Guest Editors' Introduction to the Special Issue on Bayesian Nonparametrics

Ryan P. Adams, Emily B. Fox, Erik B. Sudderth, and Yee Whye Teh

BAYESIAN nonparametric models are probabilistic models defined over infinite-dimensional parameter spaces. For Gaussian process models of regression and classification functions, the parameter space consists of a set of continuous functions. For the Dirichlet process mixture models used in density estimation and clustering, the parameter space is dense in the space of probability measures. Bayesian nonparametric models provide a flexible framework for modeling complex data and a promising alternative to classical model selection methods. Due to recent computational advances, these approaches have received increasing attention in machine learning, statistics, probability, and related application domains.

As guest editors of this special issue on Bayesian nonparametrics (BNP), we are very happy to introduce 19 papers advancing the state-of-the-art in BNP theory and practice. Our initial request for white papers drew over 60 contributions, and we are thankful for the efforts of the outstanding reviewers who assessed and improved the many excellent submissions. The resulting papers combine nonparametric modeling advances with practical learning algorithms to analyze diverse datasets including text documents, social and biological networks, financial and genetic time series, natural images, and robotic control problems. Below we group these papers into four primary themes: (i) hierarchical and topic models; (ii) feature-based alternatives to clustering; (iii) sequential models of temporal dynamics; (iv) relational models of graphs, arrays, and networks.

Due to its analytic and computational tractability, the Dirichlet process (DP) mixture model has become one of the most widely used BNP models. While DP mixtures allocate an unbounded number of latent clusters for flexible data modeling, their assumption that finite subsets of those clusters have Dirichlet-distributed weights is not always appropriate. Looking beyond the DP, Pierpaolo De Blasi, Stefano Favaro, Antonio Lijoi, Ramses Mena, Igor Pruenster, and Matteo Ruggiero ask, "Are Gibbs-type Priors the Most Natural Generalization of the Dirichlet Process?" Via prediction tasks arising in species sampling problems, this paper situates the DP within the larger Gibbs-type family of stochastic processes, and demonstrates the potential advantages of alternative BNP priors on partitions.

- *R. Adams is with the School of Engineering and Applied Sciences, Harvard University, 33 Oxford St., Cambridge, MA 02138. E-mail: rpa@seas.harvard.edu.*
- *E.B. Fox is with the Department of Statistics, Box 354322, Padelford Hall, Room B-305, University of Washington, Seattle, WA 98195-4322. E-mail: ebfox@stat.washington.edu.*
- *E. Sudderth is with the Department of Computer Science, 115 Waterman Street, Brown University, Box 1910, Providence, RI 02912. E-mail: sudderth@cs.brown.edu.*
- *Yee Whye Teh is with the Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, United Kingdom. E-mail: y.w.teh@stats.ox.ac.uk.*

For information on obtaining reprints of this article, please send e-mail to: reprints.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2014.2380478

1 Hierarchical and Topic Models

Topic models capture dependencies among related groups of data via an admixture model, in which each group is modeled as a distinct mixture of a shared set of clusters. Changyou Chen, Wray Buntine, Nan Ding, Lexing Xie, and Lan Du propose a family of "Differential Topic Models" to capture similarities and differences among collections of text or image data. Their approach improves interpretability by using the Pitman-Yor process to capture not only similarities in word co-occurrence, but also differences between sections of the corpus.

The "Supervised Hierarchical Dirichlet Process", as explored by Andrew Dai and Amos Storkey, learns topics which are predictive of response labels associated with each training document. Their approach neatly combines a hierarchical Dirichlet process with a generalized linear model with sparse covariates.

John Paisley, Chong Wang, David Blei, and Michael Jordan propose a "Nested Hierarchical Dirichlet Process" which allows learning of topics with hierarchical structure. In their approach, topics are organized in a tree, and the semantic content of each document is associated with a subtree of the overall topic hierarchy.

David Knowles and Zoubin Ghahramani show that a closely-related nested partition leads to "Pitman Yor Diffusion Trees", which can be used for Bayesian hierarchical clustering. This work generalizes the Dirichlet diffusion tree of Neal (2003) to allow for flexible non-binary tree structures that can be used for clustering, while maintaining an exchangeable distribution over the data.

2 Feature-Based Alternatives to Clustering

Standard BNP priors on random partitions, including the DP and the broader family of Gibbs-type priors, associate each observation with a unique latent cluster. In many applications, however, it may be more appropriate to represent each observation with an a priori unknown set of latent features. The Indian buffet process (IBP) by Griffiths and Ghahramani (2011) was the first Bayesian nonparametric model for such featural representations of data. A number of papers in this special issue build upon this initial work.

In a pair of papers, one by Tamara Broderick, Lester Mackey, John Paisley, and Michael Jordan on "Combinatorial Clustering and the Beta Negative Binomial Process", and the other by Mingyuan Zhou and Lawrence Carin on "Negative Binomial Process Count and Mixture Modeling", the IBP is generalized by replacing the present/absent binary features of the IBP by count-valued features. Applications include topic models and image segmentation.

An alternative framework for flexibly controlling feature frequencies is provided by the "Latent IBP Compound Dirichlet Allocation" model of Cedric Archambeau, Balaji Lakshminarayanan, and Guillaume Bouchard. This framework uses the three-parameter IBP to construct a topic model that exhibits true sparsity in topic distributions, while also exhibiting power law behavior.

Samuel Gershman, Peter Frazier, and David Blei suggest a "Distance Dependent Infinite Latent Feature Model" for data whose feature allocations are related to temporal, spatial, or other dependencies. Their approach allows a flexible user-specified distance function to bias feature generation such that similar data points share more features. Experiments study the temporal progression of Alzheimer's disease using MRI data.

Latent feature models provide a framework for learning sparse dictionary-based decompositions of high-dimensional data, as illustrated in "A Bayesian Nonparametric Approach to Image Super-resolution" by Gungor Polatkan, Mingyuan Zhou, Lawrence Carin, David Blei, and Ingrid Daubechies. They first learn a sparse model of dependencies between high-resolution and low-resolution image patches, and then quantitatively and visually assess the model's accuracy at recovering image details from low-resolution images. Experiments also compare stochastic variational learning algorithms to a baseline Gibbs sampler.

3 Temporal Dynamics and Sequential Models

The basic exchangeability assumptions justifying classical clustering models are violated in many applications, for example involving temporal or spatial data, as illustrated in "A Survey of Non-Exchangeable Priors for Bayesian Nonparametric Models" by Nicholas Foti and Sinead Williamson. In this survey, the authors cast the numerous recent developments in dependent nonparametric modeling of covariate-dependent data within a coherent framework. Their framework identifies three primary mechanisms for introducing dependency, and provides a straightforward structure for comparing models and developing new ones.

Zhiguang Xu, Steven MacEachern, and Xinyi Xu show how copula models may be used for "Modeling Non-Gaussian Time Series with Nonparametric Bayesian Model". To form stationary processes with arbitrary margins, the authors begin with classical autoregressive time series models, and apply a copula transformation to construct marginals distributed according to a target Dirichlet process mixture model. This approach leads to model coherency across different time scales, and improvements in short- and long-range predictions are demonstrated in an analysis of stock index returns.

Collections of time series with heterogeneous behavior motivate methods for "Fast Nonparametric Clustering of Dtructured Time-Series" by James Hensman, Magnus Rattray, and Neil Lawrence. Their approach uses a Dirichlet process to cluster functional data, and a hierarchical Gaussian process to capture trends within each cluster. A collapsed variational inference algorithm allows efficient analysis of gene expression data.

Statistical models of temporal dynamics provide the foundation for "Bayesian Nonparametric Methods for Partially-Observable Reinforcement Learning" by Finale Doshi-Velez, David Pfau, Frank Wood, and Nicholas Roy. Their two proposed approaches build on hierarchical Dirichlet and Pitman-Yor processes to allow an agent to refine the complexity of its knowledge representation as it gains more experience. Extensive experiments on a set of benchmark datasets explore the trade-offs between various nonparametric and parametric models.

Marc Deisenroth, Dieter Fox, and Carl Rasmussen pursue related applications of "Gaussian Processes for Data-Efficient Learning in Robotics and Control". Their approach, PILCO, uses a model-based policy search method with a Gaussian process transition model to naturally incorporate notions of model uncertainty. Long-term predictions and policy evaluation are then based on deterministic approximate inference with analytic gradients for policy learning. The significant speedup in learning without demonstrations is explored in various robotics applications including the control of unicycles and low-cost manipulators.

Computational scalability is a recurring problem in practical applications of Gaussian process models. Elad Gilboa, Yunus Saatci, and John Cunningham discuss methods for better "Scaling Multidimensional Inference for Structured Gaussian Processes". They identify cases where the Gaussian process marginal likelihood can be decomposed in ways that reduce computation without compromising accuracy, generalizing previous approaches for scalar inputs to the multidimensional setting. Applications include spatio-temporal models of land surface temperatures.

4 Relational Models of Graphs, Arrays, and Networks

Classical Bayesian nonparametric models are motivated by de Finetti's Theorem, which shows that any set of observations which are unordered or exchangeable can be modeled via a latent random measure. In applications such as the analysis of social and biological networks, observations are not of individual entities, but of relationships between them. Aldous and Hoover developed a theory of relational exchangeability, and an introduction and review of this theory are provided by Peter Orbanz and Daniel Roy in "Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures". This provides a solid foundation for developing novel Bayesian nonparametric relational models.

Konstantina Palla, David Knowles, and Zoubin Ghahramani discuss "Relational Learning and Network Modelling

Using Infinite Latent Attribute Models", a network model which represents each entity with an IBP-based feature model with an independent "Chinese restaurant process" for each feature. This increases the flexibility of the model by allowing each feature to take on one of a countable number of attribute values.

Tensor relationships among three or more variables are considered by Zenglin Xu, Feng Yan, and Yuan Qi's "Bayesian Nonparametric Models for Multiway Data Analysis". They propose an infinite Tucker decomposition of tensor relationships based on Gaussian or t processes, along with an efficient variational inference scheme. Applications include chemometrics and social network analysis.

5 Research Outlook

Several themes emerge from this collection of BNP research contributions. While Markov chain Monte Carlo methods remain the dominant approach to learning and inference, more scalable alternatives like stochastic variational inference have become competitive alternatives. A growing body of theory and modeling practice allows BNP methods to be used not just for clustering, but in applications with rich hierarchical, temporal, spatial, or relational structure. Together, these modeling and inference advances have enabled increasingly sophisticated applications to challenging problems in engineering, science, medicine, and the humanities. We believe these papers provide a rich foundation for further BNP advances.

Ryan P. Adams, *Guest Editor*
Emily B. Fox, *Guest Editor*
Erik B. Sudderth, *Guest Editor*
Yee Whye Teh, *Guest Editor*

Ryan P. Adams received the bachelor's degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2004 and the PhD degree in physics in 2009 from the University of Cambridge as a Gates Cambridge Scholar. He is an assistant professor of computer science in the School of Engineering and Applied Sciences at Harvard University. His research interests include large-scale Bayesian computation, and Bayesian nonparametrics, with applications in neuroscience, chemistry, and astronomy. He has received paper awards at ICML, AISTATS and UAI, as well as the US Defense Advanced Research Projects Agency (DARPA) Young Faculty Award, and Honorable Mention for the Leonard J. Savage Thesis Award in Theory and Methods.

Emily B. Fox received the SB, MEng, EE, and PhD degrees in 2004, 2005, 2008, and 2009, respectively, from the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology. She is the Amazon assistant professor of machine learning in statistics at the University of Washington. Her research interests include large-scale Bayesian modeling and computations, with an emphasis on time series and Bayesian nonparametrics. She received the US National Science Foundation (NSF) CAREER award, the Leonard J. Savage Thesis Award in Applied Methodology, and MIT EECS Jin-Au Kong Outstanding Doctoral Thesis Prize.

Erik B. Sudderth received the bachelor's degree (summa cum laude, 1999) in electrical engineering from the University of California, San Diego, and the master's and PhD degrees in 2006 in EECS from the Massachusetts Institute of Technology. He is an assistant professor in the Brown University Department of Computer Science. His research interests include Bayesian nonparametrics, probabilistic graphical models, and applications of statistical machine learning in computer vision and the sciences. He received the US National Science Foundation (NSF) CAREER award, and was named one of "AI's 10 to Watch" by *IEEE Intelligent Systems*.

Yee Whye Teh received the bachelor's of mathematics with double honors in computer science and pure mathematics in 1997 from the University of Waterloo and the PhD degree in computer science in 2003 from the University of Toronto. He is a professor of statistical machine learning at the Department of Statistics at the University of Oxford and a Tutorial Fellow at University College, Oxford. His research interests include machine learning and computational statistics, in particular, graphical models and Bayesian nonparametrics. He was a programme co-chair of AISTATS 2010 and MLSS Iceland 2014, and received an ERC Consolidator Grant.