On Bayesian Analysis of the Proportional Hazards Model

Sull'Analisi Bayesiana del Modello a Rischi Proporzionali

Luca La Rocca
Dipartimento di Matematica, Università di Pavia, e-mail: luca@dimat.unipv.it

Riassunto: Questo lavoro tratta il problema dell'inferenza bayesiana per il modello a rischi proporzionali di Cox (1972). Dopo una breve panoramica di carattere storico, volta a presentare la letteratura e gli aspetti salienti del problema, si propone una soluzione originale di tipo MCMC, basata su un nuovo approccio bayesiano non parametrico alla stima della funzione di intensità. La soluzione proposta, che permette di trattare non solo osservazioni esatte, ma anche osservazioni censurate da destra, viene sperimentata sui tempi di remissione leucemica originariamente considerati da Cox (1972).

Keywords: Hazard rate, Markov Chain Monte Carlo, Right censoring, Survival analysis.

1. Introduction

In the well known proportional hazards model, introduced by Cox (1972) as a tool for regression in life-table analysis, the unknown hazard rate ρ_i of survival time T_i is factored as $\rho_i = e^{<\beta,x_i>}\rho_\star$, where $< \cdot,\cdot >$ denotes the ordinary inner product in $\mathbb{R}^p$, β is a vector of regression parameters, x_i is a vector of covariates observed together with T_i, and $\rho_\star$ is the so called *baseline hazard rate*, corresponding to subjects with $x_i = 0$. Strictly speaking, a Bayesian analysis of the above model requires building a random process $\rho_\star$, together with a random vector β, so that their joint law expresses the prior beliefs on the statistical experiment; then it is necessary to compute the posterior distribution of $\rho_\star$ and β, given the observed times $T_1, \ldots, T_n$ and covariates $x_1, \ldots, x_n$. The simple treatment/placebo scenario, defined by $p = 1$ and $x_i \in \{0, 1\}$, $\forall i$, is of particular interest, the main goal being to determine whether the *hazard ratio* $\zeta = e^\beta$ is significantly different from one.

A version of the proportional hazards model was considered by Kalbfleisch (1978), who factored the minus-log-survival function of T_i as $-\log(1-\Phi_i) = -e^{\beta x_i}\log(1-\Phi_\star)$, where Φ_i is the unknown cumulative distribution function of T_i and $\Phi_\star$ is the baseline one. Following an empirical Bayes approach, Kalbfleisch (1978) proposed a gamma process prior for $-\log(1-\Phi_\star)$, then found β by maximizing its marginal likelihood and eventually was able to compute the posterior law of $-\log(1-\Phi_\star)$. The problem, as pointed out by Kalbfleisch (1978) himself, is that the gamma process almost surely gives rise to pure jump distributions, while it is for the continuous ones that the minus-log-survival function coincides with the integral function of the hazard rate, that is the cumulative hazard.

In a later work, Hjort (1990) introduced the beta process to model the integral function of $(1-\Phi_{\star-})^{-1}d\Phi_\star$, which is in all cases the cumulative hazard, even if $\Phi_\star$ has jumps; note that $\Phi_{\star-}$ in the above integrand is the left-continuous version of $\Phi_\star$. Hjort (1990) was able to compute the posterior distribution of the cumulative hazard as a function of the parameter β, working with the same factorization as Kalbfleisch (1978); he then suggested to plug-in an empirical estimate of β, thus following an empirical Bayes approach, or possibly introduce a prior for β too and look for a simulation-based solution. Note that the beta process prior, like the gamma one, almost surely produces pure jump distributions.

An MCMC solution for the above model, with a beta process prior, was subsequently carried out by Laud *et al.* (1998); meanwhile, Clayton (1991) had proposed an MCMC algorithm for a frailty version of the model at issue, adopting a gamma process prior. See the paper by Sinha and Dey (1998) for an overview of Bayesian inference for general semiparametric survival models, including frailty ones, and refer to the book by Ibrahim *et al.* (2001) for a broad review of Bayesian survival analysis. Kim and Lee (2003) have recently developed the pure jump baseline approach, by considering general neutral to the right process priors and allowing for left truncation, as well as right censoring.

In the following, only the simple proportional hazards model will be considered, but care will be taken in order to ensure that the baseline distribution be absolutely continuous, in the original spirit of Cox (1972); furthermore, the baseline hazard rate will be smooth, which can be a very desirable property. Indeed, even in the case of a discrete model, it was a preference for somewhat smooth estimates that recently led to the introduction of correlated gamma process priors by Nieto-Barajas and Walker (2002).

2. A Novel Approach

It is here suggested that the simple proportional hazards model, defined by the likelihood

$$\mathcal{L}(T \mid o, x; \rho_\star, \zeta) \;=\; \prod_{i=1}^{n} [\zeta^{x_i} \rho_\star(T_i)]^{o_i} \exp\left\{ -\zeta^{x_i} \int_0^{T_i} \rho_\star(s)\, ds \right\} \tag{1}$$

where $x_i \in \{0, 1\}$ is the group indicator for the survival time $T_i > 0$ and $o_i = 1$ if T_i is an exact observation, while $o_i = 0$ if T_i is right censored, $i = 1 \ldots n$, be analyzed taking the hazard ratio $\zeta \sim \mathcal{G}(c, d)$ independent of the baseline hazard rate

$$\rho_\star(t) \;=\; q[1 - K(t)]\xi_0 + \sum_{j=1}^{\infty} \xi_j k(t - \sigma_j), \quad t \in \mathbb{R}_+ \tag{2}$$

where $0 = \sigma_0 < \sigma_1 < \sigma_2 < \ldots$ are the event times of a homogeneous Poisson process with intensity q, independent of $\xi_0, \xi_1, \xi_2 \ldots$ which are i.i.d. with $\xi_1 \sim \mathcal{G}(a, b)$, while k is the standard normal density and K the corresponding cumulative distribution function.

The likelihood (1) is standard in this context, except for being written in terms of ζ rather than $\beta = \log \zeta$; this choice allows using a conjugate gamma prior and also a greater ease of interpretation. On the other hand, the process prior (2) is an original proposal by La Rocca (2003), basically consisting in a convolution mixture of the standard normal density by a compound Poisson process. Similar mixtures can be found in a recent paper by James (2003), where the compound Poisson process is replaced by a weighted gamma measure and a full posterior analysis of the proportional hazards model is shown to be possible. As for the Bayesian model defined by (1) and (2), the discrete nature of the compound Poisson process makes it direct to implement an *ad hoc* MCMC algorithm; see the Ph.D. thesis by La Rocca (2003) for details.

Turning to the choice of hyper-parameters, first recall that $\rho_\star$ is the hazard rate for T_i such that $x_i = 0$; therefore, it is possible to set a, b and q by eliciting a prior for the survival times in group 0 only. A procedure to impose a given constant prior expected hazard rate, while controlling the prior variability and the extent to which the model is practically nonparametric, has been suggested by La Rocca (2003). With regards to the choice of c and d, assuming data should express whether ζ is greater or lesser than 1,

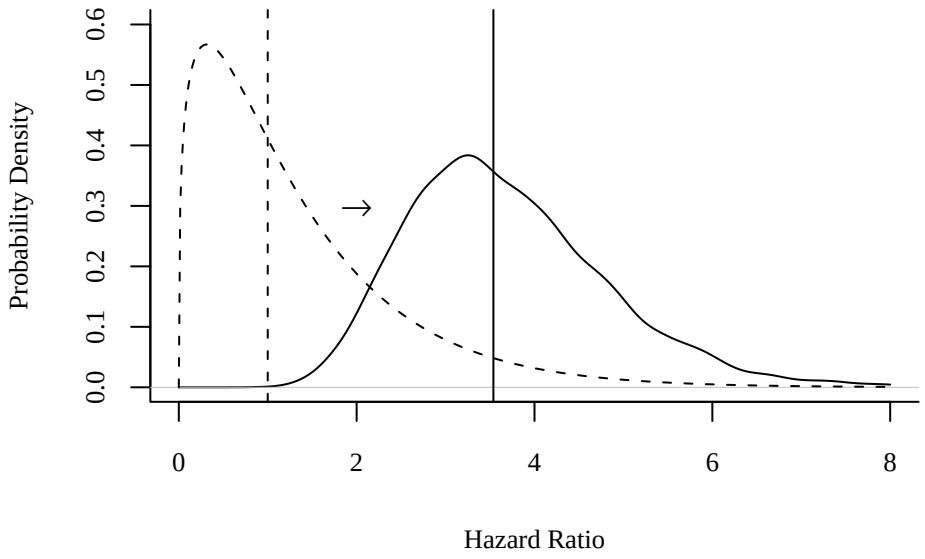
it can be wise to impose $\mathbb{P}\{\zeta > 1\} = \mathbb{P}\{\zeta < 1\}$; then c follows from d, which can be set to control prior variability. For example, setting $d = 1$ gives $c \simeq 1.31425$ and the resulting gamma prior for ζ roughly corresponds to the "standard" choice $\beta \sim \mathcal{N}(0, 1)$.

3. Data Analysis

In order to validate the suggested approach, a well known dataset is considered. These data, consisting of 21 treatment/placebo pairs of leukemia remission times, with 12 right censored observations in the group of treated patients, have also been analyzed by Cox (1972), Kalbfleisch (1978), Laud *et al.* (1998) and Ibrahim *et al.* (2001, pages 75–78). It is clear, even from visual inspection, that remission is shorter for patients receiving placebo. Andersen *et al.* (1993, pages 22–23) describe the dataset in greater detail, also noting that data are actually multivariate, since treatment/placebo pairs were coupled by initial state of remission; this feature will be neglected here—see Sinha and Dey (1998) for a suitable frailty model and its Bayesian analysis—as it is not of primary importance for validating the suggested approach and this choice allows direct comparison to the above authors.

Figure 1 displays a prior to posterior plot for the hazard ratio ζ, relative to an analysis whose details are given by La Rocca (2003). The dashed curve denotes the prior density of $\mathcal{G}(1.31425, 1)$, with the corresponding vertical line marking the prior median 1. The solid line denotes a standard kernel estimate of the posterior density, obtained from an MCMC sample, with the corresponding vertical line marking the posterior median, that is $\zeta_{0.5} = 3.54$. A shift of the density to the right is apparent, meaning a greater hazard in the group receiving placebo and therefore proving the effectiveness of treatment; in fact, the value 1 does not belong to the posterior credible interval $(\zeta_{0.025}, \zeta_{0.975}) = (1.93, 6.28)$. Note that the posterior median $\zeta_{0.5} = 3.54$ is intermediate between the prior median 1 and the empirical value 5.21 found by Cox (1972).

Figure 1: *Prior to posterior plot for ζ.*



Consider now the parameter β. It is immediate to compute $\beta_{0.5} = 1.26$ and it turns out that this value coincides with the posterior expected value $\hat{\beta}$. Laud *et al.* (1998)

found values for $\hat{\beta}$ ranging from 1.62 to 1.71, depending on a precision hyper-parameter; note that the prior expected value of β had been set to 1.5. Similarly, empirical Bayes values ranging from 1.46 to 1.61 were found by Kalbfleisch (1978). Ibrahim *et al.* (2001, pages 75–78) instead found $\hat{\beta} = 1.59$ by using a correlated gamma process prior for $\rho_\star$ and a zero mean normal prior for β. Recall that the empirical value by Cox (1972) was 1.65. Therefore, the estimate obtained by the approach proposed in this short paper is clearly the most conservative and this is because the prior distribution for ζ has been chosen in such a way that strong empirical evidence is needed if $\zeta \neq 1$ is to be inferred.

Although in the analysis of the proportional hazards model the focus is usually on the hazard ratio, rather than the baseline hazard rate, the suggested prior distribution has the advantage of selecting absolutely continuous distributions. This preserves the original interpretation of the model by Cox (1972) and implies that posterior analysis gives a smooth estimate of the baseline distribution, which was not the case in the approaches by Kalbfleisch (1978) and Laud *et al.* (1998). Note that the ability to estimate the baseline hazard rate, as well as the different hazard rates in the two groups, by disjoint analyses, allows full exploitation of the graphical technique suggested by Cox (1972) to assess the proportional hazards hypothesis.

References

Andersen P.K., Borgan O., Gill R.D. and Keiding N. (1993) *Statistical Models Based on Counting Processes*, Springer-Verlag.

Clayton D.G. (1991) A Monte Carlo method for Bayesian inference in frailty models, *Biometrics*, 47, 467–485.

Cox D.R. (1972) Regression models and life-tables, *Journal of the Royal Statistical Society, Series B*, 34, 187–220, with discussion.

Hjort N.L. (1990) Nonparametric Bayes estimators based on beta processes in models for life history data, *Annals of Statistics*, 18, 1259–1294.

Ibrahim J.G., Chen M.H. and Sinha D. (2001) *Bayesian Survival Analysis*, Springer-Verlag.

James L.F. (2003) Bayesian calculus for gamma processes with applications to semiparametric intensity models, *Sankhyā, Series A*, 65, 196–223.

Kalbfleisch J.D. (1978) Nonparametric Bayesian analysis of survival time data, *Journal of the Royal Statistical Society, Series B*, 40, 214–221.

Kim Y. and Lee J. (2003) Bayesian analysis of proportional hazard models, *Annals of Statistics*, 31, 493–511.

La Rocca L. (2003) *Stima bayesiana non parametrica della funzione di intensità con applicazione alla valutazione della pericolosità sismica*, Ph.D. thesis, Dottorato in Statistica Matematica, Università degli Studi di Pavia, XV ciclo.

Laud P.W., Damien P. and Smith A.F.M. (1998) Bayesian nonparametric and covariate analysis of failure time data, in: *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer-Verlag, 213–225.

Nieto-Barajas L.E. and Walker S.G. (2002) Markov beta and gamma processes for modelling hazard rates, *Scandinavian Journal of Statistics*, 29, 413–424.

Sinha D. and Dey D.K. (1998) Survival analysis using semiparametric Bayesian methods, in: *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer-Verlag, 195–211.