

Abstracts

Contents

Federico Camerlenghi <i>University of Milano Bicocca, Italy</i>	2
Claudio Del Sole <i>Bocconi University, Italy</i>	3
Edwin Fong <i>University of Hong Kong</i>	4
Beatrice Franzolini <i>Bocconi University, Italy</i>	5
Subhashis Ghoshal <i>North Carolina State University, USA</i>	6
Maria Fernanda Gil Leyva <i>IIMAS-UNAM, Mexico</i>	7
Li Ma <i>Duke University, USA</i>	8
Long Nguyen <i>University of Michigan, USA</i>	9
Giovanni Rebaudo <i>University of Torino, Italy</i>	10
Yan Shuo Tan <i>National University of Singapore, Singapore</i>	11

Federico Camerlenghi
University of Milano Bicocca, Italy

*Predictive characterizations of feature allocation models: general theory
and related applications*

Selecting a prior distribution is a fundamental problem of Bayesian inference, as well as one of the main critiques to the Bayesian approach by other statisticians. Recent contributions proposed to sidestep prior selection by using a “predictive approach”, whereby the statistician needs to assign a predictive rule for a new observation. In the context of species sampling methods, the interplay between the prior choice and predictive approaches is well understood in terms of W.E. Johnson “sufficientness” postulates. We extend this characterization to feature sampling models, where each observation belongs to different groups, characterizing those priors for which the probability of discovery of new features depends solely on the sample size and on the combination of sample size and total number of observed features. Palm calculus and point process theory are fundamental ingredients to obtain predictive characterizations. We finally examine several examples of feature sampling models, and we discuss significant applications in ecology and genomics.

[Back to Contents Page →](#)

Claudio Del Sole
Bocconi University, Italy

A random partition model with dependence on covariates

In the Bayesian nonparametric framework, the inclusion of continuous covariates within a fully nonparametric regression model represents a challenging task, and existing models in the literature face a trade-off between flexibility in modelling the latent partition structure, its analytical tractability, and its consistency for new observations. We introduce a novel class of covariate-dependent random probability measures, arising from the normalization of suitable random measures, which account for covariates

through a kernel structure: specifically, the jumps of a common discrete random measure are rescaled via multiplication by a similarity kernel. This construction induces a random partition model with dependence on covariates, which is characterized by great flexibility and inherent consistency for new observations, while retaining some analytical tractability, thanks to the introduction of suitable latent variables. A noteworthy example arises when the distribution of the common random measure is a specific transformation of the distribution of a stable process. Moreover, the structure of the posterior distribution implied by such specification suggests the introduction of a novel nonhomogeneous process, which extends the two-parameter Poisson-Dirichlet process and acts as quasi-conjugate prior for our proposal.

[Back to Contents Page →](#)

Edwin Fong
University of Hong Kong

Bayesian Nonparametric Inference with the Martingale Posterior

While the prior distribution is the usual starting point for Bayesian uncertainty, recent work has reframed Bayesian inference as the predictive imputation of missing observations. In particular, the martingale posterior distribution arises when the Bayesian model is a chosen sequence of predictive distributions on future observables, which then induces a posterior distribution on the parameter of interest without the need for a likelihood and prior. This generalization greatly broadens the range of nonparametric models one can use for Bayesian inference, and offers substantial advantages in computation and flexibility, while posing new challenges. In this talk, we introduce the framework and present some recent advances.

[Back to Contents Page →](#)

Beatrice Franzolini
Bocconi University, Italy

*Multivariate species sampling models for partially exchangeable data
and beyond*

Species sampling models provide a structural framework for understanding random discrete distributions in the context of exchangeable observations. However, they do not cover the case of heterogeneous data collected from related sources or distinct experimental conditions. To address this limitation, we introduce multivariate species sampling models, a general class of models characterized by their partially exchangeable partition probability function. They encompass most existing Bayesian nonparametric models for partially exchangeable data, and allow to highlight their core distributional properties and induced learning mechanisms.

We show that the dependence structure is determined by the probability of recording ties across groups with independence across sources corresponding to it being zero. Furthermore, multivariate species sampling models admit three equivalent representations in terms of either the vector of random probability measures or the collection of partially exchangeable random partitions or the sequence of observations: we provide three characterization theorems for the laws governing these objects.

The presented results enable an in-depth comprehension of the induced dependence structure of existing models and facilitate the development of new models for both partially exchangeable data and general complex data structures. By building upon this foundation, challenges posed by dynamic clustering and structured network data, among others, can be effectively addressed. The practical impact of the theoretical and modeling advances is showcased through various applications, ranging from analysis of metabolic profiles to community detection.

[Back to Contents Page →](#)

Subhashis Ghoshal
North Carolina State University, USA

*Immersion posterior: Giving Bayes an extra degree of freedom to meet
frequentist goals under structural restrictions*

A structural relation in a statistical model is often necessary for a proper scientific interpretation and includes adds valuable information. However, incorporating that in Bayesian analysis typically makes the problem considerably harder for prior construction, posterior computation, and most importantly, for studying delicate frequentist properties like coverage of credible regions. We explore an extension of the Bayesian paradigm where a structural restriction is induced by a structure complying correction map from an unrestricted posterior distribution where the later is easily computed, and can be theoretically analyzed. The resulting random measure, to be called an "immersion posterior", offers simple and effective solutions to many inference problems with structural restriction. A natural example of the correction map is often given by a projection to the restricted parameter space from a simpler unrestricted space. We give interpretations of the immersion posterior and discuss its "extent of Bayesianity" in comparison to other extensions of the Bayesian paradigm such as objective Bayesian analysis, empirical Bayes, misspecified likelihood-based posterior (such as tempered likelihood, Gibbs posterior and PAC-Bayes), and variational Bayes. The immersion posterior approach can also be combined with other extensions of the Bayesian approach. We present results for several problems including differential equation models, models with shape restriction such as monotonicity, convexity, multivariate monotonicity, sparse regression, graphical models, and others.

[Back to Contents Page →](#)

Maria Fernanda Gil Leyva

IIMAS-UNAM, Mexico

Markov stick-breaking processes

Stick-breaking has a long history and represents the most popular procedure for constructing random discrete distributions in Statistics and Machine Learning. In particular, due to their intuitive construction and computational tractability they are ubiquitous in modern Bayesian nonparametric inference. Most widely used models, such as the Dirichlet and the Pitman-Yor processes, rely on iid or independent length variables. Here we pursue a completely unexplored research direction by considering Markov length variables and investigate the corresponding general class of stick-breaking processes, which we term Markov stick-breaking processes. We derive conditions ensuring that the associated species sampling process is proper and that the distribution of a Markov stick-breaking process has full topological support, two fundamental desiderata for Bayesian nonparametric models. We also analyze the stochastic ordering of the weights and provide a new characterization of the Pitman-Yor process as the only stick-breaking process invariant under size-biased permutations, under mild conditions. Moreover, we identify two notable subclasses of Markov stick-breaking processes, that enjoy appealing properties and include Dirichlet, Pitman-Yor and Geometric priors as special cases. Our findings further include distributional results on the latent allocation variables and a posterior characterization of the length variables: besides their theoretical interest, these allow us to develop computational algorithms for posterior inference. Interesting methodological implications are drawn from numerical implementations of Markov stick-breaking mixture models.

[Back to Contents Page →](#)

Li Ma

Duke University, USA

A partial likelihood approach to Bayesian tree-based density modeling

Tree-based nonparametric processes for random probability measures are usually specified as priors on sampling models defined along a predetermined, data-independent recursive partition on the sample space. To characterize the structure of an unknown target distribution in detail, the partition must expand deeply down the tree. Such expansions often incur prohibitive computational costs and are extremely wasteful in regions where the distribution has little probability mass, which often occupy most of the sample space in multivariate applications. Consequently, existing tree-based Bayesian models typically must compromise and rely on relatively shallow trees to model the sampling distribution, forcing practitioners to give up the ability to learn highly local distributional structures and instead resort to prior regularization to smooth these structures out. This renders useless one of the most desirable features of trees—their ability to characterize local features, and consequently often leads to a loss of statistical efficiency in estimation and a decrease in power in hypothesis testing. Traditional wisdom has long held that this compromise in the depth of the tree is an inevitable price one must pay to ensure coherent Bayesian reasoning, as a data-dependent partition that expands deeper in high-density regions would induce double-dipping in the likelihood and thus lead to inconsistent inference. We show that there exists a simple strategy to restore coherency in inference while allowing the partition to be data-dependent, using Cox’s partial likelihood framework. This strategy enables the sampling model to adjust its complexity locally according to the probability density of the target distribution. Moreover, we demonstrate that under appropriate choices of the data-dependent partition strategy, the resulting partial-likelihood-based inference is asymptotically equivalent to full likelihood-based inference endowed with oracle knowledge regarding the quantiles of the unknown distribution of interest. An immediate implication is that all the theoretical guarantees in terms of posterior consistency and contraction rates for traditional tree-based models with fixed partitioning automatically extend to their partial likelihood counterparts. Our partial likelihood approach is broadly applicable to existing Bayesian nonparametric (BNP) priors on tree-based models. Through numerical experiments, we demonstrate substantial improvements in both estimation accuracy and computational efficiency of these models.

[Back to Contents Page →](#)

Long Nguyen
University of Michigan, USA

Dendrogram of mixing measures: Hierarchical clustering and model selection using finite mixture models

We present a new way to summarize and select mixture models via the hierarchical clustering tree (dendrogram) constructed from an overfitted latent mixing measure. Our proposed method bridges agglomerative hierarchical clustering and mixture modeling. The dendrogram's construction is derived from the theory of convergence of the mixing measures, and as a result, we can both consistently select the true number of mixing components and obtain the pointwise optimal convergence rate for parameter estimation from the tree, even when the model parameters are only weakly identifiable. In theory, it explicates the choice of the optimal number of clusters in hierarchical clustering. In practice, the dendrogram reveals more information on the hierarchy of subpopulations compared to traditional ways of summarizing mixture models. Several simulation studies are carried out to support our theory. We also illustrate the methodology with an application to single-cell RNA sequence analysis.

This work is joint with Dat Do, Linh Do, Scott McKinley and Jonathan Terhorst.

[Back to Contents Page →](#)

Giovanni Rebaudo
University of Torino, Italy

Graph-Aligned Random Partition Model (GARP)

Bayesian nonparametric mixtures and random partition models are powerful tools for probabilistic clustering. However, standard independent mixture models can be restrictive in some applications such as inference on cell lineage due to the biological relations of the clusters. The increasing availability of large genomic data requires new statistical tools to perform model-based clustering and infer the relationship between homogeneous subgroups of units. Motivated by single-cell RNA data we develop a novel dependent mixture model to jointly perform cluster analysis and align the clusters on a graph. Our flexible graph-aligned random partition model (GARP) exploits Gibbs-type priors as building blocks, allowing us to derive analytical results for the probability mass function (pmf) on the graph-aligned random partition. We derive a generalization of the Chinese restaurant process from the pmf and a related efficient and neat MCMC algorithm to implement Bayesian inference. We illustrate posterior inference under the GARP using single-cell RNA-seq data from mice stem cells. We further investigate the performance of the model in recovering the underlying clustering structure as well as the underlying graph by means of simulation studies.

This is a joint work with Peter Müller.

[Back to Contents Page →](#)

Yan Shuo Tan

National University of Singapore, Singapore

*The Computational Curse of Big Data for Bayesian Additive Regression
Trees: A Hitting Time Analysis*

Bayesian Additive Regression Trees (BART) is a popular Bayesian non-parametric regression model that is commonly used in causal inference and beyond. Its strong predictive performance is supported by theoretical guarantees that its posterior distribution concentrates around the true regression function at optimal rates under various data generative settings and for appropriate prior choices. In this talk, we show that the BART sampler often converges slowly, confirming empirical observations by other researchers. Assuming discrete covariates, we show that, while the BART posterior concentrates on a set comprising all optimal tree structures (smallest bias and complexity), the Markov chain's hitting time for this set increases with n (training sample size), under several common data generative settings. As n increases, the approximate BART posterior thus becomes increasingly different from the exact posterior (for the same number of MCMC samples), contrasting with earlier concentration results on the exact posterior. This contrast is highlighted by our simulations showing worsening frequentist undercoverage for approximate posterior intervals and a growing ratio between the MSE of the approximate posterior and that obtainable by artificially improving convergence via averaging multiple sampler chains. Finally, based on our theoretical insights, possibilities are discussed to improve the BART sampler convergence performance.

[Back to Contents Page →](#)