Asian Journal of Probability and Statistics

4(2): 1-14, 2019; Article no.AJPAS.50053
ISSN: 2582-0230

Review of Bayesian Analysis in Additive Hazards Model

Enrique Ernesto, Alvarez[1,2] and Maximiliano Luis, Riddick[3*]

[1] *Departamento de Matematicas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET, Instituto de Calculo, Argentina.*

[2] *Facultad de Ingeniera, Universidad Nacional de La Plata, Argentina.*

[3] *Centro de Matemática de La Plata, Facultad de Ciencias Exactas, Universidad Nacional de La Plata - CONICET, Argentina.*

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2019/v4i230112
Editor(s):
(1) Dr. Manuel Alberto M. Ferreira, Professor, Department of Mathematics, ISTA-School of Technology and Architecture, Lisbon University, Portugal.
Reviewers:
(1) Olumide Adesina, Olabisi Onabanjo University, Nigeria.
(2) Alexandre Ripamonti, Ibirapuera University, Brazil.
Complete Peer review History: http://www.sdiarticle3.com/review-history/50053

Review Article

Received: 29 April 2019
Accepted: 08 July 2019
Published: 13 July 2019

Abstract

In Survival Analysis, the focus of interest is a time T^* until the occurrence of some event. A set of explanatory variables (denoted by a vector Z) is considered to analyze if there is a relationship between any of them and T^*. Accordingly, the "hazard function" is defined:

$$\lambda(t,z) := \lim_{\Delta \downarrow 0} \frac{P[T \leq t + \Delta | T > t, Z = z]}{\Delta}.$$

Several models are defined based on this, as is the case of the additive model (among others). Bayesian techniques allow to incorporate previous knowledge or presumption information about the parameters into the model. This area grows extensively since the computationally techniques increase, giving rise to powerful Markov Chain Monte Carlo (MCMC) methods, which allow to generate random samples from the desired distributions. The purpose of this article is to offer a summary of the research developed in Bayesian techniques to approach the additive hazard models.

**Corresponding author: E-mail: mriddick@mate.unlp.edu.ar;*

Keywords: Survival analysis; Bayesian inference; Additive hazards model.

2010 Mathematics Subject Classification: 53C25; 83C05; 57N16.

1 Introduction

In Survival Analysis, the focus of interest is a time T^* until the occurrence of some event. Given $f(t)$ and $F(t)$ (the density function and the cumulative density function respectively), other functions (survival, intensity, hazard and cumulative hazard) are defined according to them. Usually, the variable of interest is observed during a time in which the event of interest has not occurred. In those cases, the observations are called "right censored data". To include those cases in the model, another variable is considered: the "censoring variable" C. Then, the observed time is $T := \min(T^*, C)$, and a non-censoring indicator $\delta = I(T^* \leq C)$ discriminates between them. Finally, the data collected considering this model takes the form (t, δ). Next, a brief review of the theory is presented, along with the notation used in the sequel.

Definition 1.1. If T^* is continuous, the ***Survival function***

$$S(t) := Pr(T^* \geq t) = 1 - F(t). \tag{1.1}$$

Definition 1.2. The ***Hazard function*** or ***hazard rate function*** of a continuous variable T^* at time t is:

$$\lambda(t) = \lim_{\Delta \downarrow 0} \frac{Pr(t < T^* \leq t + \Delta | T^* > t)}{\Delta} = \lim_{\Delta \downarrow 0} \frac{\int_t^{t+\Delta} f(u)du}{\Delta} \Big/ Pr(T^* > t) = \frac{f(t)}{S(t)}. \tag{1.2}$$

Definition 1.3. The ***Cumulative Hazard function*** of a variable T^* is:

$$\Lambda(t) = \int_0^t \lambda(s)ds = -log[S(t)]. \tag{1.3}$$

Following that:

$$S(t) = \exp[-\Lambda(t)].$$

Definition 1.4. Given a counting process

$$N(t) = \text{"number of occurrences in } [0, t]\text{"},$$

the ***Intensity function*** defined according to this process is:

$$I(t) = Y(t)\lambda(t), \tag{1.4}$$

where we have introduced the "at risk" function

$$Y(t) = 1_{(T^* \wedge C > t)} = \begin{cases} 1, \text{ if the event doesn't occur until the instant } t, \\ 0, \text{ otherwise.} \end{cases} \tag{1.5}$$

The regression analysis in survival data focuses in the relationship between the time T^* and a set of explanatory variables called "covariates" (which is suspected to have an effect in the response variable). These covariates are denoted by a vector Z, and can be considered time dependent, in which case $Z = Z(t)$.

Many classes of models are developed in Survival Analysis. A wide class of survival models utilized are the semiparametric models (composed by a parametric and a non-parametric component). The justification of its popularity is its flexibility and its large range of applications, allowed specially by the non-parametric component. In this class, the most known and studied models are the

proportional hazard model [1], the additive hazard model [2], and the accelerated failure time model [3]. An expression for the hazard function for a fixed observation with covariate vector $z = (z_1, ..., z_p)$ and its corresponding associated regression vector $\beta = (\beta_1, ..., \beta_p)$ for the above three models are:

Proportional Hazards model (PHM):

$$\lambda(t|z) = \lambda_0(t)e^{z'\beta} \tag{1.6}$$

Additive Hazard model (AHM):

$$\lambda(t|z) = \lambda_0(t) + z'\beta \tag{1.7}$$

Accelerated Failure Time model (AFTM):

$$\lambda(t|z) = \lambda_0(te^{z'\beta})e^{z'\beta} \tag{1.8}$$

where λ_0 is the **baseline hazard function** associated with the non-parametric component of the model. Full text treatises on this type of models can be found in [4], [5], among others.

The parametric component ($\beta \in \mathbb{R}^p$) is associated with p regression parameters, while the non-parametric component ($\lambda_0(.)$) is very general, in that it is only required that λ_0 be nonnegative and that it integrates to infinity. An expression usually adopted for the non-parametric component is the denominated **"piecewise constant function"**, in which given $(J-1)$ fixed points in the time interval, the baseline hazard is expressed as follows:

$$\lambda_0(t) := \begin{cases} A_1 & s_0 = 0 \leq t < s_1, \\ A_2 & s_1 \leq t < s_2, \\ \vdots & \\ A_{J-1} & s_{J-2} \leq t < s_{J-1}, \\ A_J & t \geq s_{J-1}, \end{cases} \tag{1.9}$$

where $A_1, ..., A_{J-1} \geq 0$ and $A_J > 0$, which guarantees that $\int \lambda(t)dt = \infty$, implying that T^* is not-degenerate (in the sense that $P(T^* < \infty) = 1$). The grid $(s_1, ..., s_{J-1})$ (where each $s_j, j = 1, ..., J-1$ is a given constant) must be fixed before collecting the data.

Both models can be redefined as the intensity of a counting process $N(t)$, in which case, the expression for the AHM is given by the next expression:

$$I(t|z) = Y(t)\left[\lambda_0(t) + z'\beta\right]. \tag{1.10}$$

This expression including the intensity function must be interpreted like the probability that the event of interest takes place into the time interval $[t, t+dt)$, given the fact that it has not occurred previously [6] e.g.. Usual troubles to deal with in the AHM context are:

- the restriction imposed to the parametric component of the model ($z'\beta$) to ensure the nonnegativity of the hazard function,
- the mathematically unamenable expression that usually acquires the likelihood function.

A review of a frequentist development for this model was included in [7], and into a frequentist robust context a review of the three models was addressed in [8].

From a Bayesian perspective, the PHM has been the most studied model in Survival Analysis. A review of this is included in [9]. In the AFTM, a list of references for a Bayesian treatment is included in [10]. Nevertheless, the AHM has gained attention as an alternative to the PHM,

specially in the cases that the proportional hypothesis is not satisfied. Different models present different aspects in the relation between the response variable and the covariates.

Our objective in this manuscript is to review the Bayesian literature for the AHM, focusing in the complications and in the different approaches offered by different authors.

2 Bayesian Inference in the AHM

2.1 First Approaches

The first work published in this area was exhibited by [11].

A special case of the AHM was considered by the authors, denominated "gamma polygonal model", in which the hazard function takes the next expression:

$$\lambda(t|Z=z) = \lambda_0(t) + \lambda_1(t|Z=z). \tag{2.1}$$

As in the piecewise constant hazard model, the time is partitioned in different fixed points. According to this grid, the distinct baseline values take a linear expression, implying that the entire baseline shape is polygonal. The baseline considered is:

$$\lambda_0(t) = \begin{cases} A_j + \frac{(A_{j+1}-A_j)(t-s_{j-1})}{s_j - s_{j-1}}, \text{if } s_{j-1} \leq t \leq s_j, j = 1, ..., J-1, \\ A_J, \text{if } \ t \geq s_{J-1}. \end{cases} \tag{2.2}$$

The parametric component of the model $\lambda_1(t)$ is the hazard function of a *Gamma* distribution with parameters α and β, i.e,,

$$\lambda_1(t|Z=z) = \frac{t^{\alpha-1}\exp(-\beta t)}{\int_t^\infty u^{\alpha-1}\exp(-\beta u)du}, \text{ if } t > 0. \tag{2.3}$$

An assumption imposed is that α and β are specific for each individual in the population, but related to the covariates Z through a probabilistic model. Then, a hierarchical structure is considered in the model, with the second level of hierarchy given by:

$$\frac{\alpha}{\beta}|\beta, z \sim \mathcal{LN}\left(b'z, \sigma_\alpha^2\right), \tag{2.4}$$

$$\beta \sim \mathcal{LN}(\mu_\beta, \sigma_\beta^2), \tag{2.5}$$

where $\mathcal{LN}$ denotes the *log Normal* distribution. The hyperparameters $b, \sigma_\alpha^2, \mu_\beta$ and σ_β^2 are considered unknown constants common to all individuals in the population.

To perform a Bayesian analysis, the authors selected the next level of prior hierarchy for the parameters in the model (where $\mathcal{IG}$ denotes the *Inverse Gamma* distribution) by:

$$\mu_\beta|\sigma_\beta^2 \sim \mathcal{N}(m_\beta, v_\beta^2\sigma_\beta^2), \tag{2.6}$$

$$\sigma_\beta^2 \sim \mathcal{IG}(a_\beta, b_\beta), \tag{2.7}$$

$$b|\sigma_\alpha^2 \sim \mathcal{N}_p(m_\alpha, V_\alpha\sigma_\alpha^2), \tag{2.8}$$

$$\sigma_\alpha^2 \sim \mathcal{IG}(a_\alpha, b_\alpha). \tag{2.9}$$

For the parameters of the baseline, the authors selected an autocorrelated first order process:

$$A_j = A_{j-1}.\exp(\epsilon_j), j = 2, ..., J$$

where $(\epsilon_2, ..., \epsilon_j)$ are assumed independents, with $Normal$ distribution, mean 0 and variance σ^2_ϵ, where:

$$A_1 \sim Gamma(a_A, b_A), \tag{2.10}$$

$$\sigma^2_\epsilon \sim \mathcal{IG}(a_\epsilon, b_\epsilon). \tag{2.11}$$

The posterior obtained following this model becomes intractable. Nevertheless, the conditional posterior of $(\mu_\beta, \sigma^2_\beta)$, $(\mu_\alpha, \sigma^2_\alpha)$ and σ_ϵ are available and then a simulation study using Gibbs sampling [12] or another MCMC technique is possible. The other conditionals do not have a conjugate analysis, but it is still possible to sample from them using a Metropolis-Hastings algorithm [13, 14]. Then, through a Metropolis-within-Gibbs simulation, approximate samples from the full posterior could be obtained.

Concluding remarks about this model:

- the Gamma-polygonal allows certain heterogeneity in the population (an analogous could be achieved by introducing frailty terms into the model),
- the polygonal baseline hazard function is slightly more complex than the "piecewise constant baseline hazard" (in the computationally sense) and has the advantage that it is continuous,
- considering the parametric component of the model as a mixture of *Gamma* distributions avoids the restriction of the nonnegativity of $z'\beta \geq 0$,
- if the regression parameters are time independent, the parametric component is constant in t, an then the model becomes an exponential model.

Finally, the authors applied this results to an unemployment database, in which 559 questionnaires of undergraduate of "Valencia Education and Science Council" and "Valencia University" (Spain) during the years 1978 to 1993 are considered, in which the variable of interest is the time, in months, from graduation until the first employment.

Another approach was presented by [15], which has no close relation with [11]. The approach considered by these authors is different, and the model is not exactly the same. Motivated by the drawbacks of choosing a truncated prior to avoid the trouble of the nonnegativity of the hazard function, the authors redefined the model including specified indicators applied to the covariates, which are specified according the expected incidence of the covariate. This approach is the only one in the literature which takes a multinomial prior related to the expectation of the increment or decrement of each covariate. The indicator associated to the kth covariate are denoted M_k and may takes the values:

$$M_k = \begin{cases} -1, & \text{if the } k\text{th covariate is associated with hazards reduction,} \\ 0, & \text{if the } k\text{th covariate is not associated with any hazards change,} \\ 1, & \text{if the } k\text{th covariate is associated with hazards increase.} \end{cases} \tag{2.12}$$

The covariates vector z is "standardized" so that $|z_i| \leq 1, \forall i$. This standardization can be done for bounded covariates by subtracting the minimum possible value and dividing by the range. For unbounded covariates, the sample minimum and range can be used, potentially expanding the range to include interesting values of the predictors outside of the sample range in order to use the model for extrapolation and more flexible predictions. According to those modifications, the hazard function results in:

$$\lambda(t; z_i) = \lambda^*_0(t) + \sum_{k=1}^{p} \{1_{(M_k=-1)}(1 - z_{ik}) + 1_{(M_k=1)} z_{ik}\}.\beta^*_k = \lambda^*_0(t) + z_i^{*'}\beta^*, \tag{2.13}$$

where $z^*_{ik} = 1_{(M_k=-1)}(1 - z_{ik}) + 1_{(M_k=1)}z_{ik}$, $\lambda^*_0(t) = \lambda_0(t) - \sum_{k=1}^{p} 1_{(M_k=-1)}\beta^*_k$ and β^* denotes the vector with components $\beta^*_1, ..., \beta^*_k$. The likelihood considered by the authors is based on the intensity of a counting process $N(t)$ and it is defined by:

$$\prod_{i=1}^{n}\left(\prod_{t\neq 0}[Y_i(t)\{\lambda^*_0(t) + z_i^{*'}\beta^*\}]^{dN_i(t)}\right)\exp\left(-\int_{t\neq 0} Y_i(t)\{\lambda^*_0(t) + z_i^{*'}\beta^*\}dt\right). \tag{2.14}$$

where $dN_i(t)$ denote the increment of $N_i(t)$ over the small interval $[t, t + dt)$.

Since the increment is infinitesimal, the $dN_i(t)$ contributes to the likelihood in the same manner as independent Poisson random variables even though for all t, $dN_i(t) \leq 1$.

Calling $t_1, ..., t_{J-1}$ the unique failure times observed in data set, and considering the piecewise constant hazard model with grid defined by such time points (i.e. $s_j = t_j, j = 1, ..., J - 1$), the likelihood can be re-expressed as:

$$\prod_{i=1}^{n}\prod_{j=1}^{J-1}\left(\prod_{t\in(s_{j-1},s_j]}[Y_i(t)\{\lambda^*_0(t) + z_i^{*'}\beta^*\}]^{dN_i(t)}\right)$$
$$\exp\left(-\int_{t\in(s_{j-1},s_j]} Y_i(t)\{\lambda^*_0(t) + z_i^{*'}\beta^*\}dt\right). \tag{2.15}$$

Under the assumption that the risk accrued in the interval $(s_{j-1}, s_j]$ is small (i.e., that are no ties, and that the data set is not too small),

$$\forall i, j, \qquad \int_{s_{j-1}}^{s_j} Y_i(t)\{\lambda^*_0(t) + z_i^{*'}\beta^*\}dt \approx 0.$$

Then, the likelihood contribution across this interval for individuals at risk is approximately

$$\{d\Lambda_{0j} + z_i^{*'}\beta^*(s_j - s_{j-1})\}^{dN_{i,j}}\exp(-\{d\Lambda_{0j} + z_i^{*'}\beta^*(s_j - s_{j-1})\}),$$

where $d\Lambda_{0j} = \int_{s_j-1}^{s_j}\lambda^*_0(t)dt$ and $dN_{i,j} = 1$ if subject i fails at time s_j or 0 otherwise. Hence, an approximation to the likelihood expression is given by

$$\prod_{i=1}^{n}\prod_{j:Y_j=1}\{d\Lambda_{0j} + z_i^{*'}\beta^*(s_j - s_{j-1})\}^{dN_{i,j}}\exp(-\{d\Lambda_{0j} + z_i^{*'}\beta^*(s_j - s_{j-1})\}).$$

The priors chosen for the baseline hazard $\lambda^*_0(t)$, the model indicators $M = (M_1, ..., M_p)$, and the vector of regression coefficients β^*, are:

- independent *Gamma* priors are taken for each $A_j, j = 1, ..., J$ in the baseline hazard, which are assumed independent,
- independent Multinomial distributions are taken to the M_k's:

$$M_k = \begin{cases} -1, \text{ with probability } \theta_{k,-1}, \\ 0, \text{ with probability } \theta_{k,0}, \\ 1, \text{ with probability } \theta_{k,1}, \end{cases}$$

  where $k = 1, ..., p$, and $\theta_{k,-1} + \theta_{k,0} + \theta_{k,1} = 1$,
- independent *Gamma* distributions are considered for the regression parameters.

The hyperparameters of these priors are selected in a prior elicitation context (i.e. mathematical translation of expert knowledge). This is an important contribution of the authors in the Bayesian context. The model indicators index the direction and occurrence of effects for the different covariates, but do not quantify the magnitude of the effect. The slopes are measured by the regression coefficients.

To simplify efficient posterior computation, a data augmentation approach is utilized. Under the assumed hypothesis, the dN_{ij} are independent $Poisson$ random variables:

$$dN_{ij} \overset{ind}{\sim} Poisson\left(d\Lambda_{0j} + \sum_{k=1}^{p} z_{ik}^{*'}\beta_k^*\right), \quad \forall i,j : Y_{ij} = 1. \tag{2.16}$$

This expression results in a non standard set of full conditional posterior distributions. Because of that, independent $Poisson$ latent variables are included in the model, obtaining:

$$dN_{i,j} = dM_{ij0} + \sum_{k=1}^{p} I_{M_k \neq 0} dN_{ijk}, \quad \forall i,j : Y_{ij} = 1, \tag{2.17}$$

where

$$\tau(dN_{ij0}) = Poisson(dN_{ij0}; d\Lambda_{0j}), \tag{2.18}$$

$$\tau(dN_{ijk}) = Poisson(dN_{ijk}; (s_j - s_{j-1}) z_{ik}^* \beta_k). \tag{2.19}$$

Using the property that the sum of independent $Poisson$ variable is $Poisson$, it is straightforward that integrating out the latent variables is equivalent to the previous form (2.16). This formulation allows to take advantage of $Poisson - Gamma$ conjugacy to obtain simple conditional posterior distributions.

At this point, Gibbs sampling allows to generate samples of any full conditional posterior, and then from the full posteriori itself.

The authors generalized this issues to allow a Bayesian analysis in a more general model, denoted by Additive-Multiplicative Hazard Models, with hazard function equal to:

$$\lambda(t,z) = \lambda_0(t)\exp(z'\alpha) + z'\beta, \tag{2.20}$$

where $\alpha = (\alpha_1, ..., \alpha_p)'$ are proportional hazards coefficients, and the remaining parameters defined as previously. This model is more general than the additive and the multiplicative models, including both. If $\alpha_i = 0 \quad \forall i$, the additive model is obtained, and if $\beta_i = 0 \quad \forall i$, then the expression results in the proportional model. For notational convenience, the model is defined so that the same covariates are included in the proportional and in the additive components. The prior is adapted to accommodate the multiplicative components and then generalize the Gibbs sampler accordingly.

In choosing the prior for $(\lambda_0, M, \alpha, \beta)$ (independence is assumed):

$$\tau(\lambda_0, M, \alpha, \beta) = \tau(\lambda_0)\tau(M,\beta)\tau(\alpha). \tag{2.21}$$

The priors $\tau(\lambda_0)$ and $\tau(M,\beta)$ chosen are the same as in the additive model. To induce a prior on the proportional hazards regression coefficients $\tau(\alpha)$, the authors consider a prior defined in a previous article by themselves to realize Bayesian inference on the Cox model, denoted "one-inflated truncated gamma density".

In a Simulation study section, it is verified that the method utilized works properly and can discriminate between the additive and the multiplicative components, showing that the approach has good frequentist properties. In a Data Example section, the authors carried out a study on

coronary heart disease on a data base of 1571 individuals, who were disease free at roughly age 45. It focused on age at onset of coronary heart disease in female subjects in relation to certain covariates (hypertension, above normal cholesterol levels, overweight status and obesity). The study reveals that it cannot concluded that any of the covariates has only additive or multiplicative effects. Following that, the additive-multiplicative approach seems to be the most reasonable.

A review of the previous commented articles was included in [16], who suggest a new alternatives:

- In the Gamma-polygonal model [11], he propose as an alternative to the baseline hazard, a "piecewise constant" model, and then a *Gamma* process prior for the baseline parameters.
- In the model proposed by [15], a *Gamma* process is suggested as prior to the baseline increments. Code in WinBUGS is supplied according to this model. Then, a data analysis is performed for melanoma and lung cancer databases.

2.2 Frailty Issues

The pioneers in including frailties in the model were [17].

The inclusion of frailties is motivated because the survival analysis requires the assumption that the individuals are independent given the covariates, and that the heterogeneity is explained in terms of the observed covariates. These underlying assumptions may not be satisfied due to the influence of unobserved heterogeneity among individuals in the study, and to overcome these potential limitations, a random effect (frailty) is inserted into the hazard function, giving rise to the frailty survival models. The insertion of the frailty w ($w > 0$) is usually done in a multiplicative way (i.e. the hazard function was multiplied by w or a function of w (like e^w)). In this work, the frailties are incorporated in an additive way to Aalen's model, where the data are considered sub-divided in k groups, including then the vector of frailties $w = (w_1, w_2, ..., w_k)'$ to the model, and then proceed to do a Bayesian analysis for the "additive survival model with frailty".

According to the incorporation of the frailties to the model, the expression of the intensity function results in:

$$I_i(t|z_i, w_{l_i}) = Y_i(t)\left[\lambda_0(t) + z_i'.\beta(t) + w_{l_i}\right], i = 1, \cdots, p, \tag{2.22}$$

where $N_i(t)$ is the number of occurrences of a particular event at time t, $l_i \in \{1, ..., k\}$ and $Y_i(t)$ take values 1 or 0 whether the individual i is at risk at time t, respectively. *Gamma* priors has been considered for the frailties $w_1, w_2, ..., w_k$, which are assumed i.i.d. According to this, the joint frailty distribution is given by:

$$[\Gamma(a)]^{-k} b^{k.a} \left(\prod_{l=1}^{k} w_l\right)^{a-1} \exp\left(-b\sum_{l=1}^{k} w_l\right),$$

where a and b are the shape and scale parameters, respectively. Those frailty parameters must also been estimated from the data. Unit mean for the frailties has often been assumed to avoid problems with identifiability of multiplicative frailty models. The variability of the frailties can be interpreted as the degree of heterogeneity among individuals. The authors considered time dependent covariates, contributing with extra flexibility to the model. The same idea of [15] and [16] was utilized in the prior selection for the baseline, and in the same manner, a cumulative regression function is defined to each time dependent covariate, which allows to incorporate an increasing process to them.

To model the baseline hazard function $\Lambda_0(t)$, the piecewise-constant hazard is adopted. Time dependent covariates are considered in this work, and "cumulative regression functions" are defined

according to them and to the selected grid to the baseline, as:

$$\Omega_q(t) = \int_0^t \beta_q(u)du, t \geq 0, q = 1, ..., p. \tag{2.23}$$

Independent-increment *Gamma* processes are assumed as prior for the baseline piecewise-constant parameters and the cumulative regression functions. Thus, for each $\Lambda_{0j}, \Omega_{qj}$ (wherein this parameters denote the increments of Λ_0 and $\Omega_q, q = 1, ..., p$ respectively in the interval $(s_{j-1}, s_j]$), a *Gamma* distribution with shape and scale parameters $c_0\Lambda_0^*$ and c_0 is assigned for Λ_0, and *Gamma* distribution with shape and scale parameters $c_q\Omega_q^*$ and c_q are assigned for each Ω_q.

According to this prior selection, the result posterior is proportional to:

$$\prod_{j=1}^{J-1}\left[\prod_{i\in R_j}(I_{ij}^{N_{ij}}\exp(-I_{ij}))\Lambda_{0j}^{c_0\Lambda_{0j}^*-1}\exp(-c_0\Lambda_{0j})\right.$$
$$\left.\prod_{q=1}^{p}\left(\Omega_{qj}^{c_q\Omega_{qj}^*-1}\exp(-c_q\Omega_{qj})\right)\right]\tau(w|\delta)\tau(\delta), \tag{2.24}$$

where $ds_j = s_j - s_{j-1}, I_{ij} = I_i(s_j)ds_j = Y_{ij}(\Lambda_{0j} + \beta_i'\Omega_j + w_{l_i}ds_j), Y_{ij} = Y_i(s_j)$, $\Omega_j = (\Omega_{1j}, ..., \Omega_{pj})$ and $N_{ij} = dN_i(s_j)$. R_j is the risk set in the interval $(s_{j-1}, s_j]$, $i = 1, ..., n$, $j = 1, .., J-1$, $q = 1, ..., p$, and $l_i \in \{1, ..., k\}$. The parameters of interest are $(\Lambda_0, \Omega, a, b)$, whereas the frailties w are taken as nuisance parameters that are eliminated by integration giving rise to a joint marginal posterior. This posterior expression is awkward to work. Nevertheless, the marginal posterior distributions of Λ_0, Ω, a and b can be evaluated using MCMC methods. Gibbs sampling is available to generate random samples of the joint posterior distribution.

In this article, a "Model Assessment" section is included. Both model comparison and model adequacy are important issues in survival analysis. An extensive review of Bayesian methods for studying these issues is detailed in [18], such as the Bayesian information criterion and the conditional predictive ordinate (CPO) for model comparison. The authors present a way to estimate the CPO in this model, and then it is possible calculate any measure for model comparison (for example the pseudo-Bayes factor).

In an Illustration Section, an application of this method was performed on real data sets: the first of them are 90 male patients with larynx cancer, and the second is referred to 42 leukemia patients.

In another work [19], the results obtained previously are extended to a "shared frailty" model. In this case, the frailty term w for each individual may be partitioned into two or more terms $w = w_1 + w_2 + ... + w_k$ (where w_j are frailty terms shared with other individuals, $j = 1, ..., k$) in order to assess various types of frailties within the same individual. According the new expression of the frailties, an indicator vector $a_i = (a_{i1}, a_{i2}, ..., a_{ik})'$ is considered for the individual i, in which each $a_{ij}, j = 1, ..., k$ indicates if the corresponding frailty term w_j is present or not. The expression resulting after include this modifications is:

$$I_i(t|z_i, w) = Y_i(t)\left[\lambda_0(t) + z_i'\beta(t) + a_i'w\right]. \tag{2.25}$$

The Bayesian treatment proposed for this model is analogous to the previous work, keeping the selection of the prior distributions of the parameters.

An application of this model is presented as an illustration to an adoption database of 125 families with adoptive children, aiming to compare the association of the intensity of death by infection (pneumonia) between genetic and environmental factors, concluding that the environmental conditions are the most important.

Another approach in area of Bayesian estimation in the AHM with frailty was exhibited in his Master thesis by [20]. In this work, spatial frailties are included, independent *Gamma* priors are considered to the different piecewise constant baseline parameters, and an *improper uniform* prior is selected to the regression parameters. Both covariates and frailties are considered time dependent, and the objective is estimate them according a grid, with a piecewise constant function, in the same manner than the baseline. The frailties are assumed *Gaussian* with covariance structures of either geostatistical or conditional autoregressive (CAR) type, two well known spatial dependence structures.

In the CAR model, the author develops a fully Bayesian method to estimate the parameters of the model. The frailties are assumed independent, allowing the introduction of spatial correlation in each time interval independently. The prior selected for each frailty component is $Normal$, where the variance parameters are assumed $Inverse\ Gamma$. This procedure to sample from the full conditionals of the baseline hazard, the regression and the frailty parameters relies on Metropolis-within-Gibbs simulations. Accordingly, an extensive study on the Metropolis step is done in order to obtain an efficient "proposal density". Those proposals are considered for each conditional, and the hyperparameters are selected so that the proposal and the conditional density are as similar as possible. Two methods are developed: using the mean of the full conditional, and finding the mode of the full conditional. In the first case, the normalizing constants are required, and the hyperparameters of the proposal are defined aiming at means equality. In the other method (suggested as more numerically efficient), the hyperparameters are selected using the mode of the full conditional, in order to obtain a better approach. The method is tested trough a simulation study, and then is applied to a Prostate cancer database.

With the same procedure, sampling techniques are available to sample from almost all the full conditionals of the geostatistical model. In this case, the prior considered for the frailty components is a $Multivariate\ Normal$ distribution with mean zero and the variance-covariance matrix depending on the distances between locations. In this case, priors are assigned to the variance and correlation parameters of the variance-covariance matrix. An $Inverse\ Gamma$ prior is assigned to the variance parameters, and *Gamma* priors are utilized for the correlation parameters, whit hyperparameters chosen according how strong is the belief in the priors selected. The $Inverse\ Gamma$ prior is chosen in order to make it conjugate. However, for the correlation parameters there is no available proposal. The author leaves the procedure of obtaining the suitable proposal for future research. Recently, [21] presents a similar analysis of the CAR model.

2.3 Other Proposals

In a different context, an empirical Bayesian treatment was presented in [22]. In this work, empirical estimators were developed for the regression parameters, survival curves and their correspo-nding standard deviations. Those estimators have the advantages of being easy to generate computationally, and not requiring elicitation of the hyperparameters. The method guarantees a monotone estimator for the survival curve, and can be extended easily to time dependent covariates. The priors selected are a Gamma process for the cumulative baseline hazard increments (which are supposed piecewise constant), and an improper uniform for the regression parameters. The empirical Bayes estimators of the hyperparameter vector associated to the prior process assigned to $\Lambda_0(t)$ and the regression parameters, are obtained by maximizing the likelihood defined by the model (which is proportional to a product of Poisson distributions). The posterior obtained for $\Lambda_0(t)$ conditioned on the rest of the parameters and the data does not have any standard form. However, the authors find a closed-form expression for the Laplace transform, so that the posterior moments of $\Lambda_0(t)$ and $S_0(t)$ can be evaluated.

That procedure is applied to a prospective study of 205 melanoma patients, collected between years

1962-1977. In this analysis, the only covariate chosen is gender, and the results are compared with the full Bayes estimator [15], and the method of moments estimator [23]. An improper noninformative prior process is used for the full Bayes method, suggesting that the full Bayes method is appropriate when accurate prior information is available (otherwise, the empirical Bayes method fits better). In a simulation study, the authors conclude that it is preferable to use empirical Bayes methods when the true form of λ_0 is in doubt. The full Bayes estimator presents a high level of relative bias compared with the empirical Bayes and moment estimators, especially with a moderate censoring level and noninformative priors. Additionally, extension to time dependent covariates and asymptotic properties are included in another sections.

A new class of models was considered by [24]. The authors propose a class of transformation hazard models for right censoring data. It includes the additive and proportional hazards models as special cases.

That class of transformations are based on the Box-Cox transformation [25], and takes the next expression:

$$\phi(Y) = \begin{cases} \frac{(Y^\gamma - 1)}{\gamma}, & \text{if } \gamma \neq 0, \\ \log(Y), & \text{if } \gamma = 0, \end{cases} \tag{2.26}$$

where

$$\lim_{\gamma \downarrow 0} \frac{(Y^\gamma - 1)}{\gamma} = \log(Y). \tag{2.27}$$

Within a survival semiparametric model with baseline hazard λ_0, the transformation proposed is:

$$\phi\left(\lambda(t|z_i)\right) = \phi\left(\lambda_0(t)\right) + z_i(t)'\beta, \tag{2.28}$$

where ϕ is the link function defined by the Box-Cox transformation.

The additive and multiplicative model are included into this class of transformations (the additive case when $\gamma = 1$, and the multiplicative case when $\gamma = 0$).

The first goal is to model selection on γ, by fitting different models for each value of γ and evaluating them through a model selection criterion. The value of γ that accurate a better fit is then selected. After that, it proceed to the estimation of rest of the parameters of the model. The piecewise constant model is chosen for $\lambda_0(t)$, denoting λ_0 the Jth dimensional vector $(A_1, ..., A_J)'$. In prior selection, the authors deal with the nonnegativity of the hazard restriction. To avoid this trouble, a truncated prior is selected. Following that, a truncated $Multivariate\ Normal$ distribution is selected for $(\beta|\lambda_0)$. According to this, the joint distribution of the priors is:

$$\tau(\beta, \lambda_0) = \tau(\beta|\lambda_0)\tau(\lambda_0)I(A_j^\gamma + \gamma Z_i'\beta \geq 0), \forall i, j. \tag{2.29}$$

In order to obtain the full conditional distribution of λ_0, the normalizing constant must be calculated, that involves a multivariate integral in an unwieldy nonlinear constrained parameter space. To circumvent this problem, the prior selected is modified, aiming at reducing the problem to the univariate case. This procedure is an innovate technique worthy to consider when a truncated multivariate distribution is chosen to avoid the nonnegativity restriction of the hazard. Denoting $Z_{i(-k)}$ and $\beta_{(-k)}$ the vectors Z_i and β with the kth component removed, the authors define:

$$h_\gamma(\lambda_0, \beta_{(-k)}, Z) = \min_{i,j} \left\{ \frac{A_j^\gamma + \gamma Z'_{i(-k)} . \beta_{(-k)}}{\gamma Z_{ik}} \right\}. \tag{2.30}$$

Then, a joint prior for (β, λ_0) is considered, of the form:

$$\tau(\beta, \lambda_0) = \tau(\beta_k|\beta_{(-k)}, \lambda_0) I\left(\beta_k \geq -h_\gamma(\lambda_0, \beta_{(-k)}, Z)\right). \tag{2.31}$$

This prior specification only include the restriction parameter β_k, leaving all the other parameters free.

It is usual (but not necessary) to consider $\beta_{(-k)}$ and λ_0 to be independent a priori, obtaining $\tau(\beta_{(-k)}, \lambda_0) = \tau(\beta_{(-k)})\tau(\lambda_0)$. The authors assume that the components of λ_0 are independent a priori with *Gamma* distributions. Further, a normal prior distribution is chosen for each component of $\beta_{(-k)}$. Finally, sampling from the posterior is available through a Metropolis-within-Gibbs algorithm.

The authors also presents a Model assessment section, which presents the DIC (Deviance Information Criterion) and CPO statistics according to the model.

In other articles, extensions of this model are presented, including frailties [26, 27], and adapting this issue to "cure fractions" models [28]. Several simulations and applications to real data are presented. They all show that the best fitting model correspond to some $0 < \gamma < 1$, justifying the relevance of this approach.

Currently, we are undertaking research for the AHM from a hybrid Bayesian approach which combines the classical estimating equation method of [23] with Bayesian priors.

3 Discussion

Research has been developed by many authors on Bayesian inference for the additive hazards model. The purpose of this article was to review them, showing the reader the current state of the art in this area. The advances in MCMC methods and the increase computational power gave rise to widespread Bayesian techniques, which allow to sample from distributions that were intractable before. The proposed methods are flexible in the choice of priors. The baseline hazard function is either model parametrically or nonparametrically as a piecewise constant function. In the later case, independent priors for the piecewise constants are usually chosen. Because of that approach, most of the literature omit modeling certain aspects of the baseline function that could be of interest (e.g., unimodality, concavity, monotonicity, conexity, etc). Rare exceptions are proposal for a gamma process for the baseline functions (e.g. [16], [17]) or the introduction of an AR process for the piecewise constant (e.g. [11]). This is a promising area for future research. Also, from the computing point of view, there is a fair amount of research to be done in the search to better algorithms.

Acknowledgement

The authors are grateful to the referees for their careful reading, constructive criticisms, comments and suggestions, which have helped us to improve this work significantly. M.L.R. thanks CONICET for the doctoral fellowship. Both authors thanks to Universidad Nacional de La Plata (PPID UNLP I231).

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Co DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological). 1972;34(2):187-220.

[2] Aalen Odd. A Model for nonparametric regression analysis of counting processes. Mathematical Statistics and Probability Theory. 1980;1-25.

[3] Rupert G. Miller. Least squares regression with censored data. Biometrika. 1976;63(3):449-464.

[4] Kalbfleisch JD, Prentice RL. The statistical analysis of time failure data. Hoboken: John Wiley & Sons; 1980.

[5] Fleming, Thomas R, Harrington, David P. Counting processes and survival analysis. John Wiley & Sons. 2011.

[6] Andersen PK, Borgan O, Gill RD, Keiding N. Statistical models based on counting processes. Springer; 1993.

[7] Martinussen, Torben, Scheike, Thomas H. Dynamic regression models for survival data. Springer; 2006.

[8] Álvarez, Enrique Ernesto, Ferrario, Julieta. Review of robust estimation in models Suriparam ectricas of survival. Journal of the Interamerican Statistical Institute. 2012;64(182,183):85-106.

[9] Sinha, Debajyoti, Dey Dipak K. Semiparametric bayesian analysis of survival data. Journal of the American Statistical Association. 1997;92(439):1195-1212.

[10] Jiajia Zhang, Andrew B. Lawson. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. Journal of Applied Statistics. 2011;38(3):591-603.

[11] Beamonte, Eduardo, Bermúdez, Jos D. A bayesian semiparametric analysis for additive hazard models with censored observations. Test. 2003;12(2):347-363.

[12] Geman, Stuart, Geman, Donald. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 1984;6(6):721-741.

[13] Metropolis, Nicholas, Rosenbluth, Arianna W, Rosenbluth, Marshall N, Teller, Augusta H, Teller, Edward. Equation of state calculations by fast computing machines. The Journal of Chemical Physics. 1953;21(6):1087-1092.

[14] Hastings WK. Monte carlo sampling methods using markov chains and their applications. BIOMETRIKA. 1970;57(1):97-109.

[15] Dunson, David B, Herring, Amy H. Bayesian model selection and averaging in additive and proportional hazards models. Lifetime Data Analysis. 2005;11(2):213-232.

[16] Saad AAM. Bayesian non-parametric estimation in proportional hazard models: Survival data analysis with winbugs. LAP Lambert Academic Publishing; 2012.

[17] Silva, Giovani L, Amaral-Turkman, Antnia M. Bayesian analysis of an additive survival model with frailty. Communications in Statistics - Theory and Methods. 2005;33(10):2517-2533.

[18] Ibrahim, Joseph G, Chen, Ming-Hui, Sinha Debajyoti. Bayesian survival analysis. Springer Science & Business Media; 2001.

[19] Silva, Giovani L, Amaral-Turkman, Antnia M. Additive survival models with shared frailty. Proceedings of the 22nd International Workshop on Statistical Modelling. 2007;544-548.

[20] Chernoukhov, Alexander. Bayesian spatial additive hazard model. Windsor University; 2013.

[21] Chernoukhov A, Hussein A, Nkurunziza S, Bandyopadhyay D. Bayesian inference in time-varying additive hazards models with applications to disease mapping. Environmetrics. 2018;29(5-6):e2478.

[22] Sinha, Debajyoti, McHenry, M Brent, Lipsitz, Stuart R, Ghosh, Malay. Empirical Bayes estimation for additive hazards regression models. Biometrika. 2009;96(3):545-558.

[23] Lin DY, Ying Zhiliang. Semiparametric analysis of the additive risk model. Biometrika. 1994;81(1):61-71.

[24] Yin Gousheng, Ibrahim, Joseph G, others. Bayesian transformation hazard models. Optimality. 2006;170-182.

[25] Box GEP, Cox, DR. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological). 1964;26(2):211-252.

[26] Yin, Guosheng, Ibrahim, Joseph G. Bayesian frailty models based on box-cox transformed hazards. Statistica Sinica. 2005;781-794.

[27] Yin, Guosheng, Ibrahim, Joseph G. A class of bayesian shared gamma frailty models with multivariate failure time data. Biometrics. 2005;61(1):208-216.

[28] Yin, Guosheng, Ibrahim, Joseph G. A general class of bayesian survival models with zero and nonzero cure fractions. Biometrics. 2005;61(2):403-412.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle3.com/review-history/50053