Computational Statistics and Data Analysis 93 (2016) 359–372

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Full Bayesian inference with hazard mixture models

Julyan Arbel [a], Antonio Lijoi [b,a], Bernardo Nipoti [c,a,*]

[a] Collegio Carlo Alberto, via Real Collegio, 30, 10024 Moncalieri, Italy

[b] Department of Economics and Management, University of Pavia, Via San Felice 5, 27100 Pavia, Italy

[c] Department of Economics and Statistics, University of Torino, C.so Unione Sovietica 218/bis, 10134 Torino, Italy

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form 4 December 2014
Accepted 7 December 2014
Available online 15 December 2014

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

ABSTRACT

Bayesian nonparametric inferential procedures based on Markov chain Monte Carlo marginal methods typically yield point estimates in the form of posterior expectations. Though very useful and easy to implement in a variety of statistical problems, these methods may suffer from some limitations if used to estimate non-linear functionals of the posterior distribution. The main goal is to develop a novel methodology that extends a well-established marginal procedure designed for hazard mixture models, in order to draw approximate inference on survival functions that is not limited to the posterior mean but includes, as remarkable examples, credible intervals and median survival time. The proposed approach relies on a characterization of the posterior moments that, in turn, is used to approximate the posterior distribution by means of a technique based on Jacobi polynomials. The inferential performance of this methodology is analyzed by means of an extensive study of simulated data and real data consisting of leukemia remission times. Although tailored to the survival analysis context, the proposed procedure can be adapted to a range of other models for which moments of the posterior distribution can be estimated.

1. Introduction

Most commonly used inferential procedures in Bayesian nonparametric practice rely on the implementation of sampling algorithms that can be gathered under the general umbrella of Blackwell–MacQueen Pólya urn schemes. These are characterized by the marginalization with respect to an infinite-dimensional random element that defines the de Finetti measure of an exchangeable sequence of observations or latent variables. Henceforth these will be referred to as *marginal methods*. Besides being useful for the identification of the basic building blocks of ready to use Markov chain Monte Carlo (MCMC) sampling strategies, marginal methods have proved to be effective for an approximate evaluation of Bayesian point estimators in the form of posterior means. They are typically used with models for which the predictive distribution is available in closed form. Popular examples are offered by mixtures of the Dirichlet process for density estimation (Escobar and West, 1995) and mixtures of gamma processes for hazard rate estimation (Ishwaran and James, 2004). While becoming well-established tools, these computational techniques are easily accessible also to practitioners through a straightforward software implementation (see for instance Jara et al., 2011). Though it is important to stress their relevance both in theory and in practice, it is also worth pointing out that Blackwell–MacQueen Pólya urn schemes suffer from some drawbacks which we wish to address here. Indeed, one easily notes that the posterior estimates provided by marginal methods are

* Corresponding author at: Department of Economics and Statistics, University of Torino, C.so Unione Sovietica 218/bis, 10134 Torino, Italy. Tel.: +39 011 6705023.

E-mail addresses: julyan.arbel@carloalberto.org (J. Arbel), lijoi@unipv.it (A. Lijoi), bernardo.nipoti@carloalberto.org (B. Nipoti).

http://dx.doi.org/10.1016/j.csda.2014.12.003

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 whereas in many situations of interest other choices such as absolute error or 0–1 loss functions and, as corresponding estimators, median or mode of the posterior distribution of the survival function, at any fixed time point t, would be preferable. Finally, they do not naturally allow inference on functionals of the distribution of survival times, such as the median survival time, to be drawn. A nice discussion of these issues is provided by Gelfand and Kottas (2002) where the focus is on mixtures of the Dirichlet process: the authors suggest complementing the use of marginal methods with a sampling strategy that aims at generating approximate trajectories of the Dirichlet process from its truncated stick-breaking representation.

The aim is to propose 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 to a probability distribution on $\mathbb{R}^+$. The corresponding survival and cumulative hazard functions are denoted as

$$S(t) = 1 - F(t) \quad \text{and} \quad H(t) = -\int_{[0,t]} \frac{\mathrm{d}F(s)}{F(s-)},$$

for any $t > 0$, respectively, where $F(s-) := \lim_{\varepsilon \downarrow 0} F(s - \varepsilon)$ for any positive s. If F is absolutely continuous, one has $H(t) = -\log(S(t))$ and the hazard rate function associated to F is, thus, defined as $h(t) = F'(t)/[1 - F(t-)]$. It should be recalled that survival analysis has been one of the most relevant areas of application of Bayesian nonparametric methodology soon after the groundbreaking contribution of Ferguson (1973). A number of papers in the '70s and the '80s have been devoted to the proposal of new classes of priors that accommodate for a rigorous analytical treatment of Bayesian inferential problems with censored survival data. Among these it is worth mentioning the neutral to the right processes proposed in Doksum (1974) and used to define a prior for the CDF F: since they share a conjugacy property they represent a tractable tool for drawing posterior inferences. Another noteworthy class of priors has been proposed in Hjort (1990), where a beta process is used as a nonparametric prior for the cumulative hazard function H. Also in this case, one can considerably benefit from a useful conjugacy property.

As already mentioned, the plan consists in proposing a method for full Bayesian analysis of survival data by specifying a prior on the hazard rate h. The most popular example is the gamma process mixture that has been originally proposed in Dykstra and Laud (1981) and generalized in later work by Lo and Weng (1989) and James (2005) to include any mixing random measure and any mixed kernel. Recently Lijoi and Nipoti (2014) have extended such framework to the context of partially exchangeable observations. The uses of random hazard mixtures in practical applications have been boosted by the recent developments of powerful computational techniques that allow for an approximate evaluation of posterior inferences on quantities of statistical interest. Most of these arise from a marginalization with respect to a completely random measure that identifies the de Finetti measure of the exchangeable sequence of observations. See, e.g., Ishwaran and James (2004). Though they are quite simple to implement, the direct use of their output can only yield point estimation of the hazard rates, or of the survival functions, at fixed time points through posterior means. The main goal of the present paper is to 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. The takeaway message is 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. To understand this, one can refer to a sequence of exchangeable survival times $(X_i)_{i\geq 1}$ such that $\mathbb{P}[X_1 > t_1, \ldots, X_n > t_n \mid \tilde{P}] = \prod_{i=1}^n \tilde{S}(t_i)$ where $\tilde{P}$ is a random probability measure on $\mathbb{R}^+$ and $\tilde{S}(t) = \tilde{P}((t, \infty))$ is the corresponding random survival function. Given a suitable sequence of latent variables $(Y_i)_{i\geq 1}$, a closed-form expression for

$$\mathbb{E}[\tilde{S}^r(t) \mid \boldsymbol{X}, \boldsymbol{Y}], \quad \text{for any } r \geq 1, \text{ and } t > 0, \tag{1}$$

with $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, will be provided. Our strategy consists in approximating the posterior distribution of $\tilde{S}(t)$, at each instant t, and relies on the fact that, along with the posterior mean, marginal models allow to straightforwardly estimate posterior moments of any order of $\tilde{S}(t)$. Indeed, an MCMC sampler yields a sample from the posterior distribution of $\boldsymbol{Y}$ given $\boldsymbol{X}$: this can be used to integrate out the latent variables appearing in (1) and obtain a numerical approximate evaluation of the posterior moments $\mathbb{E}[\tilde{S}^r(t) \mid \boldsymbol{X}]$. These are finally used to deduce, with almost negligible effort, an approximation of the posterior distribution of $\tilde{S}(t)$ and, in turn, to estimate some meaningful functionals of $\tilde{S}(t)$.

It is to be mentioned that one could alternatively resort to a different approach that boils down to the simulation of the trajectories of the completely random measure that defines the underlying random probability measure from its posterior distribution. In density estimation problems, this is effectively illustrated in Nieto-Barajas et al. (2004), Nieto-Barajas and Prünster (2009) and Barrios et al. (2013). As for hazard rates mixtures estimation problems, one can refer to James (2005); Nieto-Barajas and Walker (2004) and Nieto-Barajas (2014). In particular, James (2005) provides a posterior characterization that is the key for devising a Ferguson and Klass (1972) representation of the posterior distribution of the completely random measure which enters the definition of the prior for the hazards. Some numerical aspects related to the implementation of the algorithm can be quite tricky since one needs to invert the Lévy intensity to simulate posterior jumps and a set of suitable