current statistical research of most scientific data. Gaining systematic knowledge from these data is a cumulative process that benefits substantially from the integration of multiple studies and data from different technologies, and that relies critically on methods of analysis. In this work, we introduce “Bayesian multi-study factor analysis” (BMFA), a generalized version of Bayesian factor analysis that jointly models data from multiple studies in a single analysis by using (1) factors that capture common information, shared across study data, and (2) study-specific factors. The fundamental challenge of this statistical framework is the estimation of common features shared among studies (Zhao et al., (2014), Garrett-Mayer et al., (2008)) and identifying the components of variation specific to each study (Irizarry et al., (2003)). We use sparse modeling of high-dimensional factor loadings matrices, both common and specific, using nonparametric shrinkage priors (Bhattacharya and Dunson, (2011)}. We describe a computationally efficient algorithm to estimate the parameters and to select the number of relevant common and study-specific factors. We assess the operating characteristics of our method with simulation studies, and we present an application in ovarian cancer with four gene expression studies.

Kumaresh Dhara, (Department of Statistics, Florida State University)

Bayesian Single Index Model with Covariates Missing at Random

Bayesian single index model is a highly promising dimension reduction tool for an interpretable modeling of the non linear relationship between the response and its predictors. However, existing Bayesian tools in this area suffer from slow mixing of the Markov Chain Monte Carlo (MCMC) computational tool and also lack the ability to deal with missing covariates. To circumvent these practical problems, we present a new Bayesian single index model with MCMC algorithm using a mode-alignment based proposal density for the index vector for an efficient Metropolis Hastings (MH) algorithm to sample from full conditional distribution. Our method leads to an interpretable model and inference, the efficient evaluation of the likelihood, fast convergence of the MCMC, and a first time extension of inference to missing at random covariates. In addition, we provide sufficient conditions for establishing optimal rate of posterior convergence for the overall regression function.

Giuseppe Di Benedetto, (University of Oxford)

Non-exchangeable random partition model for microclustering

The task of clustering data is one of the most popular applications of Bayesian Nonparametrics. This problem can be addressed by specifying a prior distribution on the partition of the data. The most commonly used priors, such as the Chinese Restaurant Process and its two-parameters generalization, rely on some exchangeability assumption; while this assumption may be reasonable for some applications, it has strong implications on the asymptotic properties of the cluster sizes. In fact, as proved in [Kingman - The Representation of Partition Structures - 1978] and stressed in [Miller et al. - Microclustering: When the Cluster Sizes Grow Sublinearly with the Size of the Data Set - 2015], exchangeable random partitions imply the linear growth of the cluster sizes, which is not suitable for several applications. We will present a flexible non-exchangeable random partition model which is able to generate partitions whose growth of the clusters sizes is almost surely sublinear. Along with this result, we provide the asymptotic behaviour of the number of clusters of a given size. Both these rates can be tuned through a parameter and the base measure of a homogeneous completely random measure and inference on the parameters is based on a Sequential Monte Carlo algorithm.

Stefanos Dimitrakopoulos, (Oxford Brookes University)

Semiparametric Bayesian inference for time-varying parameter regression models with stochastic volatility

We develop a Bayesian semiparametric method to estimate a time-varying parameter regression model with stochastic volatility, where both the error distributions of the observations and parameter-driven dynamics are unspecified. We illustrate our methodology with an application to inflation.

David Draper, (Department of Applied Mathematics and Statistics, University of California, Santa Cruz)

A Noninformative Prior on a Space of Functions

In a given problem, the Bayesian statistical paradigm requires the specification of a prior distribution that encodes relevant information about the problem external to the data. In cases where little such information is available, the problem under study may possess an invariance under a transformation group that encodes a lack of information, leading to a unique prior. Previous successful examples of this idea have included location-scale invariance under linear transformation, multiplicative invariance of the rate at which events in a counting process are observed, and the derivation of the Haldane prior for a Bernoulli parameter. In this paper we show that this method can be extended in two ways: (1) to yield families of approximately invariant priors, and (2) to the infinite-dimensional setting, yielding families of priors on spaces of distribution functions. Our results can be used to describe conditions under which a particular Dirichlet Process posterior arises from an optimal Bayesian analysis, in the sense that invariances in the prior and likelihood lead to one and only one posterior distribution.

David Draper, (University of California, Santa Cruz)

Optimal Bayesian Analysis of A/B Tests at Big-Data Scale: The Frequentist Bootstrap is Really Bayesian Nonparametrics in Disguise

By a theorem of de Finetti, when your uncertainty about an as-yet-unobserved vector $y = (y_1, \ldots, y_n)$ of real values is exchangeable, this is functionally equivalent to the Bayesian hierarchical model $F \sim p(F), (y_i|F)\ IIDF$, where $p(F)$ is a non-parametric prior on the space $\mathcal{F}$ of all CDFs on $\mathbb{R}$. Choosing $p(F)$ to be $DP(\alpha, F_0)$ without essential loss of generality, by the usual conjugate updating the posterior for F is then $DP\left(\alpha + n, \frac{\alpha\, F_0 + n\, \hat{F}_n}{\alpha + n}\right)$, in which $\hat{F}_n$ is the empirical CDF based on y. When context suggests that little is known about F external to the data, this implies that α should be close to 0, and when n is much larger than α (as in A/B testing at massive scale, in which this situation arises in each of the A and B groups), the $DP(n, \hat{F}_n)$ posterior emerges as the basis of an *optimal Bayesian analysis*, in the sense that problem context leads to an essentially unique posterior. Sampling from $DP(n, \hat{F}_n)$ at scale with stick-breaking is a bad idea (because stick-breaking is sequential); instead, by using the P'olya-Urn representation of the predictive distribution for y_{n+1}, it's straightforward to see that an embarrassingly parallel way to sample from $DP(n, \hat{F}_n)$ is equivalent to the frequentist bootstrap!

Haosui Duanmu, (University of Toronto)

On extended admissible procedures and their nonstandard Bayes risk

For finite parameter spaces under finite loss, every Bayes procedure derived from a prior with full support is admissible, and every admissible procedure is Bayes. This relationship already breaks down once we move to finite-dimensional Euclidean parameter spaces. Compactness and strong regularity conditions suffice to repair the relationship, but without these conditions, admissible procedures need not be Bayes. Under strong regularity conditions, admissible procedures can be shown to be the limits of Bayes procedures. Under even stricter conditions, they are generalized Bayes, i.e., they minimize the Bayes risk with respect to an improper prior. In both these cases, one must venture beyond the strict confines of Bayesian analysis. Using methods from mathematical logic and nonstandard analysis, we introduce the class of nonstandard Bayes decision procedures—namely, those whose Bayes risk with respect to some prior is within an infinitesimal of the optimal Bayes risk. Among procedures with finite risk functions, we show that a decision procedure is extended admissible if and only if its nonstandard extension is nonstandard Bayes. For problems with continuous risk functions defined on metric parameter spaces, we derive a nonstandard analogue of Blyth's method that can be used to establish the admissibility of a procedure. We also apply the nonstandard theory to derive a purely standard theorem: when risk functions are continuous on a compact Hausdorff parameter space, a procedure is extended admissible if and only if it is Bayes.

Bianca Dumitrascu, (Princeton University)

A Meta Tissue Non-parametric Factor Analysis Model For Gene Co-expression Under Structured And Unstructured Noise

Gaining knowledge from high-dimensional scientific domains is a cumulative process that requires integration of multiple, often diverse studies, and relies critically on the methods used to perform the data analysis. Most measurements from high throughput experiments show both artifactual and biological sources of variation; while some artifacts reappear across studies, common biological features are more likely to be shared among the studies. We address these challenges in the context of extracting clusters of co-expressed genes essential to the identification of patterns across tissues coming from correlated brain regions of overlapping sets of individuals. Our method is a meta-analysis Bayesian nonparametric generalized latent factor model with three important features: a) an interpretable Indian Buffet Process (IBP) prior over a common loading matrix modeling the underlying gene correlation structure, b) a two component spike and slab prior over a study specific loading matrix to account for the technical structured noise, and c) an error term accounting for overlapping subjects across studies. Our computationally efficient algorithm brings together under a unified framework a traditional Gibbs sampler and a fast dynamic EM algorithm (Rockova et al, 2015). We illustrate the advantages of the proposed method through simulations and an application to a high dimensional biological data set. In particular, we identify brain specific co-regulated gene clusters within the Genotype-Tissue Expression (GTEx) project data using RNA-sequencing of samples from a total of 449 individuals across 13 different, but highly correlated, brain tissue subtypes.

Ehtsham Elahi, (Netflix)

A Nonparametric Latent Factor Model For Location-Aware Video Recommendations

In a web application we are provided with a rich view of each user. For example in a video streaming application, like Netflix, we can observe not only their preference for different types of content but also how those preferences change with respect context, such as time of day, day of week, device, and so on. An important contextual variable that influences a customer's preferences is their geographical location. It is reasonable to assume that customers who live in close proximity may have similar viewing preferences. Hence, a model is required that can capture not only a customer's latent viewing preferences, but also the relationship between those and their location. To capture both these aspects, we seek to model them in a unified model so that both location and viewing behavior can take advantage of information in each modality. For this task, we employ a bayesian latent factors model to jointly model a customer's viewing history and their geographical location. In order to tune the complexity of the model to best represent the data, we make use of bayesian nonparametric techniques. We describe a sampling based inference technique that can scale to large real-world data sets. Finally we show results obtained by applying the model to a large internal Netflix data set, that illustrates that the model was able to capture interesting relationships between viewing patterns and geographical location.

Clément Elvira, (Cristal - Ec Lille)

A Bayesian nonparametric modeling for subspace estimation
We present a Bayesian nonparametric (BNP) model to infer the subspace of a set of observations. More precisely, the observations are assumed to live in a lower-dimensional subspace, spanned by a random orthogonal basis which is assigned a prior distribution defined on the Stiefel manifold. The factor scores take benefit from an Indian buffet process prior to model the uncertainty related to the number of relevant components. This model leads to a joint posterior distribution of the subspace orthonormal basis and model parameters. Compared to existing algorithms, this method offers a more convenient implementation of a Markov Chain Monte Carlo algorithm. Besides, we obtain several theoretical results. It is shown that the natural subspace dimension estimator, namely the marginal maximum a posteriori estimator of active components, is inconsistent. As an alternative, we analytically derive the theoretical distributions of the components that are not significant. We compare these distributions with the empirical ones through a goodness-of-fit testing strategy in the spirit of model selection. As a result, we identify the low-dimensional subspace that is the most relevant to describe the signal. We complete our study with an experimental analysis of the behavior of the proposed method, and illustrate its relevance and accuracy on signal processing tasks. It appears that the proposed BNP dimension reduction approach can be efficiently coupled with a clustering problem or to a linear unmixing problem within a unique framework.

Melanie F. Pradier, (University Carlos III in Madrid)

Dynamic Poisson Factorization Model for the Productive Structure of Economies
This work aims at analyzing the productive structure and competitiveness of world economies over time, by finding an interpretable representation for international trade. Our objective is to understand economic growth of countries, i.e., which factors make countries wealthier, and how these countries acquire such factors over time. Our study is based on international trade data, with the ultimate goal of issuing economic policy recommendations. We first propose an infinite Poisson factorization model for high dimensional count data that performs non-negative doubly sparse matrix factorization in a static scenario. Our model is able to find solutions with structured sparsity in the latent space, and capture fundamental elements underlying data. We then present a time-dependent extension to analyze temporal dynamics of high-dimensional count data. We allow both the per-country activation matrix and the latent feature matrix to evolve over time according to a Markovian structure. We combine a Markov Indian Buffet process with recent advances on Gamma-Poisson autoregressive chains: the former reflects how countries acquire or lose trade capabilities, while the second captures geo-political and technological variations over time. We demonstrate the usefulness of our approach in the analysis of export portfolios over more than 30 years in three different databases. Our model is compared against traditional approaches and recent works including the Poisson-Gamma Dynamical System and Thinned Gamma Process Poisson Factor Model. Empirical results show our model's ability to give easy-to-interpret information and valuable insights on countries' economical growth.

Stefano Favaro, (University of Torino and Collegio Carlo Alberto)

Bayesian nonparametric analysis of Kingman's coalescent

Kingman's coalescent is one of the most popular model in population genetics. It describes the genealogy of a population whose genetic composition evolves in time as the Wright-Fisher model, or suitable approximations of it belonging to the class of Fleming-Viot processes. Ancestral inference under Kingman's coalescent has had much attention in the literature, both in data analysis, and from a theoretical and methodological point of view. Given a sample of individuals from the population at stationarity, most contributions have aimed at making parametric inference on quantities related to the genealogy of the observed sample. In this talk we propose a Bayesian nonparametric predictive approach to ancestral inference. That is, under the prior assumption that the composition of the population evolves in time according to a neutral Fleming-Viot process, and given an initial sample of m individuals from the population at stationarity, we estimate quantities related to the genealogy of an additional unobservable sample of size $m' \geq 1$. For instance, how many non-mutant lineages would I expect a time t ago if I enlarged my initial sample by $m' \geq 1$? How many of these non-mutant lineages are associated to the rare genetic types? Within the ancestral inference, these questions are of great interest because they relate to the speed of evolution via the rate of turnover of alleles. As a by-product of our analysis we introduce a class of estimators which can be thought of as Good-Turing type estimators for ancestral inference. The proposed approach is illustrated through an application to genetic data.

Tamara Fernandez, (University of Oxford)

Posterior Consistency for a Non-parametric Survival Model under a Gaussian Process Prior

We prove almost surely consistency of a Survival Analysis model, which puts a Gaussian process, mapped to the unit interval, as a prior on the so-called hazard function. We assume our data is given by survival lifetimes T belonging to $\mathbb{R}^+$, and covariates on $[0,1]^d$, where d is an arbitrary dimension. We define an appropriate metric for survival functions and prove posterior consistency with respect to this metric. Our proof is based on an extension of the theorem of Schwartz (1965), which gives general conditions for proving almost surely consistency in the setting of non i.i.d random variables. Due to the nature of our data, several results for Gaussian processes on $\mathbb{R}^+$ are proved which may be of independent interest.

Sarah Filippi, (Imperial College London)

Bayesian Learning of Kernel Embeddings

Representations of probability measures in reproducing kernel Hilbert spaces via the so-called kernel mean embeddings provide a flexible framework for fully nonparametric hypothesis testing, which includes nonparametric two-sample and independence tests. In practice, however, the performance of these tests can be very sensitive to the choice of kernel and its lengthscale parameters. To address this central issue, we propose a new probabilistic model for kernel mean embeddings, the Bayesian Kernel Embedding model, combining a Gaussian process prior over the Reproducing Kernel Hilbert Space containing the mean embedding with a conjugate likelihood function, thus yielding a closed form posterior over the mean embedding. The posterior mean of our model is closely related to recently proposed shrinkage estimators for kernel mean embeddings, while the posterior uncertainty is a new, interesting feature with various possible applications. Critically for the purposes of kernel learning, our model gives a simple, closed form marginal pseudolikelihood of the observed data given the kernel hyperparameters. This marginal pseudolikelihood can either be optimized to inform the hyperparameter choice or fully Bayesian inference can be used.

Jairo Alberto Fuquene Patiño, (Department of Statistics, University of Warwick)

Choosing the number of components in mixture models with skewness and heavy tails via Non-local priors

In this work we extend the framework for choosing the number of components in mixture models using Non-local priors proposed in Fuquene, Rossell and Steel (2017) to non-normal likelihoods with the dual aim of providing a more flexible framework and exploring the robustness properties induced by different choices of likelihood and prior. We will use continuous mixtures with skewness and heavy tails and the general NLPs class of priors.

Bastian Galasso, (Pontifical Catholic University)

Bayesian Flexible Modeling for Families of Random Densities

This work discusses inference for families of random densities by combining concepts and methods from Bayesian nonparametrics and functional data analysis. Specifically, we construct a Bayesian nonparametric version of a model by Kneip and Utikal (2001), where curves of interest are themselves densities, and using the Karhunen—Loeve decomposition we link all members of a family of densities $\{f_k\}_{k=1}^{K}$. The main motivation for using Bayesian nonparametric inference here goes beyond the standard argument of modeling flexibility and robustness against misspecification. Indeed, as posed by one of the discussants: “The covariance method for estimating the spectrum of the covariance operator is naturally motivated when the observed densities are considered to be a sample of random functions. This motivation is lost in the nonrandom situation considered by Kneip and Utikal” (Chiuo and Muller). However, since from a Bayesian nonparametric perspective densities are themselves random, such conceptual concern disappears. The performance of our methods is evaluated through a Monte Carlo study and an illustration is given using Galton’s Height Data. References [1] A. Kneip and K. Utikal, “Inference for densities families using functional principal component analysis (with discussion)”. Journal of the American Statistical Association, 96: 519—542 , 2001.

Ed George, (Wharton, University of Pennsylvania)

Heteroskedastic BART with Multiplicative Regression Trees

BART (Bayesian Additive Regression Trees) is a scalable regression approach which deploys a sum of regression trees to non-parametrically model the conditional mean relationship between Y and X1,. . . ,Xp. However, in spite of its remarkable flexibility for fitting the mean, the basic BART formulation assumes a constant error variance over the entire predictor space, an assumption which is often unrealistic. Heteroscedastic BART is an elaboration of BART which adds a product of regression trees, each of which determines a contribution to the overall variance, to flexibly model the conditional variance relationship between Y and X1,. . . ,Xp. Thus, Heteroscedasctic BART simultaneously models the conditional mean and conditional variance relationships between Y and X1,. . . ,Xp. The potential of Heteroscedasctic BART is demonstrated on a simple low-dimensional simulated dataset, a higher-dimensional dataset of used car prices, a fisheries dataset and data from an alcohol consumption study.

Clara Grazian, (Nuffield Department of Medicine, University of Oxford)

Approximate Bayesian inference in semiparametric copula models

A copula is a multivariate probability distribution for which the marginal probability distributions are uniform. Copulas are used to flexibly describe the dependence between random variables because it has been shown (Sklar’s theorem) that any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula which absorbs all the dependence structure between the variables. We describe a simple method for making inference on a functional of the copula. The method is particularly useful when the model is only partially specified: the interest of the analysis is, in fact, on a particular function of the copula and it is possible to avoid the choice of the complete copula structure, in order to prevent estimation errors due to model miss-specification. In this sense, the complete dependence structure is considered as a nuisance parameter which belongs to an infinite dimensional metric space and we derive an approximate integrated likelihood for the parameter of interest. The method implicitly assumes for the nuisance parameters a prior process which favours distributions with a high level of entropy. Moreover, we show that nonparametric estimates of the marginal distributions lead, in general, to better asymptotic properties. We apply the method to several dependence measures, as the Spearman's ρ and the lower and upper tail dependence indices λ_L and λ_L. A great improvement in the literature on copula models is the fact that, in a general multivariate setting of dimension k, the proposed approach produces a quantification of the uncertainty of the estimates, while there exist no satisfactory method to do this in a frequentist context.

Rajarshi Guhaniyogi, (Department of Applied Math and Stat, University of California Santa Cruz)

Multiscale Spatial Kriging with Tree Shrinkage Priors for Massive Geostatistical Data

With the recent advancements in GIS software and related tools, spatial modelers are increasingly encountering massive spatial datasets which entail computations that become prohibitive. There is a burgeoning literature on approaches for analyzing large spatial datasets. In this article, we propose a mutiscale spatial regression model that employs a divide-and-conquer strategy within the Bayesian paradigm. We develop a multiscale spatial kernel convolution technique with higher order functions to capture fine scale local features and lower order terms to capture large scale features. To achieve parsimony, the coefficients in the multiscale kernel convolution model is assigned a new class of “Tree shrinkage prior" distributions. Tree shrinkage priors have an unique feature exerting increasing shrinkage on the coefficients as resolution grows so as to adapt to the necessary degree of resolution at any sub-domain. Our proposed model has a number of significant features over the existing multi-scale spatial models for big data. In contrast to the existing multiresolution spatial models, tree shrinkage prior auto-tunes the degree of resolution necessary to model a subregion in the domain. Secondly, compactly supported kernel functions allow local updating of the model parameters which achieves massive scalability by suitable parallelization. Thirdly, we theoretically characterize the set of spatial functions generated from the proposed model and prove consistency of the posterior distribution under mild regularity conditions. The approach is intuitive, easy to implement, and is buttressed by theoretical results that we develop here. Excellent empirical performances are illustrated using several simulation experiments and a geostatistical analysis of the sea surface temperature data from the pacific ocean.

Luis Gutierrez, (Pontificia Universidad Católica de Chile)

Bayesian nonparametric hypothesis testing for comparing several treatments with a control

We propose a flexible strategy to test for differences between a control group and several treatment regimes. While the existing tests for this type of comparison are based on the differences between location parameters, our approach identifies differences across the entire distributions. The procedure is developed under a Bayesian nonparametric approach, which avoids strong modelling assumptions over the distributions of each treatment or population. As a by-product, we make inferences on the shift function, which allows us to quantify the differences between two populations and identifies which part of the population has a treatment effect. The procedure is illustrated and compared to current hypothesis testing procedures under simulated scenarios. An application to a real data set is also presented.

Jarno Hartog, (University of Amsterdam)

Nonparametric label prediction on a graph

We consider classification problems on graphs, where the graph is given by the application context and the vertices are of different types, coded by two possible values. The available data are noisy observations of some of the labels. The goal is to classify the vertices correctly, including those for which there is no observation available. The idea is that typically, the location of a given vertex in the graph, in combination with (noisy) information about the labels of vertices close to it, should have predictive power for the label of the vertex of interest. We present an implementation of a nonparametric Bayesian approach to solve these binary classification problems on graphs. We consider a hierarchical Bayesian approach with a randomly scaled Gaussian prior. We have implemented the procedure with a theoretically optimal prior and a variant with a prior that exploits partial conjugacy and has some more flexibility and illustrate our methods with simulated data and real data examples.

Boris Hejblum, (Inserm, Bordeaux Population Health Research Center, UMR 1219 ; Univ. Bordeaux, ISPED ; INRIA SISTM ; VRI)

Sequential Dirichlet Process Mixtures of Multivariate Skew t-distributions for Model-based Clustering of Flow Cytometry Data

We propose a Bayesian nonparametric approach using Dirichlet process mixtures (DPM) of multivariate skew t-distributions to perform model based clustering. DPM models directly estimate the number of clusters from the data, avoiding model selection issues, and skew t-distributions provides robustness to outliers and non-elliptical shape of clusters. In addition, we developed a sequential strategy relying on a parametric approximation of the posterior to accommodate repeated measurements. This sequential strategy allows the analysis of multiple samples and is especially beneficial for fitting small clusters. We apply this method to flow-cytometry data, a high-throughput technology used to quantify single cells markers to identify cell sub-types, and to count the number of cells sampled for each sub-type. Improvements of this technology lead to the ability of describing millions of individual cells from a blood sample, resulting in high-dimensional datasets whose manual analysis is highly time-consuming and poorly reproducible. While several methods have been developed to perform automatic recognition of cell populations, most of them treat and analyze each sample independently. However, in practice, individual samples are rarely independent (e.g. longitudinal studies). We illustrate the good performance of our method on simulations, on an experimental benchmark dataset, and on new longitudinal data from the DALIA-1 trial, a clinical trial evaluating a therapeutic vaccine against HIV. On the benchmark dataset, the sequential strategy outperforms all other methods evaluated, and similarly, leads to improved performance on the DALIA-1 data. This method has been implemented in the R package NPflow, available on CRAN.

Katherine Heller, (Duke University)

Accurate and Real Time Prediction of Sepsis Using an End-to-End Gaussian Process RNN Classifier

We present a scalable end-to-end classifier that uses streams of multivariate physiological data to accurately predict onset of sepsis, a life-threatening complication from infections. Early detection and aggressive treatment of sepsis can substantially improve patient outcomes, motivating the development of sepsis early warning scores. Existing early warning scores in the medical literature are overly simplistic, assign independent scores to a small number of vitals, and as such suffer from extremely high false alarm rates. However, our method utilizes over 30 vital and lab variables, baseline covariates, and 8 different classes of administered medications. Our proposed framework overcomes current limitations using multitask Gaussian processes to model the trajectory of labs and vitals during a patient encounter, learning both temporal structure and correlations between variables. The Gaussian processes seamlessly account for the uncertainty and potential missingness in the multivariate clinical data. Information on medications is incorporated into the mean function, allowing the model to learn how medications affect the trajectory of the measured labs and vitals. The latent function values from the process, along with baseline covariates and indicators for medications, are then directly fed into a deep recurrent neural network to classify encounters by risk of developing sepsis. We train the model discriminatively in an end-to-end fashion using backpropagation, jointly learning Gaussian process and neural network parameters. In a cohort of 55,000 inpatient encounters at the Duke University Health System we find that it outperforms several established baselines, with substantially lower false alarm rates

Gudmund Hermansen, (University of Oslo)

Bernshtein–von Mises theorems for a class of Bayesian nonparametrics setups for stationary time series
There exist various parametric and nonparametric modelling strategies for stationary Gaussian time series. For most parametric approaches it is fairly easy to apply Bayesian techniques, where the large-sample behaviour is well understood, associated with Bernshtein–von Mises theorems. It is not as obvious how to proceed with a Bayesian nonparametric approach, however. Also, establishing good large-sample behaviour, like posterior consistency or Bernshtein–von Mises theorems, for nonparametric constructions becomes much more challenging. Here we will apply a certain class sample size dependent priors that model the spectral density of the time series as constant over an increasing refined partition of frequency domain. This enables Bayesian nonparametric analysis of stationary Gaussian time series. Furthermore, inference is easily carried out via approximations or simulations, and we are also able to establish precise and sufficient conditions needed to obtain Bernshtein–von Mises type of results for this particular construction.

Reyhaneh Hosseini, (University of Ottawa)

A Bayesian nonparameric estimation of mode
The Bayesian nonparametric plays a crucial role in statistical inference. The Dirichlet process perhaps is the most popular prior in Bayesian nonparametric statistics and it has been applied in many different areas of statistical inference. The most common applications of Dirichlet process are in density estimation. One of the applications of density estimation is to estimate the mode of a distribution. This paper adopts a Bayesian nonparametric approach to estimating the mode of a unimodal distribution. The method employed here is inspired by considering a Dirichlet process prior on the distribution of data. Then, we estimate the mode of the distribution by the Dirichlet process posterior. An example with simulation results is also provided.

Jonathan Huggins, (MIT)

Truncated Random Measures
Completely random measures (CRMs) and their normalizations are a rich source of Bayesian nonparametric priors. Examples include the beta, gamma, and Dirichlet processes. In this paper we detail two major classes of sequential CRM representations—series representations and representations—within which we organize both novel and existing sequential representations that can be used for simulation and posterior inference. These two classes and their constituent representations subsume existing ones that have previously been developed in an ad hoc manner for specific processes. Since a complete infinite-dimensional CRM cannot be used explicitly for computation, sequential representations are often truncated for tractability. We provide truncation error analyses for each type of sequential representation, as well as their normalized versions, thereby generalizing and improving upon existing truncation error bounds in the literature. We analyze the computational complexity of the sequential representations, which in conjunction with our error bounds allows us to directly compare representations and discuss their relative efficiency. We

include numerous applications of our theoretical results to commonly-used (normalized) CRMs, demonstrating that our results enable a straightforward representation and analysis of CRMs that has not previously been available in a Bayesian nonparametric context.

Colman Humphrey, (University of Pennsylvania, Wharton School)

Analysis of Urban Vibrancy and Safety in Philadelphia: Matching on Intersections
Urban data analysis has been recently improved through publicly available high resolution data, allowing us to empirically investigate urban design principles of the past half-century. Following our initial case study of Philadelphia, we further explore the association between safety and business vibrancy, constructed from our business database, placing it on a causal level. We develop a non-parametric matching method based on the city's intersections, using Bayesian methods to match intersections to best replicate a randomized experiment. We now find a strong causal link between longer opening hours and fewer crimes. In contrast to our associative initial work, we no longer find a significant relationship between business density and crime.

Matteo Iacopini, (Ca' Foscari University of Venice and Université Paris 1 - Panthéon-Sorbonne)

Bayesian Nonparametric prior for sparse time-varying graphs
The study of graphical models has become a central topic in statistics in the last few years, with a vast range of applications especially in machine learning, biostatistics and econometrics. The complexity of real world phenomena under study has highlighted the limits of existing tools as well as the need for flexible statistical models able to describe several properties of real networks such as sparsity and clustering and their dynamic evolution. Bayesian Nonparametric models have proven to be a powerful and flexible instrument capable to learn complex features from data under minimal assumptions. We aim at contributing to this literature by proposing a new random measure as a prior for the density of time-varying graphs. We build on the theory of completely random measures and exploit the underlying Poisson Process to create a measure able to generate sparse graphs, where the number of edges grows sub-quadratically with respect to the number of nodes. To capture both the presence of clusters and the variation over time of the network resulting from the disappearance of edges or the change of the clusters, we specify a discrete-time Dependent Dirichlet Process on the edge parameters.

Vanda Inacio de Carvalho, (University of Edinburgh)

The receiver operating characteristic (ROC) surface is a popular tool for evaluating the accuracy of medical diagnostic tests that classify individuals into one of three ordered classes. We propose a fully nonparametric method based on the Bayesian bootstrap for conducting inferences for the ROC surface and its functionals, such as the volume under the surface. The proposed estimator is based on a simple, yet interesting, representation of the ROC surface in terms of placement variables. A simulation study and a real data application concerning the diagnosis of neurological impairment in HIV patients are provided.

Moumita Karmakar, (Icahn School of Medicine at Mount Sinai, New York)

Bayesian estimation of sufficient dimension reduction space

Subspace estimation in complex non-linear models is of growing interest in modern biomedical applications. Classical approaches to this problem are based on adhoc multistage approaches which do not fully account for the uncertainty in estimation. To address this, we develop a Bayesian model based on inverse regression to estimate the Sufficient Dimension Reduction (SDR) subspace. The key component of the model is the reduction matrix, an element in the Grassmannian, which spans the SDR subspace. Using a uniform prior distribution on the Grassmannian, we develop a fully Bayesian specification and associated posterior sampling algorithm. Consistency of the Bayes estimators are established under reasonable assumptions on the data generating process. In addition, a method for variable selection is proposed based on the posterior samples of the reduction subspace. The efficacy of the Bayesian procedure is demonstrated via numerous simulation studies and a real data example.

Milad Kharratzadeh, (Columbia University)

Hierarchical Bayesian Spline Models

We study the problem of Bayesian nonparametric regression using hierarchical spline models. We start by introducing a Bayesian framework for fitting spline models in Stan probabilistic programming language. Implementing splines in Stan allows us to take advantage of its Hamiltonian Monte Carlo (HMC) sampler to efficiently explore the parameter space. We then study the effects of different informative and weakly informative priors. Specifically, we are interested in hierarchical structures where the spline parameters for different variables are related in higher levels. We also address the process of selecting spline knots; in practical problems, it is not always clear how to choose the number/location of the knots. In absence of appropriate priors, choosing too many/too few knots may lead to overfitting/underfitting. We introduce random-walk priors that alleviate these problems to a great extent; by enforcing consecutive B-spline coefficients to be not too far from each other, we enforce smoothness over the fitted curves. Finally, we use our hierarchical spline models for fitting two real-world problems. The first one is the opinion of people on gay marriage over the years which we regress on demographic information. The second problem is the survival analysis of patients with acute myeloid Leukemia (AML). Through these two problems, we study different aspects of our Bayesian spline models.

Samer Kharroubi, (American University of Beirut)

Rationale: The SF-6D was proposed by Brazier et al as a health state descriptor, derived from the SF-36, for the purposes of assigning a HRQoL utility to health states. However, their utility measure suffers from problems of non-monotonicity and failure to characterise states with very low utility. Objectives: The objectives of this research were (a) to develop a novel Bayesian statistical method to analyse individual health state utility data, and (b) to apply this method to the original SF-6D sample data in order to derive an improved utility measure for the SF-6D. Method: A new Bayesian nonparametric model was built to describe the intrinsic characteristics of individual health state valuation data more faithfully than previous statistical models have done. In particular, the model allows for the repeated measures feature, expresses the skewness that is found in such data and incorporates a covariance structure in the prior distribution that mitigates against non-monotonicity. Results: The new model is found to validate very well. The new UK SF-6D utility measure fits the data better than the previous analysis, avoids non-monotonicity and estimates lower utilities for the worst health states. The method also characterises predictive uncertainty in the utility measure. Conclusions: The new Bayesian nonparametric method is a powerful technique that might be applied to devise HRQoL utility measures for a wide range of health state descriptive systems. Its application to the SF-6D overcomes problems in the previous utility measure and should be the preferred UK valuation for future uses of this generic instrument.

Alisa Kirichenko, (University of Amsterdam)

Function estimation on a large graph using Bayesian Laplacian regularisation

We study a Bayesian approach to estimating a smooth function in the context of regression or classification problems on large graphs. We present a mathematical framework that allows to study the performance of nonparametric function estimation methods on large graphs and we obtain minimax convergence rates for these problems within the framework. We derive theoretical results that show how asymptotically optimal Bayesian regularisation can be achieved under an asymptotic shape assumption on the underlying graph and a smoothness condition on the target function, both formulated in terms of the graph Laplacian. The priors we study are randomly scaled Gaussians with precision operators involving the Laplacian of the graph. We also consider the case of missing observations and investigate the generalisation performance for various missing mechanisms.

Ilja Klebanov, (Zuse Institute Berlin)

Information-theoretic concepts for empirical Bayes methods

When estimating a probability density within the empirical Bayes framework, the non-parametric maximum likelihood estimate usually tends to overfit the data. This issue is usually taken care of by regularization, which, from the Bayesian perspective, can be seen as choosing a hyperprior on the set of possible priors. However, the majority of regularizations currently in use lack invariance under transformation of the parameters (reparametrization) and measurements, resulting in inconsistent statistical inference. We will show how information-theoretic concepts can provide transformation-invariant regularizations that favor non-informativity, show their relation to reference priors and compare them to Dirichlet process hyperpriors.

Guillaume Kon Kam King, (Università degli Studi di Torino, Collegio Carlo Alberto)

Bayesian Nonparametric modelling of locally trended functional time series

Functional time series naturally appear in contexts where phenomena are measured regularly. Examples include the income distribution over time, the evolution of molecular size distribution during polymerisation, or daily demand/offer curves in an exchange market. Trends are common in these series: higher incomes might tend to increase while lower incomes stagnate or decrease, polymerisation increases molecule sizes globally, and prices commonly show rising or falling trends. The functional nature of the data raises a challenge for the inference and indeed, the likelihood can be intractable in the case of fully observed functions. We present a likelihood-free approach for functional data forecast with a trend phenomenon. We develop a bayesian nonparametric model based on a dependent process. It builds on particle system models, which originate from population genetics. This construction provides a means to flexibly specify the correlation of the dependent process. We take advantage of the expressiveness of interacting particle models to embed a local and transient trend mechanism. To this aim, we draw inspiration from interaction potentials between physical particle systems in molecular dynamics. We perform the likelihood-free inference by means of Approximate Bayesian Computation (ABC). We discuss the elicitation of informative summary statistics for stochastic processes building on the idea of semi-automatic summaries. Coupled with a population ABC, this results in a very versatile inference method. We show the increased robustness of the trended model and comment on the generality of our approach for building functional forecast models.

Athanasios Kottas, (University of California, Santa Cruz)

Bayesian semiparametric methods for emulation and calibration of stochastic computer simulators
Stochastic computer simulators are increasingly used in technology and science to model random systems, e.g., in population dynamics, biological processes, and nuclear interactions. We propose a framework for emulation and calibration of complex stochastic simulators. The emulator is built from a general nonparametric mixture model for the joint distribution of the simulator inputs and output(s). For calibration, we develop a semiparametric method to link the simulator data with field observed data using a modular, two-stage approach. In the first stage, the posterior distribution of the emulator parameters is estimated based solely on the simulator data. Then, in the second stage, the posterior distribution of the calibration parameters is approximated based on the field data as well as the posterior distribution of the emulator parameters obtained in the first stage. The methodology is applied to PROPSET, a stochastic simulator that models the bombardment of a device with high-energy protons in order to study the effect of radiation on spaceborne microelectronics devices.

Suprateek Kundu, (Emory University)

Scalable Bayesian Variable Selection for Structured High Dimensional Data
Variable selection for structured covariates lying on an underlying known graph is a problem motivated by practical applications and has been a topic of increasing interest. However, most of the existing methods may not be scalable to ultra-high dimensional settings involving tens of thousands of variables lying on known pathways such as the case in genomics studies. We propose an adaptive Bayesian shrinkage approach which incorporates prior network information by smoothing the shrinkage parameters for connected variables in the graph so that the corresponding coefficients have a similar degree of shrinkage. The proposed approach induces data adaptive shrinkage, where connected variables are encouraged to have an equal importance in the model, without requiring the corresponding regression coefficients to have similar magnitudes. We fit our model via a computationally efficient expectation maximization algorithm which scalable to high dimensional settings ($p \sim 100{,}000$ and beyond). Theoretical properties such as the oracle property and variable selection consistency are established for fixed as well as increasing dimensions, even when the number of variables increases exponentially with the sample size, and even when the true graph is misspecified. We demonstrate the advantages of our approach in terms of variable selection, prediction, and computational scalability via a simulation study, and apply the method to a cancer genomics study. Numerical results illustrate that the proposed approach is computationally slightly slower than standard lasso-based regression approaches, however, they result in substantial gains in performance compared to existing structured and unstructured variable selection approaches.

Minjung Kyung, (Duksung Women's University)

In the field of molecular biology, it is often of interest to analyze microarray data for clustering genes based on similar profiles of gene expression to identify genes that are differentially expressed under multiple biological conditions. One of the notable characteristics of a gene expression profile is that it shows a cyclic curve over a course of time. To group sequences of similar molecular functions, we propose a Bayesian Dirichlet process mixture (DPM) of linear regression models with a Fourier series for the regression coefficients, for each of which a spike and slab prior is assumed. A full Gibbs-sampling algorithm is developed for an efficient MCMC posterior computation. Due to the so-called "label-switching" problem and different numbers of clusters during the MCMC computation, a post-process approach of Fritsch and Ickstadt (2009) is additionally applied to MCMC samples for an optimal single clustering estimate by maximizing the posterior expected adjusted Rand index with the posterior probabilities of two observations being clustered together. The proposed method is illustrated with two simulated data and one real data of the physiological response of fibroblasts to serum of Iyer et al. (1999).

Juhee Lee, (University of California, Santa Cruz)

Joint Bayesian Semiparametric Regression Analysis of Recurrent Adverse Events and Survival

We propose a Bayesian semiparametric joint regression model for a recurrent event process and survival time. Assuming independent latent subject frailties, marginal models for each subject are defined for the recurrent event process and survival distribution as functions of the subject's frailty and covariates. Averaging over the frailty distribution yields a joint model. A robust Bayesian model is obtained by assuming a Dirichlet process for the frailty distribution. The model is applied to analyze an observational dataset from esophageal cancer patients treated with radiation therapy, including re- current effusions of fluid to the heart or lungs. A simulation study is presented that compares posterior estimates under the joint model to simpler Bayesian models that ignore either the recurrent event process or survival time, and also to a frequentist joint model. The simulations show that the proposed Bayesian joint model does a good job of correcting for treatment assignment bias, is robust to a mild violation of the frailty model assumption, and has favorable estimation reliability and accuracy compared with each of the Bayesian sub-models and the frequentist joint model.

Juho Lee, (POSTECH)

Finite-dimensional BFRY processes for power-law models

Bayesian nonparametric methods based on the Dirichlet Process (DP), gamma process and beta process, have proven effective in capturing aspects of various datasets arising in machine learning. However, it is now recognized that such processes have their limitations in terms of the ability to capture power law behavior. As such there is now considerable interest in models based on the Stable Processs (SP), Generalized Gamma process (GGP) and Stable-Beta Process (SBP). These models present new challenges in terms of practical statistical implementation. In analogy to tractable processes such as the finite-dimensional Dirichlet process, we describe a class of random processes, we call iid finite-dimensional Bertoin-Fujita-Roynette-Yor (BFRY) processes. We show that the three types of finite-dimensional BFRY processes converge to SP, GGP and SBP by showing the convergence of Laplace functionals. Our approach enables one to begin to develop efficient posterior inference algorithms such as variational Bayes that readily scale to massive datasets.

Didong Li, (Duke University)

Bayesian manifold learning using locally curved basis functions

It is very common to have an interest in identifying lower dimensional structure in high-dimensional data. A typical approach is to approximate the manifold or subspace by a set of hyperplanes. A key disadvantage of such approaches is the lack of statistical efficiency and parsimony of the representation, particularly when the true manifold has large curvature. For example, if the manifold is a circle with very small radius, a huge number of tangent lines will be needed to fit the circle well. Taking curvature into consideration, a reasonable basis for the manifold is space form, with three special cases: sphere, hyperbolic space and hyperplane. With this locally curved basis, a pseudo distance between any two sets of points is defined based on how well they can be fit by a space form. In some sense, this model generalizes the linear model, since hyperplane is a special space form, with zero curvature, but the new models will perform much better when the manifold has large curvature. Taking a nonparametric Bayes approach to inference, theoretical support is provided, including results showing significant gains in reduction in the number of components needed to provide an approximation within a given error tolerance. In addition, an efficient Markov chain Monte Carlo algorithm is developed for posterior computation, while also developing a simple and fast optimization algorithm as an alternative. The methods are assessed in comparison to a variety of frequentist and Bayesian competitors through toy simulation examples, and applications to several real-world data sets. Substantial practical gains are shown in these examples, including in cases in which the true subspace is multiple separated manifolds.

Frank Liu, (Monash University)

Robust Bayesian inference for moment condition models

This paper proposes a new robust Bayesian exponentially tilted empirical likelihood (RBETEL) inferential methodology suitable for moment condition models for data that may be contaminated by outliers. The work builds upon the foundations developed by Schennach (2005) who proposed the original Bayesian exponentially tilted empirical likelihood (BETEL) method, justified by the fact that an empirical likelihood can be interpreted as the nonparametric limit of a Bayesian procedure when the implied probabilities are obtained from maximizing entropy subject to some given moment constraints. After first demonstrating through a simulation exercise that BETEL posteriors are susceptible to interference from outliers, a Markov chain Monte Carlo framework is developed, incorporating outlier information derived from a robust frequentist estimator of multivariate location and scatter (Rocke,1996). Controlled simulation experiments are conducted to investigate the performance of the proposed RBETEL method, and finds that the new approach improves on BETEL when outliers are present.

Shuhei Mano, (Institute of Statistical Mathematics)

Multiplicative Measures on Partitions and A-Hypergeometric System

Multiplicative measures on partitions is an important class of exchangeable partition probability function, which include Gibbs partition and appear as sampling distributions in Bayesian nonparametrics. The conditional measure is an algebraic exponential family, whose normalization constant is the partial Bell polynomial. Especially, it is the A-Hypergeometric polynomial associated with the rational normal curve. The maximum likelihood estimator (MLE) of the full and the curved exponential families are studied in terms of the information geometry of the Newton polytopes. Especially, it is shown that the MLE does not exist for the full exponential family. Algebraic methods for the numerical computation are also provided.

Nicolo' Margaritella, (University of Edinburgh)

Bayesian Modelling of fMRI data

The modelling of brain activity has a long history but only in recent years it started benefiting from the extraordinary technological advances in the fields of neurophysiology and neuroimaging. With a remarkable amount of spatiotemporal data recordable simultaneously from several parts of the brain, researchers are challenged to find models that can capture meaningful patterns and structures behind such complexity. In this context, the Bayesian approach allows for a flexible modelling of the data and it is able to exploit information from previous studies to narrow down its posterior uncertainties. Thus motivated, we aim at modelling data from fMRI recordings employing functional data analysis within a Bayesian perspective; in particular, we want to find suitable solutions to account for the spatiotemporal nature of the data as well as to profit from an increasingly informative prior knowledge. We fit the constructed models in JAGS and present the corresponding results obtained.

Asael Fabian Martínez, (Centro de Investigación en Matemáticas, A.C.)

Mixture models and allocation schemes

In this work, we study mixture models under a random partition approach and making use of some of the theory of allocation schemes. With this framework, we explore some of the main drawbacks of mixture models such as label switching and non identifiability. Our proposed model is tested using synthetic and real datasets.

Gianluca Mastrantonio, (Polytechnic University of Turin)

Semi-parametric Bayesian change-point model
We introduce a semi-parametric Bayesian change-point model. In our proposal the model is expressed as a constrained hidden Markov model defining the latent time dynamic using a transformation of a Dirichlet process. The number of change points is then treated as random and estimated along with all the other model parameters. Posterior inference is carried out using a Markov chain Monte Carlo algorithm based on a marginalized version of the proposed model. We test the model using simulated examples and two real datasets, namely the coal mining disaster data, that is a widely used dataset for illustrative purpose, and a dataset of indoor radon recording. With the simulated examples we highlight the ability of the model in recovering the parameters and number of change points, and we compare our results with the ones of the state of the art models, showing a clear improvement in terms of change points identification. The results obtained on the coal mining disaster data are coherent with previous literature and the radon data study shows a clear segmentation of the observed time series.

Alexander Meier, (Otto-von-Guericke-University Magdeburg)

Bayesian nonparametric analysis of multivariate time series
While there is an increasing amount of literature about Bayesian time series analysis, only few nonparametric approaches to multivariate time series exist. Most notably, Rosen and Stoffer (2007) and Shibin Zhang (2016) rely on Whittle's likelihood, involving the second order structure of a time series by means of the spectral density matrix. The latter is modeled with a smoothing splines prior for the components of the Cholesky decomposition to ensure positive definiteness. Such a componentwise approach makes it difficult to employ prior knowledge about certain matrix characteristics such as eigenvalues. A different idea is to model the spectral density with the aid of random measures. This is in the spirit of Choudhuri et al. (2004), who model the normalized spectral density of a univariate time series by a Dirichlet process mixture of beta densities. Conceiving the Dirichlet process as a normalized Gamma process, an extension of this idea to multivariate time series is not straightforward, because the canonical matrix variate extensions of the gamma distribution lack infinite divisibility (Gupta and Nagar, 1999). Instead, we use the infinitely divisible matrix Gamma distributions from P{'e}rez-Abreu and Stelzer (2014) to construct a matrix Gamma process mixture of beta densities prior for the spectral density matrix. We investigate the performance of this prior in conjunction with Whittle's likelihood. We also consider a matrix Dirichlet process as obtained by normalization and a multivariate extension of the corrected parametric likelihood introduced by Kirch et al. (2017).

Xenia Miscouridou, (University of Oxford)

Exchangeable Random Measures for Sparse and Modular Graphs with Overlapping Communities.
We propose a novel statistical model for sparse networks with overlapping community structure. The model is based on representing the graph as an exchangeable point process, and naturally generalizes existing probabilistic models with overlapping block-structure to the sparse regime. Our construction builds on vectors of completely random measures, and has interpretable parameters, each node being assigned a vector representing its level of aliation to some latent communities. We develop methods for simulating this class of random graphs, as well as to perform posterior inference. We show that the proposed approach can recover interpretable structure from two real-world networks and can handle graphs with thousands of nodes and tens of thousands of edges

Anuj Mishra, (University of Pune)

Bayesian non-parametric modeling of durations in high frequency finance.

Recently, there has been a growing interest in studying the autoregressive conditional duration (ACD) models, originally introduced by Engle and Russell (1998). ACD models are useful for modeling the time between the events, especially, in the financial context, the time between trading of stocks. Though useful, various parametric forms of ACD models limit the scope and types of inference that can be drawn. In this paper, we relax parametric assumptions and introduce a fully non-parametric model for durations. The new duration model is based on a dependent Dirichlet process prior on a family of random probability measures indexed by the lagged durations. Posterior inference is accomplished using three different MCMC schemes. The first is the marginal method which relies on integrating out the random distribution function from the model, thus removing the infinite-dimensional parameters. The second method is the truncation method where, instead of integrating out the DP, an approximation to DP is considered. The third one is based on slice sampling which works by sampling a finite but sufficient number of variables at every iteration. The model is tested under a simulation study and is illustrated with the analysis of the high-frequency data for the year 2015 from National Stock Exchange (NSE), India.

Masahiro Nakano, (NTT communication science labratories)

Infinite Number Place

We present a Bayesian nonparametric model of Number Place (also called Sudoku) puzzle. Number Place is a combinatorial number-placement puzzle: Given a $n^2 \times n^2 (n \in \mathbb{N})$ grid, the objective is to fill them with digits so that each column, each row, and each of the $n \times n$ sub-grids contains all of the digits from 1 to n. We discuss the construction of such stochastic processes with combinatorial constraints. Currently, most of the stochastic processes can be described by the projective limit representations based on Kolmogorov's extension theorem, involving stochastic enumeration algorithms. However, to our knowledge, there have been no successful constructions for the stochastic processes with combinatorial constraints which leads to no available enumeration. To address such difficult problems, we propose an indirect construction: we first introduce the auxiliary projective system using the measure factorization proposed in NIPS, 2010, then extend it to a projective limit, and finally describe the expanded stochastic process as a mapping from the projective limit of the auxiliary system. We apply this strategy to the construction of the Bayesian nonparametric model of Number Place puzzle.

Zacharie Naulet, (University of Toronto)

Application of Sparse Exchangeable Graphs to Infinite Latent Feature Modelling

Bayesian non-parametric feature allocation schemes are built on top of a prior distribution over infinite binary matrices; typically using the Beta-Binomial process or its combinatorial structure the Indian Buffet Process. Recent work in statistical network modelling has introduced sparse exchangeable graph distributions: a large class of new probability distributions over random graphs with an infinite number of vertices. A bipartite graph drawn from one of these distributions may be viewed as defining an infinite feature allocation by identifying one partition with features and the other partition with the objects to which features are assigned. As such, sparse exchangeable graph distributions define Bayesian non-parametric priors for feature allocations. We show that this framework naturally includes previous models, including the Indian buffet process, and we demonstrate that the model class can accommodate a range of behaviours outside the purview of traditional Bayesian non-parametrics. We also show how sampling schemes for these distributions can be used to make inferences in latent feature allocation tasks.

Yang Ni, (The University of Texas at Austin)

Bayesian nonparametric reciprocal graphical models
We develop novel hierarchical reciprocal graphical models to infer gene networks from heterogeneous data. In the case of data that can be naturally divided into known groups, we propose to connect graphs by introducing a hierarchical prior across group-specific graphs, including a correlation on edge strengths across graphs. Thresholding priors are applied to induce sparsity of the estimated networks. In the case of unknown groups, we use Pitman-Yor process mixtures to cluster subjects into subpopulations and jointly estimate cluster-specific gene networks. We illustrate the proposed approach by simulation studies and two applications in multiplatform genomic data for multiple cancers.

Luis Nieto-Barajas, (ITAM)

A Bayesian nonparametric dynamic AR model for multiple time series analysis
In this talk we present a Bayesian nonparametric model for the analysis of multiple time series. We consider an autoregressive structure of order p for each of the series and borrow strength across the series by considering a common error population that is also evolving in time. The error populations (distributions) are assumed nonparametric whose law is based on a series of dependent Polya trees with zero median. This dependence is of order q and is achieved via a dependent beta process that links the branching probabilities of the trees. We study the prior properties and show how to obtain posterior inference. The model is tested under a simulation study and is illustrated with the analysis of the economic activity index of the 32 States of Mexico.

Bernardo Nipoti, (Trinity College Dublin)

Stratified survival regression with Bayesian nonparametric mixtures
The stratified proportional hazards model accounts for heterogeneity in the baseline hazard rate among different groups or strata. Usually strata are determined by means of some categorical covariate. We consider a stratification that takes into account all regression variables and is induced by a Bayesian nonparametric mixture model. We specialise our approach to the frameworks of the proportional hazards and the accelerated life models. Conditionally on the stratification induced by the model, inference is produced for each stratum. We illustrate the performance of our approach by analysing both synthetic and real data.

Akihiko Nishimura, (Duke University)

Gaussian process for functional regression and causal inference: application to the study of long-term air pollution exposure to cardiovascular diseases
We present a novel application of Bayesian non-parametric methods to an analysis of a complex observational data from a large-scale epidemiological study. The goal of the study is to assess the cumulative effect of long-term air pollution exposure to one's cardio-vascular health. We employ Gaussian process priors to 1. develop a non-linear distributed lag model that is flexible yet familiar and accessible to epidemiologists and 2. control for unobserved confounding factors with temporal structure. Both a latent dose-response curve and log-transformed lag weight are given Gaussian process priors to allow for non-linearity. The model is amenable to posterior inference via Hamiltonian Monte Carlo after suitable reparametrization and can be easily implemented with a probabilistic programming language Stan. One complication in analyzing the data is the presence of unobserved confounding factors with temporal structure. Since air pollution levels exhibit a strong temporal pattern, its effect on an outcome variable cannot be estimated properly without controlling for the temporal confounding factors. We address this problem by introducing a temporal Gaussian process to our model, whose range parameter is weakly constrained to be larger than a typical range of correlations

in air pollution levels. As we demonstrate, this leads to a more biological interpretable estimate of the dose-response curve.

Andriy Norets, (Brown University)

Sparsity and Dimension Reduction in Bayesian Conditional Density Estimation
We study asymptotic behavior of the posterior distribution in a Bayesian nonparametric model for conditional densities when the number of covariates is very large. The model is based on a mixture of normal distributions with mixing weights that depend on covariates

$$p(y|x) = \sum_{j=1}^{m} \frac{\alpha_j \exp\{-0.5||x-\mu_j^x||^2/\sigma^2\}}{\sum_{i=1}^{m} \alpha_i \exp\{-0.5||x-\mu_i^x||^2/\sigma^2\}} \frac{\exp\{-0.5||y-\mu_j^y||^2/\sigma^2\}}{(2\pi\sigma^2)^{d_y/2}}$$

A prior on m is specified. The marginal distribution of x is not modeled. When the covariates dimension d_x is fixed, Norets and Pati (Econometric Theory, 2016) show that the posterior in this model contracts at an adaptive optimal rate up to a log factor; moreover, when y depends only on first $d_x^0 \leq d_x$ components of x, the rate depends on d_x^0 rather than d_x. In the present paper, we assume that d_x^0 is also fixed but the number of potential covariates d_x can increase with the sample size n, $d_x \leq e^{n^{\alpha}}$. We show that under a sparse prior on the covariates selection similar to the one in Shen and Ghosal (Bernoulli, 2016), the rate is bounded above by

$$\max\{n^{-\beta/(2\beta+d_y+d_x^0)}(\log n)^t, (\log d_x/n)^{0.5}\},$$

where β is the smoothness level of the true conditional density and $t > 0$ is a constant. Our ongoing work involves an extension to these results to allow for d_x^0 that increases with n. We also consider dimension reduction settings with fixed d_x and the true conditional density dependent only on Bx, where B is a $d_x^0 \times d_x$ matrix. We show that when scalar σ^2 is replaced by a scaling matrix in the mixing weights, the posterior contraction rate can be bounded by $n^{-\beta/(2\beta+d_y+d_x^0)}$ up to a log factor.

Nurzhan Nurushev, (VU Amsterdam)

Uncertainty quantification for biclustering model
We study the problem of inference on the unknown parameter in the biclustering model by using the penalization method which originates from the empirical Bayes approach. The underlying biclustering structure is that the high-dimensional parameter consists of a few blocks of equal coordinates. The main inference problem is the uncertainty quantification (i.e., construction of a conference set for the unknown parameter), but on the way we solve the estimation and posterior contraction problems as well. We pursue a novel local approach in that the procedure quality is characterized by a local quantity, the oracle rate, which is the best trade-off between the approximation error by a biclustering structure and the best performance for that approximating biclustering structure. The approach is also robust in that the additive errors in the model are not assumed to be independent with some known distribution, but (in general dependent) only satisfying certain mild exchangeable exponential moment conditions. We introduce the excessive bias restriction under which we establish the local (oracle) confidence optimality of the proposed confidence ball. Adaptive minimax results (for the graphon estimation and posterior contraction problems) follow from our local results. The results for the stochastic block model follow, with implications for network modeling.

Kurdistan Omar, (University of Leicester, Mathematics Deparment)

Nonparametric method for functional regression with multivariate functional responses

Ferraty, Van Keilegom, and Vieu (2012) proposed the noparametric regression model when the response and the covariate are both functional. In this paper, we expand upon their research by proposing two new models: we first extend the Individual functional responses to multivariate functional responses; secondly we take the correlation among different functional responses into account. We utilize the functional kernel with an automatic bandwidth selection by the cross-validation procedure and a semi-metric build on the functional principle component analysis for the measure of the proximity between curves. The performance of these models is measured by calculating mean square error. The proposed models are illustrated through a simulation data and a real data example in UK weather. After we checked all the scenarios, the multivariate functional responses with correlation model performs better than other models, in the sight that the mean square error is lower. The multivariate functional responses with correlation model takes advantage of covariance between different functional responses and therefore improves the accuracy of prediction.

- Ferraty, F., Van Keilegom, I., and Vieu, P. (2012),'Regression when both response and predictor are functions', *Journal of Multivariate Analysis*, 109, 10–28.

Garritt Page, (Brigham Young University)

Interaction Discovery using Covariate Informed Product Partition Models

It is common in studies that collect covariates on subjects/experimental units that learning which and how covariates interact is of principal interest. For example, it would be very useful to medical personnel diagnosing presence/absence of a disease to know how time varying physiological and baseline covariates (e.g., gender, age, blood pressure, etc.) interact. However, with no *a priori* guidance from experts finding interactions becomes increasingly difficult particularly as the number of covariates grow. This statistical problem has been studied extensively with the majority of methods proposing some sort of multiplicity test corrections or some stochastic search. Our approach (though exploratory in nature) carries out a very general interaction search without expert guidance and without explicitly including interactions in a data model. This is done by connecting covariates to partitions through a covariate dependent partition prior and employing association rules to discover potential cluster-specific interactions. We illustrate the procedure with a small simulation study and a real world data example.

Freddy Palma, (Instituute of Research in Applied Mathematics and Systems (IIMAS) - National Autonomous University of Mexico (UNAM))

A dual Markov construction for filtering problems

Within a hidden Markov framework, the law driving the signal process is of major interest in a wide variety of fields. Indeed, in such a context, the rol played by the transition probability associated such a process turns out to be crucial for the computation of the filters. Here, we study a construction of continuous-time reversible Markov process with arbitrary but given invariant distribution, that, when assumed for the signal, leads us to computable and tractable expressions for the filters. Also, it allows us to calculate statistics associated to the filters. Furthermore, the construction guarantees the existence of a dual to the signal, which can be used to derive an alternative filter recursions. Some well know models fall within our construction.

Sayan Patra, (Duke University)

SINC processes: Avoiding discrete mixtures in modeling unknown densities

This article proposing a novel class of Spline INverse Cdf (SINC) processes for unknown densities. The SINC approach relies on approximating the inverse cdf for the unknown density using splines. Including a nugget and introducing uniform latent variables, one obtains a simple Gaussian nonparametric latent variable regression model. Efficient posterior computation is deceptively hard under this model, but we develop an approach that marginalizes out the unknown regression function to obtain an analytic form for the posterior distribution of the latent variables. Leveraging on this form, an efficient algorithm is developed for posterior computation. Theoretical support is provided on accuracy as the number of knots in the spline approximation increases, while also developing asymptotic theory on statistical performance in estimating the unknown density. Comprehensive simulation studies are run comparing SINC with common competitors ranging from frequentist kernel density estimation to Dirichlet process mixtures of Gaussians. It is shown that there are substantial computational advantages to our implementation of SINC, which bypasses thorny issues related to the use of discrete mixture models, such as label switching. In addition, a substantial advantage is ease of application and generalization of SINC in challenging cases ranging from conditional density estimation to multivariate unknown densities to hierarchical models involving one or more unknown densities. Some such extensions are described and illustrated through applications.

Iliana Peneva, (Mathematics for Real-World Systems CDT, University of Warwick)

A Bayesian Nonparametric Model for Finding Novel Pancreatic Cancer Subtypes

Pancreatic cancer is one of the most common causes of cancer-related death in the developed world and is projected to become the second most common in the Western societies by 2030. In this work, we develop a novel Bayesian nonparametric model for combined data integration and clustering called BayesCluster which aims to identify novel subtypes of pancreatic cancer. The classical clustering methods such as k-means and hierarchical clustering fail to capture the common structure shared by the different datasets about the same set of patients, and also to offer insight into the molecular mechanisms underlying the biology of pancreatic cancer. The proposed method can integrate and use the information from multiple different datasets, and can be easily extended to richer biological structure by using sparsity priors and the kernel trick. We incorporate feature learning in the model because of the large number of predictors and use a Dirichlet Process mixture modelling approach and MCMC sampler to produce the patient subgroups. We apply the model to the datasets from the Cancer Genome Atlas (TCGA) Pancancer Project of Pancreatic adenocarcinoma, which contain clinical and biological data about PC patients, followed up to 5 years. By combining all available information we are able to be better identify the different subtypes of pancreatic cancer, opening the way for new personalised treatments.

Valerio Perrone, (University of Warwick)

Poisson Random Fields for Dynamic Feature Models

We present the Wright-Fisher Indian buffet process (WF-IBP), a probabilistic model for time-dependent data assumed to have been generated by an unknown number of latent features. This model is suitable as a prior in Bayesian nonparametric feature allocation models in which the features underlying the observed data exhibit a dependency structure over time. More specifically, we establish a new framework for generating dependent Indian buffet processes, where the Poisson random field model from population genetics is used as a way of constructing dependent beta processes. Inference in the model is complex, and we describe a sophisticated Markov Chain Monte Carlo algorithm for exact posterior simulation. We apply our construction to develop a nonparametric focused topic model for collections of time-stamped text documents and test it on the full corpus of NIPS papers published from 1987 to 2015.

Valerie Poynor, (California State University, Fullerton)

Bayesian nonparametric modeling for order constrained mean residual life functions

In survival analysis, two populations are mean residual life (mrl) ordered if the mean residual life function of one population is higher than the other for the entire survival domain. There are certain applications that focus on the comparison of survival distributions in which the researcher expects that the average remaining lifetime for one population is higher than that of the other population given survival up to a particular time. For instance, in treatment and control experimental groups, researchers may have knowledge that the population receiving the treatment has a longer remaining life expectancy at all times. Incorporating this information improves inferential accuracy and certainty. Contrary to stochastic, hazard rate, and a number of other types of ordering constraints, modeling under mrl order constraint has been comparably neglected. We present a Bayesian nonparametric model for inference on mrl ordered populations that does not require the stronger hazard rate order assumption. We use a mixture of gamma distributions with common scale parameter, giving rise to stochastically ordered kernel distribution, and mix over the shape parameter via hazard rate ordered mixing distributions. The mixing distributions are constructed from two independent Dirichlet process priors. We are able to demonstrate the posterior benefits of our model by comparing results from a data set on small cell lung cancer, consisting of a number of right censored survival times, under analogous Bayesian nonparametric models that do not incorporate the mrl ordering assumption.

Anthony Quinn, (Department of Statistics, University of California at Berkeley)

Fully Probabilistic Design of Nonparametrically Conditioned Distributions

We address the problem of conditioning on stochastic knowledge, K, for the purpose of updating a distribution of a quantity of interest, θ. Standard Bayesian conditioning accomplishes this, *if* a joint model of θ and K is specified. However, this is an unreasonable assumption when conditioning on a nonparametric process. It is precisely the situation we confront in many instances of (Bayesian) transfer learning, where K is made available in the form of an externally supplied and computed distribution. The minimum KLD principle has a long history as an optimal strategy for probabilistic model design in incompletely specified contexts, such as this one. Variants such as MaxEnt provide an estimate of the conditional (i.e. knowledge-constrained) distribution, $\hat{f}(\theta|K)$, replacing a prescriptive Bayes' rule in these incompletely modelled contexts. In this contribution, we will adopt a fully hierarchical setting, in which the incompletely specified conditional distribution is, itself, modelled as a nonparametric process, something which has not been done in standard deployments of the minimum KLD principle in the past. It allows exploration and quantification of the uncertainty in the conditional design. We will examine some applications in (statically) interacting Kalman filters, and in centralized merging of predictors in distributed probabilistic knowledge processing.

Fernando Quintana, (Pontificia Universidad Católica de Chile)

Covariate-Dependent Mixture Models Induced by Determinantal Point Processes and Some Applications

We consider mixture models where latent parameters are a priori encouraged to separate from each other (a property we refer to as repulsion), thus inducing well separated clusters. We specifically consider the case where repulsion is induced via stationary Determinantal Point Processes (DPPs). The DPPs need not have explicit spectral representations. We consider a covariate-dependent clustering assignment induced by the model and illustrate the results using synthetic and publicly available data. Posterior inference is carried out using MCMC and reversible jumps methods when needed.

Kolyan Ray, (Leiden University)

A Bayesian approach to log-concave density estimation

Log-concave densities have numerous applications, including in estimating mixture models, tail index estimation, clustering, regression and independent component analysis. We propose a Bayesian method to estimate a log-concave density using a prior based on an exponentiated Dirichlet process mixture. We show that the posterior distribution converges to the log-concave truth at the minimax rate in Hellinger distance. We also propose a computationally more feasible relaxation that can be sampled using an MCMC scheme and illustrate our approach in simulations.

Yordan Raykov, (Aston University)

Efficient block inference for BNP models using Beta distributed slices

We extend the current slice sampling methods for DPMMs and HDP-HMMs by modifying the corresponding probabilistic models with different auxiliary variables. Instead of assuming uniformly distributed auxiliary variables to obtain efficient 'slices', we study constructions which use more general beta distributed variables. Similar to earlier work, the proposed approach allows us to derive straightforward samplers without collapsing over the infinite random measures in BNP models. This enables scalable, memory efficient inference methods which use block updates for the model variables making most of the conditional updates inherently parallelizable. As the resulting samplers are tractable without any Rao-Blackwellization steps they are also really useful for inference non-conjugate BNP models. The auxiliary variable construction is further used to derive efficient and rigorous maximum-a-posteriori (MAP) algorithms for deterministic inference in DPMMs and HDP-HMMs. We demonstrate the scalability of the MAP method for inference in DPMMs on a large single cell gene expression data. When applied to HDP-HMMs, the proposed approach allows us to derive a nonparametric version of Viterbi algorithm which takes advantage of dynamic programming implementation. The method converges in just a few iterations to a locally optimal solution and can provide a useful modal approximation for low dimensional problems. We demonstrate a streaming application of the MAP algorithm where data is segmented online on board of a severely resource-constraint embedded hardware used for data collection of sensor data.

Alan Riva Palacio, (University of Kent)

Scalable Bayesian nonparametric inference for a class of survival regression models

In this work, we introduce a semi-parametric regression model for survival analysis data. The construction of the model relies on a vector of dependent completely random measures, allowing the borrowing of information among different covariates. The model proposed has non-proportional hazards and it ensures more flexibility in modelling survival data. We are able to characterize explicitly the posterior distribution allowing the implementation of the inferential procedure. In particular, when compound random measures are considered (Griffin and Leisen, 2017), we propose a novel scalable algorithm for addressing posterior inference.

Jenovah Rodrigues, (University of Edinburgh)

Bayesian Inverse Problems with Heterogeneous Noise

We review linear, ill-posed inverse problems in separable Hilbert spaces with noisy observations. A Bayesian solution with Gaussian regularising priors will be studied; the aim being to select the prior distribution in such a way that the solution achieves the optimal rate of convergence, when the unknown function belongs to a Sobolev space. Consequently, we will focus on obtaining the rate of convergence, for the rate of contraction, of the whole posterior distribution to the forementioned unknown function. Current results in the literature, ("Bayesian inverse problems with Gaussian priors" by Knapik et al), refer to a Gaussian white noise error model with homogeneous variance, which is investigated using the spectral decomposition of the operator defined in the inverse problem. We will extend these results to a heterogeneous Gaussian error model, in the hope of applying them to a Poisson error model at a future date.

Smit Rohan, (University of Pune)

Bivariate model for price change and duration
This paper studies the non-linear behavior of high-frequency financial data and employs bivariate model for price change and duration (PCD model). There has been a growing interest in such models in the recent past to study the market microstructure. The PCD model has been originally introduced by McCulloch and T say (2000). We have four simple conditional models to handle the dynamic structure within a trading day.We study the effect of the number of transactions without price change on the size of price change. Price reversal is also an another important factor in this analysis wherein we study if the behavior of the previous directions of the price change (whether the price goes up or comes down) has a role in explaining the price reversal for stocks from different sectors. The order of dependence also changes with sector and hence problem of order selection is considered. The Local trend in price movement has been studied by a general autoregressive structure. Instead of modelling durations by an ACD structure, we introduce additional explanatory variables in the model and investigate their significance. To estimate the parameters we have used Markov Chain Monte Carlo method for which we have used Gibbs sampling and Metropolis-Hasting Algorithm. We estimated this model using high-frequency data from NSE, India.

Luca Rossini, (Ca' Foscari University of Venice)

Bayesian nonparametric sparse seemingly unrelated regression model
Seemingly unrelated regression (SUR) models are used in studying the interactions among economic variables of interest. In a high dimensional setting and when applied to large panel of time series, these models have a large number of parameters to be estimated and suffer of inferential problems. In order to avoid overparametrization and overtting issues, we propose a hierarchical Dirichlet process prior for SUR models, which allows shrinkage of SUR coefficients toward multiple locations and identication of group of coefficients. This new multiple shrinkage prior model allows us to extract network structures from panel data and to cluster the network edges. In the macroeconomic application, we show the presence of contagion changes over lags and identify strong and weak contagion effects between European (core and periphery countries) and rest-of-the-world countries.

Massimiliano Russo, (Department of Statistical Sciences, University of Padua, Via C. Battisti 241, Padova, Italy)

Inferring structure in many way tables via grouped tensor factorizations
It is common in many fields to collect high-dimensional multivariate categorical data that can be organized as a many-way contingency table. The focus of inference in analyzing such tables is often the dependence structure, with log-linear and latent structure models providing popular approaches for statistical inference. In ultra-sparse cases in which the number of cells in the table vastly exceeds the sample size, such approaches tend to break down, and a variety of low rank tensor factorization algorithms have been proposed as alternatives. This article proposes an important new class of *GROuped Tensor (GROT)* factorizations, which have superior performance in terms of data compression, using relatively few components to represent the joint probability mass function of the data. While popular PARAFAC factorizations rely on mixing together independent components, GROT mixes together grouped

factorizations, equivalent to replacing vector arms in PARAFAC with low-dimensional tensor arms. Taking a Bayesian approach to inference with Dirichlet priors on the mixing weights and arm components, a combined low rank and sparse structure can be achieved, while facilitating efficient posterior computation via Markov chain Monte Carlo. The proposed GROT approach is shown to out perform a variety of existing approaches in simulations, and real data analysis.

Abhijoy Saha, (Department of Statistics, The Ohio State University)

A Geometric Variational Approach to Bayesian Inference

We propose a novel Riemannian geometric framework for variational inference in Bayesian models based on the nonparametric Fisher-Rao metric on the manifold of probability density functions. Under the square-root transform representation, the manifold with the Fisher-Rao metric reduces to the unit Hilbert hypersphere with the standard $\mathbb{L}^2$ metric. As the approximating family to the posterior distribution, we consider the family of all probability densities on the parameter space which factorize. We do not restrict our setup to just parametric families, as is common in the current literature. Further, in contrast to existing approaches based on the Kullback-Leibler divergence, we approximate the posterior by a member of an appropriate class closest to the posterior with respect to the α-divergence. As a consequence, in comparison with existing methods, our procedure leads to a tighter lower bound on the marginal density of the data. Our procedure also leads to an upper bound on the marginal density, which cannot be obtained from approaches based on Kullback-Leibler divergence. To solve the variational Bayes problem, we define a gradient ascent algorithm utilizing the geometry of the space of probability density functions. We provide several examples that validate the proposed framework. In particular, we consider classification via Bayesian logistic regression on few data sets and show that the performance of our method is comparable to other classification approaches.

S. Yaser Samadi, (Southern Illinois University, Carbondale)

Nonparametric Bayesian Factor Model for Tensor Time Series Data

Many data sets from across the sciences collect sequences of matrix- and tensor-structured data; we refer to such data as tensor time series. We are particularly motivated by electrophysiology studies in which electrical activity at multiple locations in the brain is measured over time. It is typical to pre-process such data to obtain tensors for each short time interval representing the level of coherence between each pair of brain regions at each spectral frequency. There is a pressing need in this application area to obtain flexible stochastic models of dynamic changes in neural subnetworks over time, with such models serving as an important starting point for analyses incorporating variation across study subjects and association with traits of the subjects and behavior. We propose a flexible class of nonparametric Bayesian factor models for tensor time series data, which reduce dimensionality and maintain interpretability through the incorporation of sparsity constraints. In particular, each dynamically varying factor only impacts a subset of the pairs of brain regions and frequency bands, and hence defines a subnetwork. Starting with a stochastic differential equation (SDE) representation, we define an efficient Markov chain Monte Carlo algorithm for posterior computation using a simulation smoother. The ability to accurately infer dynamically changing subnetworks is shown through simulations, and the methods are applied to mouse electrophysiology data.

Terrance Savitsky, (U.S. Bureau of Labor Statistics)

Bayesian Nonparametric Multiresolution Estimation for the American Community Survey

Bayesian hierarchical methods implemented for small area estimation focus on reducing the noise variation in published government official statistics by borrowing information among dependent response values. Even the most flexible models confine parameters defined at the finest scale to link to each data observation in a one-to-one construction. We propose a Bayesian multiresolution formulation that utilizes an ensemble of observations at a variety of coarse scales in space and time to additively nest parameters we define at a finer scale, which serve as our focus for estimation. Our construction is motivated by and applied to the estimation of $1-$ year period employment totals, indexed by county, from statistics published at coarser areal domains and multi-year periods in the American Community Survey (ACS). We construct a nonparametric mixture of Gaussian processes as the prior on a set of regression coefficients of county-indexed latent functions over multiple survey years. We evaluate a modified Dirichlet process prior that incorporates county-year predictors as the mixing measure. Each county-year parameter of a latent function is estimated from multiple coarse scale observations in space and time to which it links. The multiresolution formulation is evaluated on synthetic data and applied to the ACS.

Ingmar Schuster, (FU Berlin)

In recent years, there has been a high interest in Bayesian methods for high dimensional models. In particular, sparse Gaussian sequences have been well studied in the literature. In a recent paper, Tansey et al. (2014) proposed a new approach derived from the Spike and Slab for models with a known underlying structure, borrowing ideas from the Fused Lasso. In this work we propose a new approach in this setting, using shrinkage prior that has been proved to have good theoretical properties for the sequence model (Castillo et al. 2015). For this new model, we propose two numerical approaches to detect signal, one based on a Taylor approximation of the high dimensional posterior density in a transformed space and another based on an efficient sample using adaptive MCMC. On both simulated and real data experiments, our approach appears to be more flexible and competitive with the ones proposed in the literature.

Catia Scricciolo, (Department of Economics, University of Verona)

Nonparametric Bayes and maximum likelihood estimation in measurement error models
The problem of recovering the latent distribution from observations additively contaminated with errors having a Laplace distribution in a standard measurement error model is considered. Assuming the mixing distribution to be completely unknown leads to a nonparametric deconvolution problem. We study the behaviour of the Bayes' and maximum likelihood estimators to answer the question of whether there are functionals of the mixing distribution for which asymptotically efficient estimation can be performed. First, global rates of convergence in Hellinger distance for estimating the sampling density, which is a Laplace mixture with a possibly unbounded set of locations, are assessed: the rate of convergence for the Bayes' estimator corresponding to a Ferguson-Dirichlet process prior over the space of mixing distributions matches, up to a logarithmic factor, with the $n^{-3/8}\log^{1/8} n$ rate for the MLE. This global rate is then used to establish asymptotic efficiency of certain linear functionals of the MLE, which can be $\sqrt{n}$-consistent, asymptotically normal and efficient, even though the unknown mixing distribution can, at worst, only be recovered at a slower $n^{-1/8}$-rate in the L^1-Wasserstein distance. As for the Bayesian counterpart, linear functionals can be estimated at an almost parametric rate $n^{-1/2}K_n$, with K_n an arbitrarily slowly diverging sequence, when the mixing distribution is compactly supported on an unknown number of atoms, which is bounded by a finite known N, and a truncated stick-breaking prior is considered on the space of mixing distributions with at most N support points. The latent mixing distribution can instead be recovered at a nearly optimal rate of order $O_P(n^{-1/4}K_n^{1/2})$ in the L^1-Wasserstein or Kolmogorov-Smirnov distance.

Paulo Serra, (Eindhoven University of Technology)

Regression with correlated noise: non-parametric approach

Regression models, particularly of the "signal+noise" variant, play a central role in statis- tics and are a fundamental tool in many applied fields. Typically, the noise terms are as- sumed to be independent but this is often not a realistic assumption. Methods for selecting bandwidths/smoothing parameters for kernel/spline estimators like generalised cross-validation (GCV) break down even if the correlation is mild. To deal with this, two common approaches are to either "robustify" the criteria for selecting bandwidth/smoothing parameters, or making a parametric assumption on the noise. Unfortunately, both approaches are very sensitive to misspecification. The approach I will talk about is fully non-parametric. In this talk I will focus on penalised spline estimators – essentially smoothing splines with relatively few knots. I will show how they can be interpreted as Bayesian estimators (correspond- ing to a certain prior on the regression function). An alternative interpretation is as best linear unbiased predictors (BLUPs) in a linear mixed-effects model (LMM). The spline parameters are estimated via the empirical Bayes approach. I will talk a bit about some implementation issues, and about the asymptotics of the estimators. These asymptotics make explicit the influence of the correlation structure on the smoothing parameters of the penalised spline, and introduce some non-trivial constraints on the order of the splines. I will close with some numerical exper- iments where I compare our approach to two kernel estimators, and to a standard R procedure based on a (parametric) assumption on the noise structure.

Siva Sivaganesan, (University of Cincinnati)

Bayesian Causal Inference Using Gaussian Process
Bayesian approach to causal inference has traditionally been modeling the outcome mechanism and ignoring the fact that treatments are selectively assigned. On the other hand, frequentist approach has been relying on removing treatment selection bias via baseline covariates matching or propensity score methods, both of which are two-step methods. This article proposes a Bayesian semiparametric approach for estimating the population mean treatment effect using Gaussian process, which accomplishes matching and modeling outcome mechanism in a single step. We demonstrate a close relationship between the matching method and the GP regression for estimating average treatment effect. The proposed method utilizes a distance similar to Mahalanobis distance but determines the range of matching automatically without imposing a caliper arbitrarily. The results from our simulation study suggest that GP regression leads to an accurate and more efficient estimate than the linear regression modeling with adjustment for propensity score, inverse probability weighting and Bayesian additive regression trees. We also illustrate the proposed method using data obtained from the Pediatric Rheumatology Collaborative Outcome Improvement Network for estimating the effect of early aggressive biological disease modifying anti-rheumatic drugs.

Sanvesh Srivastava, (The University of Iowa)

Bayesian Causal Inference Using Gaussian Process
Bayesian approach to causal inference has traditionally been modeling the outcome mechanism and ignoring the fact that treatments are selectively assigned. On the other hand, frequentist approach has been relying on removing treatment selection bias via baseline covariates matching or propensity score methods, both of which are two-step methods. This article proposes a Bayesian semiparametric approach for estimating the population mean treatment effect using Gaussian process, which accomplishes matching and modeling outcome mechanism in a single step. We demonstrate a close relationship between the matching method and the GP regression for estimating average treatment effect. The proposed method utilizes a distance similar to Mahalanobis distance but determines the range of matching automatically without imposing a caliper arbitrarily. The results from our simulation study suggest that GP regression leads to an accurate and more efficient estimate than the linear regression modeling with adjustment for propensity score, inverse probability weighting and Bayesian additive regression trees. We also illustrate the proposed method

using data obtained from the Pediatric Rheumatology Collaborative Outcome Improvement Network for estimating the effect of early aggressive biological disease modifying anti-rheumatic drugs.

Rebecca Steorts, (Duke University)

Methods for Quantifying Conflict Casualties in Syria
Information about social entities is often spread across multiple large databases, each degraded by noise, and without unique identifiers shared across databases. Record linkage—reconstructing the actual entities and their attributes—is essential to using big data and is challenging not only for inference but also for computation. In this talk, I motivate record linkage by the current conflict in Syria. It has been tremendously well documented, however, we still do not know how many people have been killed from conflict-related violence. We present a novel approach that uses machine learning techniques and a novel Bayesian nonparametric method and property allowing us to estimate death counts in Syria. We crucially rely on the microclustering property, whereby the size of the largest cluster grows sublinearly as the number of records grows and propose a class of new models for record linkage satisfying this property. Our model quantifies the uncertainty in the inference and propagates this uncertainty into subsequent analyses. Finally, we provide initial death estimates for the Syrian conflict, comparing these to supervised methods. We remark to the success and challenges of solving a problem that is at the forefront of national headlines and news.

Briana Stephenson, (University of North Carolina at Chapel Hill)

Robust model-based clustering for multivariate and grouped data via local deviation processes
Model-based clustering has become a great utility in modeling systematic differences across subgroups within a large population. Traditional clustering methods operate under the assumption that two subjects allocated to the same cluster will respond identically to all measured variables. Yet, aberrations between some individual measured variables can yield valuable information. Furthermore, these models often realize an increasing number of clusters that expand with the dimensionality of sample size (n) and number of variables (p). This may lead to a loss in interpretability, due to a large number of clusters and an oversensitivity to minor deviations that exist among groups. Our goal is to develop a robust and parsimonious clustering method to address these complexities. Motivated from a partial partition process framework, we propose a new method that allows subjects to aggregate at two levels: (1) globally, where subjects are allotted to a single cluster via a standard Dirichlet process and (2) locally, where individual measured items can deviate from their global indicators via a Beta-Bernoulli process to adapt for differences across groups of individuals. Using data obtained from the 1997-2009 National Birth Defects Prevention Study, we apply this method for the derivation of dietary patterns of pregnant women in the United States, while adjusting for potential state-level differences.

Justin Strait, (Department of Statistics, The Ohio State University)

Bayesian Models for Automatic Landmark Detection on Elastic Curves
Given a collection of shapes from a population, one may wish to identify important points (landmarks) on these shapes which are preserved across the population. Landmarks are frequently used in statistical shape analysis, both in early and recent literature. The goal of this work is to provide an automated, model-based approach to detecting landmarks on shapes (represented as either open or closed curves). The model is based on a linear reconstruction of the collection of shapes, passing through the specified points. Inference is done in a Bayesian setting in order to efficiently estimate posterior distributions and quantify uncertainty in the landmark locations. The question of how many landmarks to select is also addressed in two different ways: a criteria-based approach, as well as incorporation into the Bayesian model (leading to a variable dimension parameter space). Efficient methods for posterior sampling are also discussed.

Martin Trapp, (Austrian Research Institute for Artificial Intelligence (OFAI) and Graz University of Technology)

infinite Sum-Product Networks

Sum-Product Networks (SPNs) are tractable probabilistic models which neatly combine the fields of deep learning and probabilistic modelling. While guaranteeing linear time inference in their representation size, these models have shown impressive results on a broad variety of applications ranging from image completion to language modelling. Despite the promising results, learning in SPNs is difficult as defining an appropriate network structure is crucial. In this work, we derive a prior on so-called induced trees which build the basis of every valid SPN. This allows us to infer network structures and provides interpretability of the latent variables of such networks. Interleaving the nested Dirichlet process with nested variable subset selection enables us to specify the generative process for infinite sum-product trees (iSPTs). Our construction guarantees for consistent and decomposable SPNs warranting the tractability of such networks. We show how to utilize the tractability of SPNs to efficiently sample from the posterior of iSPTs by deriving a distributed slice sampling approach. Furthermore, we discuss the inference in general trees and show how the posterior of infinite sum-product networks (iSPN) can be derived using our prior construction. To evaluate the effectiveness of iSPTs, we give quantitative comparisons on various datasets.

Laura Turbatu, (University of Geneva)

Accuracy and validity of posterior quantiles in Bayesian inference using empirical likelihoods

In this work we investigate the validity and the accuracy properties of posterior quantiles in Bayesian statistics when using empirical likelihoods based on a set of estimating equations. At first we study the validity of the posterior distributions derived using a continuous prior distribution and a pseudo-likelihood based on a set of M-type estimating equations of location constructed with weights derived as minimum distant from the empirical distribution in the Cressie-Read family of divergences. We subsequently compare in terms of accuracy different members of the Cressie-Read family for the location model in the exponential family, notably Owen's empirical likelihood (EL) and the exponential tilting empirical likelihood (ET). In the frequen- tist setup, the EL is Bartlett correctable while the ET is not, and we translate this relation in the Bayesian framework by comparing the accuracy of the posterior quantiles when we are using each of these empirical likelihoods. In the final part we propose an empirical noninformative prior distribution by matching the posterior coverage probability to the corresponding frequentist empirical coverage probability, such that we address the Bayesian inference from a complete empirical based perspective.

Willem van den Boom, (Duke University)

Flexible Bayesian Feature Extraction from Varying Length Functional Data

Although there is a rich literature on methods for assessing the impact of functional predictors, the focus has been on approaches for dimension reduction that can fail dramatically in certain applications, such as when the size of the domain varies between values for the functional predictor. Examples of standard approaches include functional linear models, functional principal components, and cluster-based approaches, such as latent trajectory analysis. We propose a Bayesian functional feature extraction model motivated by blood pressure trajectories during surgery as predictors for post-surgery adverse outcomes. Our model allows for the varying length of the blood pressure trajectories due to varying surgery duration. The functional predictor is modelled by a Gaussian mixture autoregressive model that allows for rich functional dynamics while its flexible stationary distribution is available analytically. Features that can be extracted from functions can be related to dynamics as wells as features of the stationary distribution.

Johan Van Der Molen Moris, (University of Edinburgh)

Bayesian semi-parametric density estimation for nonregular models

We consider a Bayesian semi-parametric model for estimating a nonregular parameter such as the end point of the support of an unknown density with a discontinuity. For this purpose we consider two types of prior models for the density, namely, a sieve prior for a basis expansion and a Dirichlet Process mixture prior. The aims are to select appropriate hyperprior distributions so that the estimators perform well in practice, and to investigate their theoretical performance, more specifically a BvM-type theorem. We will present our preliminary results on both prior models for a monotone non-increasing density with no interaction with the location parameter with illustration on simulated and real data. This is joint work with Natalia Bochkina, Judith Rousseau and J. B. Salomond.

Jan van Waaij, (Korteweg-De Vries Institute for Mathematics)

The Best Of Both Worlds: Adaptive Posterior Convergence Rates For Empirical Bayes Methods In Nonparametric Diffusion Models

Observe $X^T = \{X_t : t \in [0,T]\}, T > 0$ from an SDE $dX_t = \theta_0(X_t)dt + dW_t, \theta_0$ is a 1-periodic function. We endow this space with a prior Π_λ defined by $\theta = \lambda \sum_{k=1}^{\infty} k^{-1/2-\alpha} Z_k \phi_k$ with $\alpha > 0$ fixed and $\lambda > 0$. The Z_k are iid $N(0,1)$ and (ϕ_k) is the standard Fourier basis. For fixed λ, the asymptotic posterior convergence rates are suboptimal, unless the Sobolev smoothness β of the unknown true parameter matches the smoothness α of the prior (see [1,4]). Scaling with a time-dependent λ changes the apparent smoothness of the prior. We estimate an optimal λ from the data and use the posterior of the prior $\Pi_{\hat{\lambda}_n}$ for the inference. Define the Marginal Maximum Likelihood Estimator (MMLE) $\hat{\lambda}_n$ as a value $\lambda \in \Lambda := [n^{-1/(4+4\alpha)}, n^\alpha]$ that maximises $\int p\theta(XT) d\Pi_\lambda(\theta)$ over Λ. Intuitively $\Pi_{\hat{\lambda}_n}$ is the prior that favours the true parameter the most, by putting the most mass around θ_0 of all priors $\{\Pi_\lambda : \lambda \in \Lambda\}$. We follow the approach of [2,3] and show that when $\beta \in (0, \alpha + 1/2]$, the posterior $\Pi_{\hat{\lambda}_n}(.|X^T)$ concentrates eventually almost all its mass in $MT^{-\beta/(1+2\beta)}$-L^2-balls around θ_0, with probability going to 1, where $M > 0$ is a big constant. This rate is minimax optimal, up to a constant. [1] Castillo, I. (2008) Lower bounds for posterior rates with Gaussian process priors [2] Donnet, S., Rivoirard, V., Rousseau, J., Scricciolo, C. (2016) Posterior concentration rates for empirical Bayes procedures, with applications to Dirichlet process mixtures [3] Rousseau, J., Szabo, B. (2017) Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator [4] van der Vaart, A. W., van Zanten, J. H. (2008) Rates of contraction of posterior distributions based on Gaussian process priors

Elodie Vernet, (University of Cambridge)

Posterior concentration rates for emission distributions in finite nonparametric mixture models and hidden Markov models

Finite hidden Markov models (HMMs) and mixture models have been widely used in diverse fields such as speech recognition, genomics, econometrics. Because parametric modeling of emission distributions, that is the distributions of an observation given the latent state, may lead to poor results in practice, in particular for clustering purposes, recent interest in using nonparametric latent models appeared in applications. Yet little thoughts have been given to theory in this framework. Here we are interested in the quality of estimation of the emission distributions. We consider finite state space mixture models and hidden Markov models and we assume we know this finite number of possible states; so that the distribution of the latent states (which are i.i.d. categorical in mixture models and form a Markov chain in HMMs) is described by a parameter in a finite-dimensional space. Yet the emission distributions (the distributions of an observation given the associated latent state) are not constrained to live in a finite-dimensional space. A common difficulty in HMMs and mixture models is to understand how the marginal distribution of the observations is linked to the parameters (in particular to the emission distributions). Recently this understanding increased thanks to general identifiability results. Yet more precise results are needed when we want to derive rates for the parameters. We will present a control of the (L1-)distance between emission distributions associated with close marginal distribution of the observations. This control allows us to derive posterior concentration rates for emission distributions.

Anqi Wu, (Princeton Neuroscience Institute)

Brain Kernel: A covariance function for fMRI data using a large-scale Gaussian process latent variable model

The study of correlated neural activity in the brain represents an easily accessible marker of brain functional architecture and has given rise to fundamental insights on brain organization. In this work, we describe a novel covariance function for full-brain activity patterns, or "brain kernel", which we learn from large functional Magnetic Resonance Imaging (fMRI) datasets. The most common choice of covariance function for data that exhibits spatial smoothness like fMRI is the RBF kernel, which uses the coordinates of voxels in 3D Euclidean space to set the a priori correlation of the brain activity. However, Euclidean distance is a poor proxy for functional connectivity in real brains, and voxels that are far apart often exhibit strong correlations. To address this shortcoming, we use a Gaussian process (GP) to infer a continuous nonlinear mapping from 3D brain coordinates to a latent space so that the RBF kernel can accurately capture the covariance, and employ a GP latent variable model to account for the covariance of brain activity. Learning this model has two main challenges: i) the latent embedding must accurately capture the true full-brain correlation; ii) full-brain data is extremely high dimensional (containing 10-100K voxels). To overcome these two challenges, we use a block coordinate descent algorithm for exact and scalable inference, and provide a scalable spectral formulation for the hyper-parameter estimation for the GP kernel. We show that the model can learn a good latent embedding representation to construct a brain kernel function which can generalize better to unseen data than alternative methods for high-dimensional covariance estimation. To our knowledge, this is the first report of a covariance function of large scale spontaneous brain activity.

Jieren Xu, (Duke University)

General Mode Decompositions and Spectral Analysis Using a Mechanistic Hierarchical Based Deep Gaussian process

General mode decompositions and spectral analysis for multiple components general-mode time series $sum_i s_i(\pi\phi_i(t))$, especially retrieving the time-variant frequency and phase functions $\phi_i(t)$ for each components separately, are fundamental problems in harmonic analysis and signal processing. This paper provides a novel mechanistic hierarchical based deep Gaussian process(GP) approach to extract different time series components $s_i(\pi\phi_i(t))$ from their superposition input, and estimate the spectral information as well as the general mode function $s_i(t)$ for each components. There are two layers of hierarchical GP structure in our model. The first layer is a deep Gaussian process(Damianou et al.) with two unknown Gaussian process, the mode-function $s_i(t)$ and the phase $\phi_i(t)$, composed together and a horizontal components structure to be estimated. They second layer is a mechanistic hierarchical Gaussian process model(Dunson et al.) with an estimated first order derivative information of $\phi_i(t)$, i.e. the frequency function, which is obtained by a newly developed Harmonic analysis method called ConCeFT(Daubechies et al.). The first layer of the problem is expensive, since deep GP is difficult to be scaled. Methods include Nested variational compression in deep Gaussian process(Hensman et al.) and the Variational auto-encoded deep GP(Dai et al.) are developed to conquer this defect. However, since we apply the mechanical hierarchical model to adapt local smoothness information for the second inner layer deep GP and the first inner layer deep GP, the mode-function $s_i(t)$, is periodic, the original deep GP becomes more "trackable". Our hierarchical model leads to a framework for the generalized mode decomposition problem under a weak well-separation condition. Numerical examples of synthetic and real data are provided to demonstrate the fruitful applications of these methods.

Junyu Xuan, (University of Technology Sydney)

Cooperative Hierarchical Dirichlet Processes

The cooperative hierarchical structure is a common and significant data structure observed in, or adopted by, plenty of research areas, such as: text mining (author-paper-word), recommender systems (genre-movie-user), and multi-label classification (label-instance-feature). Renowned Bayesian approaches for hierarchical structure modeling are mostly based on topic models. However, these approaches suffer from a serious issue in that the number of hidden topics/factors needs to be fixed in advance which may lead to overfitting or underfitting. Although an elegant way to resolve this issue has already been developed - Bayesian nonparametric learning - existing work in this area cannot be used for cooperative hierarchical structure modeling, i.e., hierarchical Dirichlet process is only designed for non-cooperative hierarchical structure. In this paper, we propose a cooperative hierarchical Dirichlet process (CHDP) to fill this gap. Each node in a cooperative hierarchical structure is assigned a Dirichlet process to model the weights of nodes on the hidden factors, which require no prior knowledge of the factor number. Addition to Inheritance, a new measure operation – Cooperation - is proposed and merged into hierarchical Dirichlet process to capture the many-to-many relationships in the cooperative hierarchical structure. Two different types of Cooperation are defined: Superposition and Maximization. Furthermore, two constructive representations, i.e., stick-breaking and international restaurant process, are designed for CHDP to facilitate the model inference.

Juan Yan, (University of Manchester)

Parallel Bayesian computing in wind power forecasting based on SpiNNaker

With the increasing penetration of wind power, it is imperative to make accurate wind power forecasting for the power dispatch to guarantee the stability of power grid. The current forecasting techniques usually combines a network of different algorithms and becomes less-competitive in forecasting efficiency. The biology-inspired parallel computing machine SpiNNaker offers a good option in changing the current situation. However, SpiNNaker was designed to simulate neuromurphy models. It's validation and efficiency in non-neuron models remains to be examined. In this paper, a Bayesian-based parallel computing method will be proposed and implemented on SpiNNaker for wind power forecasting. The algorithm will be applied to real wind farm data for test of its effectiveness.The result will show the improvement in forecasting efficiency and also the validation of SpiNNaker for non-neuron models.

Yun Yang, (Florida State University)

Bayesian model selection consistency and oracle inequality with intractable marginal likelihood

In this presentation, we investigate large sample properties of model selection procedures in a general Bayesian framework when a closed form expression of the marginal likelihood function is not available or a local asymptotic quadratic approximation of the log-likelihood function does not exist. Under appropriate identifiability assumptions on the true model, we provide sufficient conditions for a Bayesian model selection procedure to be consistent and obey the Occam's razor phenomenon, i.e., the probability of selecting the “smallest" model that contains the truth tends to one as the sample size goes to infinity. In order to show that a Bayesian model selection procedure selects the smallest model containing the truth, we impose a prior anti-concentration condition, requiring the prior mass assigned by large models to a neighborhood of the truth to be sufficiently small. In a more general setting where the strong model identifiability assumption may not hold, we introduce the notion of local Bayesian complexity and develop oracle inequalities for Bayesian model selection procedures. Our Bayesian oracle inequality characterizes a trade-off between the approximation error and a Bayesian characterization of the local complexity of the model, illustrating the adaptive nature of averaging-based Bayesian procedures towards achieving an optimal rate of posterior convergence. Specific applications of the model selection theory are discussed in the context of high-dimensional nonparametric regression and density regression where the regression function or the conditional density is assumed to depend on a fixed subset of predictors.

Mingan Yang, (San Diego State University)

Bayesian variable selection in semiparametric mixed effects models
In this article, we addres the problem of variable selection for both fixed and random effects in the mixed effects models. Generally, it is assumed that the fixed effects and random effects follow a parametric distribution, e.g, a normal distribution. However, substantial violation of the normality assumption can potentially impact the subset selection and result in poor interpretation and even incorrect results. In this article, we use the semiparatric priors for both the fixed and random effects and implement an efficient fully Bayesian variable selection. The approach is illustrated using a simulation and a real data example.

William Weimin Yoo, (Mathematical Institute, Leiden University)

Bayesian inference in infinite-dimensions: Adaptive procedures and their implications
Bayesian nonparametrics (BNP) studies Bayesian inference on nonparametric models, where the parameter of interest is infinite-dimensional. For example, we study posterior distributions of regression functions and densities. Applications of BNP are numerous, ranging from image analysis, neural networks, machine learning to climate modelling, genetics and speech recognition. Current research in BNP is driven by attempts to answer two fundamental questions in Bayesian analysis: `How do you put priors on infinite-dimensional parameters?" and`How to quantify uncertainty when estimating these quantities¿‘. My research involves developing a solid theoretical understanding of these two issues, and to formulate priors that adapt to the smoothness of the parameters to be estimated. In this talk, I will discuss wavelet spike-and-slab priors and a Bayesian Lepski's method to estimate regression functions. I will also talk about adaptive credible bands and how it can be used as confidence bands in practice. Other new topics that I will touch upon include priors for discrete wavelet transforms, self-similarity constraint using splines, and a lower limit of adaptation for Bayesian procedures.

Mahmoud Zarepour, (University of Ottawa)

A bayesian nonparametric goodness of fit test
The Bayesian nonparametric inference and Dirichlet process are popular tools in statistical methodologies. We employ the Dirichlet process in hypothesis testing to propose a Bayesian nonparametric chi-squared goodness-of-fit test. The null hypothesis of the classical Pearson's chi-squared goodness-of-fit test examines how well the distribution F_0 fits a set of observations. In our Bayesian nonparametric approach, we consider the Dirichlet process as the prior for the distribution of data and carry out the test based on the Kullback-Leibler distance between the updated Dirichlet process and the hypothesized distribution F_0. We prove that this distance asymptotically converges to the same chi-squared distribution as the classical test does. Similarly, a Bayesian nonparametric chi-squared test of independence for a contingency table is provided. Also by computing the Kullback-Leibler distance between the Dirichlet process and the hypothesized distribution, a method to obtain an appropriate concentration parameter for the Dirichlet process is proposed.

Mahmoud Zarepour, (University of Ottawa)

A consistent Bayesian bootstrap for chi-squared goodness-of-fit test using a Dirichlet prior
The Bayesian nonparametric inference and Dirichlet process are popular tools in Bayesian statistical methodologies. In this paper, we employ the Dirichlet process in a hypothesis testing to propose a Bayesian nonparametric chi-squared goodness-of-fit test. In our new Bayesian nonparametric approach, we consider the Dirichlet process as the prior for the distribution of the data and carry out the test based on the Kullback-Leibler distance between the updated Dirichlet process and the hypothesized distribution. We prove that this distance asymptotically converges to the same chi-squared distribution as the classical frequentist's chi-squared test does. Indeed, the result indicates that the Bayesian bootstrap is consistent for this goodness-of-fit chi-squared test. Moreover, the results are generalized to chi-squared test of independence for a contingency table. In addition, a method to obtain an appropriate concentration parameter for the Dirichlet process is presented.

Michael Zhang, (The University of Texas at Austin)

Embarrassingly parallel inference for Gaussian processes
Training Gaussian process (GP) based models typically involves an $O(N^3)$ computational bottleneck. Popular methods for overcoming the matrix inversion problem include sparse approximations of the covariance matrix through inducing variables or through dimensionality reduction via "local experts". However, these type of models cannot account for both long and short range correlations in the GP functions. Furthermore, these methods are often ill-suited for cases where the input data is not uniformly distributed. We present an embarassingly parallel method that takes advantage of the computational ease of inverting block diagonal matrices, while maintaining much of the expressivity of a full covariance matrix. By using importance sampling to average over different realizations of low-rank approximations of the GP model, we ensure our algorithm is both asymptotically unbiased and embarrassingly parallel.

Mingyuan Zhou, (University of Texas at Austin)

Softplus Regressions and Convex Polytopes
To construct flexible nonlinear predictive distributions, I introduce a family of softplus function based regression models that convolve, stack, or combine both operations by convolving countably infinite stacked gamma distributions, whose scales depend on the covariates. Generalizing logistic regression that uses a single hyperplane to partition the covariate space into two halves, softplus regressions employ multiple hyperplanes to construct a confined space, related to a single convex polytope defined by the intersection of multiple half-spaces or a union of multiple convex polytopes, to separate one class from the other. The gamma process is introduced to support the convolution of countably infinite (stacked) covariate-dependent gamma distributions. For Bayesian inference, Gibbs sampling derived via novel data augmentation and marginalization techniques is used to deconvolve and/or demix the highly complex nonlinear predictive distribution. Example results demonstrate that softplus regressions provide flexible nonlinear decision boundaries, achieving classification accuracies comparable to that of kernel support vector machine while requiring significant less computation for out-of-sample prediction.

Shuang Zhou, (Florida State University)

Gaussian process with error in covariates
Gaussian process is a very popular tool for nonparametric regression. However, in many applications, the covariates are not directly observed and are often contaminated with error. This is common in spatial applications where the locations are observed with error. There are two key challenges associated with such models. First, the posterior computation

involves sampling the unobserved locations which is computationally expensive and even prohibitive when the sample size is moderate to high. We propose a random-cosine approximation to the Gaussian process which allows fast and numerically stable posterior updates. Second, we obtain optimal posterior convergence rate for estimating the true regression function under different choices of the contaminated error density. Simulation examples demonstrate the efficacy of the approach and the methods are applied to estimate global world temperature in presence of location errors.

List of Participants