Foundations of Objective Bayesian Methodology

Gutiérrez-Peña, Eduardo (IIMAS-UNAM, Mexico),
Müller, Peter (University of Texas at Austin, United States),
Robert, Christian (Paris Dauphine University, France),
Broderick, Tamara (Massachusetts Institute of Technology, United States)

November 28 - December 3, 2021

1 Overview of the Field

The aim of the workshop was to bring together researchers working on objective Bayesian methodology, Bayesian nonparametric methods and machine learning. The main objective was to allow some of the leading experts in these fields to discuss foundations and important problems in objective Bayesian methodology and how these relate to current research in Bayesian nonparametrics and machine learning.

Objective Bayesian methods are, for the most part, oriented towards the development of Bayesian procedures that can be used automatically, i.e., methods that do not require subjective input other than the specific probabilistic model chosen to describe the data. Recent research directions in the area include L^η, that is weighted log-likelihood techniques; related ideas in variational Bayes methods; approaches to variable selection and model averaging; modularization methods; objective Bayes in graphical models; model assessment and alternatives to p-values; approximate Bayesian computation; and computational challenges and implementations for objective Bayes methods.

As is evident from this list, objective Bayes is evolving from a historically narrower focus on issues related to prior specification to research themes that are more broadly related to default methods in Bayesian analysis. This naturally leads to increasing overlap with Bayesian nonparametrics and machine learning. The proposed workshop aimed to exploit this increasing intersection by bringing together participants from all three research communities. Recent meetings on objective Bayesian analysis have been held in Shanghai, China, 2011; Durham, NC, USA, 2013; Valencia, Spain, 2015; and Austin, TX, 2017. This workshop was planned to continue in the tradition of these previous meetings, but adding a focus on the intersection with Bayesian nonparametrics and machine learning.

Objective Bayes methodology is of dramatically increasing relevance today since application of Bayesian analysis is rapidly growing among nonspecialists, most of whom seek automatic, non-subjective, that is, objective Bayesian procedures. At the same time, computational advances have allowed Bayesian methodology to be employed in problems of such complexity that determination of serious subjective priors is difficult or impossible. For the same reasons machine learning and Bayesian nonparametric methods are increasingly used in many applications, leading to many similar research problems as in objective Bayes. Responding to these developments, the meeting programme emphasized foundational, computational, and theoretical aspects in the three areas, with a focus on themes that are most relevant to the application of methods and perspectives from one of these area in others.

We invited talks on topics including robust, default Bayesian analysis, reproducibility, variable selection, big data and nonparametric Bayes. The programme explored related objective Bayes, Bayesian nonpara-

metric and machine learning methods. To facilitate this aim, the (greatly limited) list of invited participants included a balanced mix of researchers across the three research communities. The same aim is reflected in the composition of the Organizing Committee.

Finally, we note that while the formal organizing committee are the four names stated in the title, we did work closely with Ramses Mena (UNAM, Mexico), who essentially functioned as a 5th member of the organizing committee. Only because the formally recorded committee can not include more than one member from the same institution, we could not include his name.

2 Recent Developments and Open Problems

The workshop programme touched on several recent developments and open problems across the three areas. These included, in particular:

(1) Bayesian nonparametric models over dependent random probability measures, and suitable methods of combining evidence across related samples (related talks mostly on Monday morning);

(2) Machine learning and other computational methods to implement actual inference (related talks mostly on Monday afternoon);

(3) objective and generally robust Bayesian prior models for mixtures (Tuesday morning);

(4) computational challenges for inference in such models (Tuesday afternoon);

(5) alternatives to model-based inference (Wednesday morning);

(6) graphical models (Thursday morning);

(7) applications of machine learning (Thursday afternoon).

Due to scheduling constraints, some talks were scheduled outside this general programme of focused themes.

3 Presentation Highlights

The workshop programme included 30 talks, with 12 talks delivered in-person, and the remaining 18 on-line (see below for our recommendation for future workshops under pandemic-related constraints). Each talk was scheduled for 45 minutes, including usually 35-40 minutes presentation and 5-10 minutes discussion. Additional time slots after the morning session and on Friday morning were reserved for collaborative work and discussions. The programme included a good mix of junior (Ph.D. students, postdoctoral trainees and junior faculty) and senior presenters. The organizers paid attention to including a good number of female speakers and maintaining geographic diversity. The latter was a bit restricted by pandemic-related travel restrictions.

Day 1. The workshop started with a talk by *Igor Prünster* (Bocconi University, Milano, Italy) developing a unified view of prior models for dependent random probability measures, building on the always fascinating notion that a random discrete non-parametric prior is inducing a distribution on the partitions (EPPF). Discussion touched on the applicability in mixtures and their generalisations.

Beatrice Franzolini (Bocconi University, Milano, Italy) discussed specific ways to create dependent distributions based on independent samples.

Marta Catalano (Warwick University, UK) presented her work on partial exchangeability and optimal transportation, including comments on the dependence of the Wasserstein distance on the parameterisation.

Isadora Antoniano-Villalobos (Ca Foscari, Venice, Italy) highlighted the use of BNP methods for extreme value data.

In the afternoon, *Julyan Arbel* (INRIA, Grenoble, France) continued with reference priors for extreme value distributions, with the surprising observation of an ill-defined Jeffreys prior in a particular case (reminding some of the participants of the work of Clara Grazian on Jeffreys priors for mixtures, where some models were not allowing for Fisher information to exist). In a second part of this talk the speaker discussed a modified local versions of Gelman & Rubin's (1992) $\hat{R}$ convergence diagnostic, and the recent modification proposed by Aki Vethari *et al.* (2019). This work appears in [6].

Trevor Campbell (UBC, Vancouver, Canada) developed a practicable and computationally cleverly arranged approach to parallel tempering. This work appears in [8]. The powers of a geometric mixture are optimised by piecewise linear spline functions of the temperature.

María Gil-Leyva (UNAM, Mexico, and Bocconi University, Milano, Italy) presented her original and ordered approach to mixture estimation.

The final talk of the day was by *Anirban Bhattacharya* (Texas A&M University, USA) on high-dimensional Bayesian regression and coupling techniques for checking convergence. The approach uses a very elaborate construct of coupling strategies within a Gibbs sampler, with some steps relying on optimal coupling and others on the use of common random generators.

Day 2. The morning session started with *Helen Ogden* (University of Southhampton, UK) exploring cross-validation for estimating the number k of components for finite mixtures, when using the likelihood as an objective function.

Diana Cai (Princeton University, USA) presented a counter-argument in the sense that she proved, in joint work with Trevor Campbell and Tamara Broderick, that the posterior on k diverges to infinity with the number n of observations if a mixture model is misspecified for said data.

Alexander Ly (University of Amsterdam, Netherlands) discussed Bayes factors for multiple data sets, with asymptotics showing consistency for some (improper) priors if one sample size grows to infinity, while actually attaining the same rate under both hypotheses.

Luis Nieto-Barajas (ITAM, Mexico) presented an approach on uncertainty assessment based on Kullback-Leibler divergence for random probability measures. The approach required a calibration of the Kullback-Leibler divergence in this setting (as it does not enjoy a uniform scale).

Chris Holmes (Oxford University, UK) presented recent work with Edwin Fong and Steven Walker on a predictive approach to Bayesian inference. This work appears in [3]. It is a very original proposal where likelihoods and priors are replaced by the sequence of posterior predictives and only parameters of interest get simulated. The Bayesian flavour of the approach is delicate to assess, though.

In the afternoon session, *Judith Rousseau* (Oxford University, UK) presented recent work on cut posteriors for semi-parametric hidden Markov models. With interesting outcomes for efficiently estimating the transition matrix, the component distributions, and the smoothing distribution.

Sinead Williamson (University of Texas at Austin, USA) spoke on distributed Markov chain Monte Carlo for Bayesian nonparametrics, returning to inference in clustering similar to problems considered by earlier speakers.

Michele Guindani (University of California at Irvine, USA) spoke about clustering distributions, using common atoms mixture models. The discussion included interesting applications to microbiome and calcium imaging (including a mice brain in action). He discussed links with mixtures of mixtures in [4].

Giovanni Rebaudo (University of Texas at Austin, USA) presented a generalised notion of clustering aligned on a graph, with some observations located between the nodes corresponding to clusters. Represented as a random measure with common parameters for the clusters and separated parameters outside. The discussion included a consideration of random partitions, Pólya urns, and species sampling.

Day 3. The day started with *David Rossell* (Pompeu Fabra University, Barcelona, Spain) presenting an empirical Bayes approach to generalised linear model choice with a high degree of confounding, using approximate Laplace approximations. See [5]. Some participants raised the question about the apparent lack of a fully (and objective) Bayesian alternative.

Veronika Rockova (University of Chicago, USA) discussed work on approximate Metropolis-Hastings by classification. Her approximate Bayesian computation implementation replaces the tolerance step by an exponential of minus the estimated Kullback-Leibler divergence between the data density and the density

associated with the current value of the parameter. The classification step need be run at every iteration, which could be sped up by subsampling.

Jack Jewson (Pompeu Fabra University, Barcelona, Spain) exposed his work on generalised Bayesian and improper models.

Rajesh Ranganath (New York University, USA) completed the morning session with a talk on the difficulty of connecting Bayesian models and complex prediction models, proposing instead a game-theoretic approach with Brier scores under censoring.

The final group discussion on Day 3 included a short presentation by *Elena Bortolato* (University of Padova, Italy), posing questions about variations of Jeffreys prior, including the use of fractional powers of Jeffreys prior.

Day 3 afternoon was free for group discussion, co-working and excursions. Working with a local travel agent, Casa Matemática Oaxaca organized a tour to the archaeological site at Monte Alban. About eight participants joined the tour, making it a good opportunity for an informal exchange in a different environment.

Day 4. *Noiritt Chandra* (University of Texas at Austin, USA) presented a novel concept of Bayesian factor analysis for precision matrices. Some participants raised the question of the choice of the number q of factors.

Daniele Durante (Bocconi University, Milano, Italy) spoke about analytical posteriors for probit models using unified skew-Normal priors (based on a 2019 Biometrika paper, [2]).

Filippo Ascolani (Bocconi University, Milano, Italy) introduced trees of random probability measures, with each mother node defining the distribution of the atoms for the descendant nodes [1].

Yang Ni (Texas A&M) concluded the morning with a talk on Bayesian causal graphs.

In the afternoon several speakers highlighted applications of Bayesian nonparametrics, machine learning and objective Bayes methods to substantial scientific inference problems. *José Antonio Perusquia* (University of Kent, UK) reviewed applications of statistical inference on reliability of computer systems, including the use of nonparametric Bayesian priors for binary matrices to detect malware.

Katherine Heller (Google LLC) discussed some examples of machine learning and Bayesian inference in computational health. She discussed in more detail a system to predict sepsis infection in emergency rooms. The system has been deployed at Duke University Hospital, and has greatly improved over existing solutions that were based on an unmodified implementation of a system developed in the UK.

Mengyang Gu (University of California at Santa Barbara, USA) introduced the generalized probabilistic principal component analysis, which generalizes traditional principal component analysis by assuming for each factor a Gaussian process prior. The model allows closed-form maximum maximum likelihood estimation for factor loadings.

In the last talk of the afternoon, *Alan Riva-Palacio* (University of Kent, UK, and UNAM, Mexico) considered inference for multivariate dependent time series data, such as event times under different settings. The talk was based on [7].

4 Scientific Progress Made and Outcome of the Meeting

There was a general consensus that alternative approaches to inference and decision making in the face of uncertainty are viable. Most of the talks explored several versions of a generalized Bayesian inference paradigm and clever, innovative computational schemes. Specific opportunities and links were highlighted in the talks and discussed.

Theory and model developments in Bayesian nonparametrics over the past few years have focused largely on problems related to multiple dependent random probability measures. Related talks were *Igor Prünster*, *Marta Catalano*, *Diana Cai*, *Beatrice Franzolini* and others. Models with well understood theory and practicable computational strategies have developed to a point where they are interesting and applicable for large-scale machine learning and data science applications. Some applications were highlighted in talks by *Giovanni Rebaudo* and others.

An interesting stream of research in objective Bayes and robust, data-driven inference is the construction of methods beyond model-based inference. Great examples are the new approaches presented by *Chris*

Holmes and *Veronika Rockova*. The predictive approach proposed by *Chris Holmes* remains meaningful under any black-box prediction scheme. The discussion included statement of requirements for meaningful consistency in the prediction, essentially just a martingale property of the prediction. In the context of understanding the Bayesian nonparametric, objective Bayes and machine learning interface, this result could be described as building well defined inference based on any good predictive machine, including in particular popular machine learning tools. *Veronika Rockova* introduced an interesting notion of using any universal classifier, including in particular classifiers based on Bayesian additive regression trees or any other machine learning methods, to drive posterior Markov chain Monte Carlo methods. The idea is intriguing. The approach can alternatively be described as introducing a novel summary for approximate Bayesian methods. It is attractive for problems with difficult (or impossible) to evaluate likelihood.

A long-standing area of research on the intersection of machine learning, Bayesian nonparametric and objective Bayes is related to mixture models. Several talks reported progress on current research in this area. *Diana Cai* discussed problems of inconsistency with mixture models. Essentially, under model misspecification, such as using normal kernels to fit mixtures of Laplace kernels, inference on the (unknown) size of the random mixture is inconsistent, not suprisingly. *Michele Guindani* introduced two substantial applications, to inference for microbiome data and for neuroscience data similar to spike trains, and developed interesting novel common atoms mixture models for these problems. The applications are good examples of Bayesian nonparametric methods empowering data science techniques for massive data problems.

Important for the meaningful use of Bayesian nonparametric and machine learning methods in objective Bayes methodology and vice versa is the availability of practicable computational methods. Several talks reported on recent progress. *Trevor Campbell* introduced an implementation of parallel tempering using paths that are indexed in terms of two exponents only. Optimal paths can then be found by optimizing over paths on these pairs of exponents, which is implemented as piece-wise linear splines, making the search for optimal paths entirely feasible.

Some of the most useful contributions of Bayesian nonparametrics and machine learning for robust Bayesian inference occur for high dimensional data applications. Several talks introduced novel methods and computational solutions. *Noirrit Chandra* introduced an interesting approach to factor analysis for precision matrices (rather than the conventional covariance matrix factor analysis). The approach pushes applicability of such models way beyond currently feasible limits in dimensions. Obvious applications are to inference for large graphical models.

Comments and recommendations about programming under pandemic restrictions. CMO-BIRS provided excellent facilities which helped create a favourable atmosphere for collaborations and exchanges of ideas. The workshop gave us the opportunity to bring together researchers from seemingly different backgrounds but who are working on similar problems. It also allowed young researchers to interact with senior researchers in a relaxed, informal way. We believe the workshop brought fresh insights to the table. No doubt it will spawn new projects and collaborations among some of the participants.

Overall the hybrid format with only 15 in-person participants worked. Key was to maintain across all sessions a good mix of in-person and on-line talks. Several participants commented on the benefit of this opportunity for an in-person research meeting, after almost 2 years of pandemic related constraints. There is consensus about the importance of maintaining in-person meetings to the extent possible. This is particularly important for junior researchers, and colleagues from smaller research groups who work in more isolated environments. Everyone appreciated the efforts of CMO-BIRS to make this happen, in a safe environment in compliance with local pandemic-related policies.

General recommendations are:

(i) Increase the in-person limit to perhaps 20, if feasible by sanitary constraints. We felt the limited number of in-person participants created in some instances a limitation in the ensuing interactions.

(ii) Restrict workshops to 4 day programmes, in recognition of the reduced participation.

(iii) Provide more microphones to use for participant comments around the conference room. We feel there should be at least 2 cameras: one for (in-person) speakers, and a second camera for the audience. This, together with a microphone to pass around for comments and questions, will surely foster more spontaneous interactions.

(iv) Provide an additional screen in the conference room showing all on-line participants attending the current zoom session. While most in-person participants bring along their devices, not all of them join the Zoom session and might therefore miss some of the reactions of the on-line participants.

Last, but not least, we would like to express our gratitude to Claudia Arias (Conference Programme Coordinator) and Miguel Altamirano (technical staff) for their constant support that contributed enormously to the success of the workshop. We also thank the staff at Hotel Hacienda Los Laureles; they were always attentive and helpful.

References

[1] F. Ascolani, B. Franzolini, A. Lijoi, and I. Prünster. On the dependence structure in Bayesian nonparametric priors. In *Book of Short Papers*. Italian Statistical Society, 2021.

[2] Daniele Durante. Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika*, 106(4):765–779, 08 2019.

[3] Edwin Fong, Chris Holmes, and Stephen G. Walker. Martingale posterior distributions, 2021.

[4] Sylvia Frühwirth-Schnatter, Gertraud Malsiner-Walli, and Bettina Grün. Generalized Mixtures of Finite Mixtures and Telescoping Sampling. *Bayesian Analysis*, 16: 1–29, 2021.

[5] Miquel Torrens i Dinarès, Omiros Papaspiliopoulos, and David Rossell. Confounder importance learning for treatment effect inference, 2021.

[6] Théo Moins, Julyan Arbel, Anne Dutfoy, and Stéphane Girard. Discussion of the paper "Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC". *Bayesian Analysis*, 16: 711–712, 2021.

[7] Alan Riva Palacio and Fabrizio Leisen. Bayesian nonparametric estimation of survival functions with multiple-samples information. *Electronic Journal of Statistics*, 12: 1330–1357, 2018.

[8] Saifuddin Syed, Vittorio Romaniello, Trevor Campbell, and Alexandre Bouchard-Côté. Parallel tempering on optimized paths, 2021.