Bayesian Hazard Mixture Models Based on Moments Characterization

Julyan Arbel[1], Antonio Lijoi[1,2], Bernardo Nipoti[1,3]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Collegio Carlo Alberto,
via Real Collegio, 30, 10024 Moncalieri, Italy
`<julyan.arbel, bernardo.nipoti>@carloalberto.org`

[2] Department of Economics and Management,
University of Pavia, Via San Felice 5, 27100 Pavia, Italy
`lijoi@unipv.it`

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

Abstract

Bayesian nonparametric marginal methods typically yield point estimates in the form of posterior expectations. Though very useful and easy to implement, these methods may suffer from some limitations if used to estimate non–linear functionals of posterior distributions, such as credible intervals. This is particularly relevant in survival analysis where various estimators for the same survival function can lead to different estimates. The main goal of the paper summarized in this extended abstract is to develop a novel methodology for hazard mixture models in order to draw approximate inference on survival functions that is not limited to the posterior mean.

Keywords: Bayesian nonparametrics; hazard mixture models; moment-based approximations; survival analysis; completely random measures.

In this extended abstract, we announce results which will be extensively presented and proved in Arbel et al. (2014). We briefly review in Section 1 hazard mixture models, then provide in Section 2 our main result on the characterization of the posterior moments of the survival function, and finally show in Section 3 how such a piece of information can be used in order to carry out Bayesian inference not restricted to the posterior mean.

1 Introduction

A well–known nonparametric prior for the hazard rate function $\tilde{h}$ within multiplicative intensity models used in survival analysis arises as a mixture of *completely random measures* (CRMs). If $k(\cdot\,;\cdot)$ is a transition kernel on $\mathbb{R}^+ \times \mathbb{Y}$, a prior for $\tilde{h}$ is the distribution of the random hazard rate (RHR)

$$\tilde{h}(t) = \int_{\mathbb{Y}} k(t;y)\tilde{\mu}(\mathrm{d}y), \tag{1}$$

where $\tilde{\mu}$ is a CRM on $\mathbb{Y}$. We observe that, if $\lim_{t\to\infty}\int_0^t \tilde{h}(s)\mathrm{d}s = \infty$ with probability 1, then one can adopt the following model

$$\begin{aligned} X_i \,|\, \tilde{P} &\overset{\text{iid}}{\sim} \tilde{P} \\ \tilde{P}((\,\cdot\,,\infty)) &\overset{\text{d}}{=} \exp\left(-\int_0^{\,\cdot} \tilde{h}(s)\,\mathrm{d}s\right) \end{aligned} \tag{2}$$

for a sequence of (possibly censored) survival data $\boldsymbol{X} = (X_i)_{1\le i\le n}$. This means that $\tilde{h}$ in (1) defines a random survival function

$$t \mapsto \tilde{S}(t) = \exp(-\int_0^t \tilde{h}(s)\mathrm{d}s).$$

In this setting, Dykstra and Laud (1981) characterize 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 Lo and Weng (1989) 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 James (2005) for any CRM $\tilde{\mu}$ and kernel $k(\,\cdot\,;\,\cdot\,)$.

2 Hazard mixture models and moment approximations

A useful augmentation suggests introducing latent random variables $\boldsymbol{Y} = (Y_1,\dots,Y_n)$ such that, building upon the posterior characterization derived by James (2005), we can derive expressions for the posterior moments of the random variable $\tilde{S}(t)$, conditionally on $\boldsymbol{X}$ and $\boldsymbol{Y}$. To this end, define $K_x(y) = \int_0^x k(s;y)\mathrm{d}s$ and $K_{\boldsymbol{X}}(y) = \sum_{i=1}^n K_{X_i}(y)$. Also, the almost sure discreteness of $\tilde{\mu}$ implies there might be ties among the Y_i's with positive probability. Therefore, we denote the distinct values among $\boldsymbol{Y}$ with $(Y_1^*,\dots,Y_k^*)$, where $k \le n$, and, for any $j = 1,\dots,k$, we define $C_j = \left\{l : Y_l = Y_j^*\right\}$ and $n_j = \#C_j$, the cardinality of C_j.

Proposition 1 *Denote by $\nu(\mathrm{d}s,\mathrm{d}y) = \rho(s)\,\mathrm{d}s\,c\,P_0(\mathrm{d}y)$ the Lévy intensity of $\tilde{\mu}$. Then for every $t > 0$ and $r > 0$,*

$$\begin{aligned} \mathbb{E}[\tilde{S}^r(t)\,|\,\boldsymbol{X},\boldsymbol{Y}] = \exp&\left\{-\int_{\mathbb{R}^+\times\mathbb{Y}} \left(1 - \mathrm{e}^{-rK_t(y)s}\right)\mathrm{e}^{-K_{\boldsymbol{X}}(y)s}\nu(\mathrm{d}s,\mathrm{d}y)\right\} \\ &\times \prod_{j=1}^k \frac{1}{B_j}\int_{\mathbb{R}^+} \exp\left\{-s\left(rK_t(Y_j^*) + K_{\boldsymbol{X}}(Y_j^*)\right)\right\} s^{n_j}\rho(s)\mathrm{d}s, \end{aligned} \tag{3}$$

where $B_j = \int_{\mathbb{R}^+} s^{n_j}\exp\left\{-sK_{\boldsymbol{X}}(Y_j^*)\right\}\rho(s)\mathrm{d}s$, for $j = 1,\dots,k$.

3 Bayesian inference

By an extensive simulation study, we show that the posterior distribution of $\tilde{S}(t)$, for each $t > 0$, can be approximated with the N first posterior moments estimated in Proposition 1 with a precision increasing with N. This allows us to devise an algorithm for carrying out full Bayesian inference on survival data. This is applied on a dataset involving leukemia remission times (see for example Cox, 1972).

Particular attention in the illustration is dedicated to posterior mean, posterior median, posterior mode and highest posterior density (HPD) intervals, plotted in Figure 1. However, any functional of interest of the posterior distribution of $\tilde{S}(t)$ can be estimated by the proposed algorithm. By inspecting the left part of Figure 1, it is apparent that, for large values of t, posterior mean, median and mode show significantly different behaviors, with posterior mean being more optimistic than posterior median and mode. It is worth stressing that such differences, while very meaningful for clinicians, could not be captured by marginal methods for which only the posterior mean would be available. Furthermore, marginal methods generally underestimate the uncertainty associated to posterior estimates. This is clearly shown in the right part of Figure 1 where we have compared the estimated 95% HPD intervals for $\tilde{S}(t)$ and the intervals corresponding to the marginal method.

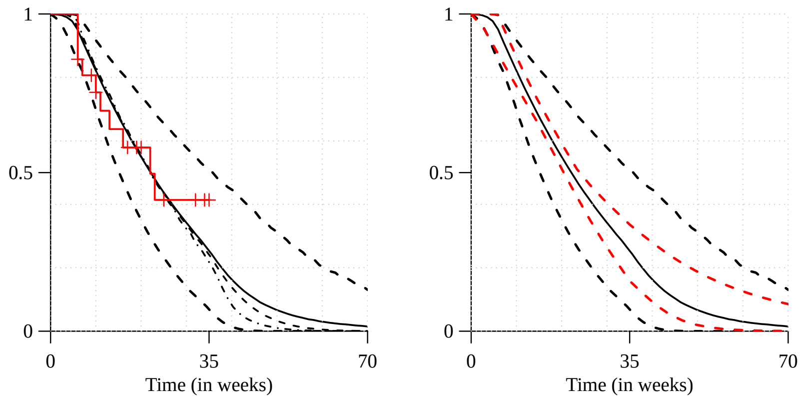

Figure 1: Left: posterior mean (solid line), median (dashed line), mode (point dashed line), 95% highest posterior density credible intervals (thick dashed line), and Kaplan-Meier estimate (red) for $\tilde{S}(t)$. Right: 95% highest posterior density credible interval (dashed black line) and 95% marginal credible interval (dashed red line) for $\tilde{S}(t)$.

References

Arbel, J., Lijoi, A., and Nipoti, B. (2014). Full Bayesian inference with hazard mixture models. *arXiv preprint arXiv:1405.6628.*

Cox, D. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. A*, 34:187–202.

Dykstra, R. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.*, 9:356–367.

James, L. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.*, 33:1771–1799.

Lo, A. and Weng, C. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.*, 41:227–245.