Springer Proceedings in Mathematics & Statistics

Sylvia Frühwirth-Schnatter
Angela Bitto
Gregor Kastner
Alexandra Posekany *Editors*

Bayesian Statistics from Methods to Models and Applications

Research from BAYSM 2014

Springer

Springer Proceedings in Mathematics & Statistics

Volume 126

More information about this series at http://www.springer.com/series/10533

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of select contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Sylvia Frühwirth-Schnatter • Angela Bitto
Gregor Kastner • Alexandra Posekany
Editors

Bayesian Statistics from Methods to Models and Applications

Research from BAYSM 2014

Springer

Editors
Sylvia Frühwirth-Schnatter
Institute for Statistics and Mathematics
WU Vienna University of Economics and Business
Vienna, Austria

Angela Bitto
Institute for Statistics and Mathematics
WU Vienna University of Economics and Business
Vienna, Austria

Gregor Kastner
Institute for Statistics and Mathematics
WU Vienna University of Economics and Business
Vienna, Austria

Alexandra Posekany
Institute for Statistics and Mathematics
WU Vienna University of Economics and Business
Vienna, Austria

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-16237-9 ISBN 978-3-319-16238-6 (eBook)
DOI 10.1007/978-3-319-16238-6

Library of Congress Control Number: 2015937960

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

The Contribution of Young Researchers to Bayesian Statistics II

Proceedings of BAYSM 2014

Preface

BAYSM 2014—the second Bayesian Young Statisticians Meeting—took place at the WU Vienna University of Economics and Business, Austria, on September 18–19, 2014. The conference was hosted by the Institute for Statistics and Mathematics of the Department of Finance, Accounting and Statistics. It attracted more than 100 participants from 25 different countries spread over five continents.

Following BAYSM 2013, the first meeting of this kind in Milan, Italy, BAYSM 2014 continues to establish a scientific forum for the next generation of researchers in Bayesian statistics. This inspiring scientific meeting provided opportunities for M.S. students, Ph.D. students, postdoctoral scholars, young researchers, and interested parties from the industry to get in touch with the Bayesian community at large, to expand their professional network, to interact with colleagues, and to exchange ideas.

The scientific program reflected the wide variety of fields in which Bayesian methods are currently employed or could be introduced in the future. Three brilliant keynote lectures by Chris Holmes (University of Oxford), Christian Robert (Université Paris-Dauphine), and Mike West (Duke University) were complemented by 24 plenary talks covering the major topics *Dynamic Models*, *Applications*, *Bayesian Nonparametrics*, *Biostatistics*, *Bayesian Methods in Economics*, and *Models and Methods*, as well as a lively poster session with 30 contributions. The presence of numerous "matured" Bayesians, be it keynote speakers, members of the scientific committee, or senior discussants, provided invaluable inspiration for all attendant young researchers. Throughout the whole workshop, participants were able to discuss open questions, received helpful feedback on their current research, and were encouraged to pursue their line of research.

This volume comprises a peer-reviewed selection of young researchers' contributions presented at BAYSM 2014. It is structured in the following way: The first part, entitled *Theory and Methods*, is dedicated to mathematical statistics, model building, and methodological works, demonstrated by examples. The second part, entitled *Applications and Case Studies*, focuses on the applications of complex methods to real-world problems and data. We want to thank all the authors for their excellent contributions to this volume. Thanks are also due to all reviewers for dedicating time and efforts to the improvement of these young researchers' scientific attempts.

We would like to take this opportunity to express our gratitude to all those people who made BAYSM 2014 an outstanding scientific event and an enjoyable experience. We wish to thank our profound keynote speakers, Chris Holmes, Christian Robert, and Mike West for their inspiring talks and their most valued contributions to a lively meeting. Sincere thanks are given to all participants for the high quality of their presentations. Special thanks go to all the senior discussants for their valuable feedback, especially Jesus Crespo Cuaresma, Bettina Grün, Helga Wagner, and the current President of the International Society for Bayesian Analysis, Sonia Petrone. Finally, we are deeply grateful for the outstanding support we received from the organizing committee, chaired by Karin Haupt, Deputy Head of Office FAS D4, the WU Vienna University of Economics and Business, as well as our sponsors, Accenture, Google, ISBA, and UNIQA.

Hosting this meeting was an exciting and most rewarding experience for us, and we are very pleased that BAYSM 2014 could continue the great success of the first meeting. This extraordinary series of scientific meetings for young researchers in Bayesian statistics will be resumed in June 2016 in Florence, Italy, with BAYSM 2016. Further information can be found on the BAYSM websites at baysm2014.wu.ac.at and baysm.org.

Vienna, Austria
December 2014

Sylvia Frühwirth-Schnatter
Angela Bitto
Gregor Kastner
Alexandra Posekany

Organization

Conference Chairs

Sylvia Frühwirth-Schnatter
Angela Bitto
Gregor Kastner
Alexandra Posekany

Scientific Committee

Raffaele Argiento
Angela Bitto
Jesus Crespo Cuaresma
Sylvia Frühwirth-Schnatter
Bettina Grün
Paul Hofmarcher
Francesca Ieva
Gregor Kastner
Ettore Lanzarone
Alexandra Posekany
Laura Vana
Helga Wagner

Contents

Editors' Biographies

Sylvia Frühwirth-Schnatter is a Professor of Applied Statistics and Econometrics at the Department of Finance, Accounting and Statistics at the WU Vienna University of Economics and Business, Austria. She received her Ph.D. in Mathematics from the Vienna University of Technology in 1988. She has published in many leading journals in applied statistics and econometrics on topics such as Bayesian inference, finite mixture models, Markov switching models, state space models, and their application in economics, finance, and business. In 2014, she became elected member of the Austrian Academy of Science.

Angela Bitto holds a Masters in Mathematics and is currently working on her Ph.D. in Statistics at the Vienna University of Technology. Her research focuses on the Bayesian estimation of sparse time-varying parameter models. Prior to joining the Institute of Statistics and Mathematics at the WU Vienna University of Economics and Business, she worked as a research analyst for the European Central Bank.

Gregor Kastner is an Assistant Professor at the WU Vienna University of Economics and Business and a Lecturer at the University of Applied Sciences in Wiener Neustadt, Austria. He holds Masters in Mathematics, Computer Science, Informatics Management, and Physical Education; in 2014, he received his Ph.D. in Mathematics. Gregor researches the Bayesian modeling of economic time series, in particular the efficient estimation of univariate and high-dimensional stochastic volatility models. His work has been published in leading journals in computational statistics and computer software.

Alexandra Posekany is an Assistant Professor at the Institute of Statistics and Mathematics, WU Vienna University of Economics and Business, Austria. She holds a Ph.D. in Mathematics from the Vienna University of Technology. Her research includes applications of Bayesian analysis in computational biology and econometrics, as well as the development of algorithms and statistical methods in Bayesian computing and big data analysis.

Part I
Theory and Methods

Chapter 1
Bayesian Survival Model Based on Moment Characterization

Julyan Arbel, Antonio Lijoi, and Bernardo Nipoti

Abstract Bayesian nonparametric marginal methods are very popular since they lead to fairly easy implementation due to the formal marginalization of the infinite-dimensional parameter of the model. However, the straightforwardness of these methods also entails some limitations. They typically yield point estimates in the form of posterior expectations, but cannot be used to estimate non-linear functionals of the posterior distribution, such as median, mode or credible intervals. This is particularly relevant in survival analysis where non-linear functionals such as the median survival time play a central role for clinicians and practitioners. The main goal of this paper is to summarize the methodology introduced in (Arbel, Lijoi and Nipoti, Comput. Stat. Data. Anal. 2015) for hazard mixture models in order to draw approximate Bayesian inference on survival functions that is not limited to the posterior mean. In addition, we propose a practical implementation of an R package called **momentify** designed for moment-based density approximation. By means of an extensive simulation study, we thoroughly compare the introduced methodology with standard marginal methods and empirical estimation.

Key words: Bayesian nonparametrics, Completely random measures, Hazard mixture models, Median survival time, Moment-based approximations, Survival analysis

---

J. Arbel (✉)
Collegio Carlo Alberto, Moncalieri, Italy
e-mail: julyan.arbel@carloalberto.org

A. Lijoi
Department of Economics and Management, University of Pavia, Pavia, Italy

Collegio Carlo Alberto, Moncalieri, Italy
e-mail: lijoi@unipv.it

B. Nipoti
Department of Economics and Statistics, University of Torino, Torino, Italy

Collegio Carlo Alberto, Moncalieri, Italy
e-mail: bernardo.nipoti@carloalberto.org

S. Frühwirth-Schnatter et al. (eds.), *Bayesian Statistics from Methods to Models and Applications*, Springer Proceedings in Mathematics & Statistics 126,
DOI 10.1007/978-3-319-16238-6_1

1.1 Introduction

With *marginal methods* in Bayesian nonparametrics we refer to inferential procedures which rely on the integration (or marginalization) of the infinite-dimensional parameter of the model. This marginalization step is typically achieved by means of the so-called Blackwell–MacQueen Pólya urn scheme. We consider the popular example of the Dirichlet process [4] to illustrate the idea. Denote by $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ an exchangeable sequence of random variables to which we assign as a prior distribution a Dirichlet process with mass parameter M and base measure G_0, that is

$$Y_i | G \overset{\text{iid}}{\sim} G,$$
$$G \sim DP(M, G_0).$$

The marginal distribution of $\boldsymbol{Y}$, once G has been integrated out, can be derived from the set of predictive distributions for Y_i, given $(Y_1, \ldots, Y_{i-1})$, for each $i = 1, \ldots, n$. In this case, such conditional distributions are linear combinations between the base measure G_0 and the empirical distribution of the conditioning variables and are effectively described through a Pólya urn sampling scheme. Marginal methods have played a major role in the success of Bayesian nonparametrics since the Pólya urn generally leads to ready to use Markov chain Monte Carlo (MCMC) sampling strategies which, furthermore, immediately provide Bayesian point estimators in the form of posterior means. A popular example is offered by mixtures of the Dirichlet process for density estimation; for the implementation, see, e.g., the R package **DPpackage** by Jara et al. [9]. However, the use of marginal methods has important limitations that we wish to address here. Indeed, one easily notes that the posterior estimates provided by marginal methods are not suitably endowed with measures of uncertainty such as posterior credible intervals. Furthermore, using the posterior mean as an estimator is equivalent to choosing a square loss function which does not allow for other types of estimators such as median or mode of the posterior distribution. Finally, marginal methods do not naturally lead to the estimation of non-linear functionals of the distribution of a survival time, such as the median survival time. For a discussion of these limitations, see, e.g., Gelfand and Kottas [5].

The present paper aims at proposing a new procedure that combines closed-form analytical results arising from the application of marginal methods with an approximation of the posterior distribution which makes use of posterior moments. The whole machinery is developed for the estimation of survival functions that are modeled in terms of hazard rate functions. To this end, let F denote the cumulative distribution function (CDF) associated with a probability distribution on $\mathbb{R}^+$. If F is absolutely continuous, then the corresponding survival function and cumulative hazard rate are defined, respectively, by $S(t) = 1 - F(t)$ and $H(t) = -\log(S(t))$, and the hazard rate function is given by $h(t) = -S'(t)/S(t)$. Let us recall that survival analysis has been a very active area of application of Bayesian nonparametric methodology: neutral to the right processes were used by [2] as a prior for the

CDF F, and beta processes by [6] as a prior for the cumulative hazard function H, both benefiting from useful conjugacy properties. Here we specify a prior on the hazard rate h. The most popular example is the gamma process mixture, originally proposed in [3]. More general models have been studied in later work by [10] and [8]. Bayesian inference for these models often relies on a marginal method, see, e.g., [7]. Although quite simple to implement, marginal methods typically yield estimates of the hazard rate, or equivalently of the survival function, only in the form of the posterior mean at a fixed time point. Working along the lines of Arbel et al. [1], we show that a clever use of a moment-based approximation method does provide a relevant upgrade on the type of inference one can draw via marginal sampling schemes. We should stress that the information gathered by marginal methods is not confined to the posterior mean but is actually much richer and, if properly exploited, can lead to a more complete posterior inference.

Let us briefly introduce Bayesian hazard mixture models. Random parameters, such as the hazard rate and survival function, are denoted with a tilde on top, e.g. $\tilde{h}$ and $\tilde{S}$. We endow $\tilde{h}$ with a prior distribution defined by the distribution of the random hazard rate (RHR)

$$\tilde{h}(t) = \int_{\mathbb{Y}} k(t;y)\tilde{\mu}(\mathrm{d}y), \tag{1.1}$$

where $\tilde{\mu}$ is a completely random measure (CRM) on $\mathbb{Y} = \mathbb{R}^+$, and $k(\,\cdot\,;\cdot\,)$ denotes a transition kernel on $\mathbb{R}^+ \times \mathbb{Y}$. Under suitable assumption on the CRM $\tilde{\mu}$, we have $\lim_{t\to\infty} \int_0^t \tilde{h}(s)\mathrm{d}s = \infty$ with probability 1. Therefore, we can adopt the following model

$$\begin{aligned} X_i \,|\, \tilde{P} &\overset{\text{iid}}{\sim} \tilde{P} \\ \tilde{P}((t,\infty)) &\overset{\text{d}}{=} \tilde{S}(t) \overset{\text{d}}{=} \exp\left(-\int_0^t \tilde{h}(s)\,\mathrm{d}s\right), \end{aligned} \tag{1.2}$$

for a sequence of (possibly censored) survival data $\boldsymbol{X} = (X_1,\ldots,X_n)$. In this setting, [3] characterizes the posterior distribution of the so-called *extended gamma process*: this is obtained when $\tilde{\mu}$ is a gamma CRM and $k(t;y) = \mathbb{1}_{(0,t]}(y)\,\beta(y)$ for some positive right-continuous function $\beta : \mathbb{R}^+ \to \mathbb{R}^+$. The same kind of result is proved in [10] for *weighted gamma processes* corresponding to RHRs obtained when $\tilde{\mu}$ is still a gamma CRM and $k(\,\cdot\,;\cdot\,)$ is an arbitrary kernel. Finally, a posterior characterization has been derived by [8] for any CRM $\tilde{\mu}$ and kernel $k(\,\cdot\,;\,\cdot\,)$.

The rest of the paper is organized as follows. In Sect. 1.2, we provide the closed-form expressions for the posterior moments of the survival function. We then show in Sect. 1.3 how to exploit the expression for the moments to approximate the corresponding density function and sample from it. Finally, in Sect. 1.4 we study the performance of our methodology by means of an extensive simulation study with survival data.