Journal of the Korean Statistical Society (2002), **31**: *4, pp 509–518*

Semiparametric Bayesian Regression Model for Multiple Event Time Data†

Yongdai Kim[1]

ABSTRACT

This paper is concerned with semiparametric Bayesian analysis of the proportional intensity regression model of the Poisson process for multiple event time data. A nonparametric prior distribution is put on the baseline cumulative intensity function and a usual parametric prior distribution is given to the regression parameter. Also we allow heterogeneity among the intensity processes in different subjects by using unobserved random frailty components. Gibbs sampling approach with the Metropolis-Hastings algorithm is used to explore the posterior distributions. Finally, the results are applied to a real data set.

Keywords. Frailty, Gibbs sampling, Poisson process, proportional intensity.
AMS 2000 subject classifications. Primary 67C10; Secondary 62N01.

1. Introduction

In many medical studies, we frequently meet a data set in which a subject experiences multiple events. An example can be found in a bladder cancer study (Byar, 1980) conducted by the Veterans Administration Cooperative Urological Research Group. In this study, all patients had superficial bladder tumors when they entered the trial. These tumors were removed transurethrally and patients were randomly assigned to one of three treatments: placebo, thiotepa and pyridoxine. One feature of this data set is that many patients had multiple recurrences of tumors during the study.

Several regression models have been proposed in the literature to deal with situations where individuals experience repeated failures such as multiple tumor recurrences. Oakes (1992) reviews several models of multiple event time data

Received December 2001; accepted October 2002.
†This work was supported by Korea Research Foundation Grant (KRF 2001-041-D00045).
[1]Department of Statistics, Ewha Womans University, Seoul 120-750, Korea

including Andersen and Gill (1982) and Lagakas *et al.* (1978). In this paper, we propose a Bayesian semiparametric regression model for multiple event data. Our model is based on the Andersen-Gill model for the Poisson process, in which covariates enter in a proportional intensity fashion. Also, unobserved random frailty is added into the model to allow events coming from a same patient to be correlated.

The parameters to be estimated in the model are the regression coefficients and the baseline cumulative intensity function. We give a nonparametric prior distribution on the baseline cumulative intensity function regarding it as a realization of a gamma process (Kalbfleish, 1978; Lo, 1982), and give usual parametric prior distributions to the regression coefficients. Gibbs sampling algorithm is presented to explore the posterior distributions of the regression coefficients as well as the baseline cumulative intensity function.

A Bayesian model for the Poisson process without covariates was considered by Lo (1982). Sinha (1993) proposed a Bayesian semiparametric model for discrete time scales. In this view, our model provides unification and extension of the previously considered Bayesian models for multiple event time data, for our model is time continuous and allows having covariates. Slightly different Bayesian models for counting processes are considered by Kim (1999).

The paper is organized as follows. Section 2 describes the Bayesian model for multiple event time data in detail. Section 3 describes the Gibbs sampling method to generate samples from the posterior distributions. Section 4 exemplifies the methodology with the data set of bladder tumors. Discussion is presented in Section 5.

2. The Model

Let $N_1, \ldots, N_n$ be independent counting processes which record the event history of n patients. That is, $N_i(t)$ counts the number of events in the i^{th} patients up to time t. We assume that N_i is observed continuously on time interval $[0, \tau_i]$ where $\tau_i > 0$. For given covariates $Z_i \in R^p$, N_i is assumed to be a Poisson process with intensity Λ_i where

$$\Lambda_i(t) = \int_0^t w_i Y_i(s) \exp(\beta Z_i) d\Lambda(s), \tag{2.1}$$

$Y_i(s) = I(s \leq \tau_i)$ and $\Lambda(t)$ is the baseline intensity function. In other words, N_i is a process with independent increments such that $N_i(t) - N_i(s)$ is distributed according to the Poisson distribution with mean $\Lambda_i(t) - \Lambda_i(s)$ for all $s < t$. In

(2.1), w_i is a multiplicative random frailty on the intensity function associated with the i^{th} individual. It is assumed that $w_1, \ldots, w_n$ are *iid* gamma variates with mean unity and unknown variance α. The assumption that the frailties are gamma variates follows previous research on unobserved heterogeneity that makes wide use of this distribution (Clayton, 1978; Vaupel *et al.*, 1979; Sinha, 1993). The advantages of the gamma distribution are its flexible shape and its analytical tractability.

We use the random frailties $w_1, \ldots, w_n$ to accommodate heterogeneity among the intensity functions in different subjects. In multiple event time data, it is frequently observed that a subject vulnerable to a specific risk experiences more frequent events than a subject immune to the risk. That is, a certain subject can have a larger intensity function than others and *vice versa*. This heterogeneity is modelled by the frailties nicely. We can control this heterogeneity of the risks by use of the frailty terms. However, the population average of the frailties is 1, which means that the overall risk of the event in time is represented by the baseline intensity function. The variance of the frailties represents the degree of heterogeneity of the risk among subjects.

Let $\pi(\beta)$ and $\kappa(\alpha)$ be prior distributions of β and α respectively. For a prior distribution of the unknown baseline cumulative intensity function $\Lambda(t)$, we use a gamma process with mean $\Lambda_0(t)$ and scale parameter $c(t) > 0$. Let $\{t_k\}$ be the countable set of points of discontinuity of $\Lambda_0(t)$, that is, $\Delta\Lambda_0(t_k) = \Lambda_0(t_k) - \Lambda_0(t_k-) > 0$, and let Λ_{0c} be the continuous part of Λ_0, that is,

$$\Lambda_{0c}(t) = \Lambda_0(t) - \sum_{t_k \le t} \Delta\Lambda_0(t_k).$$

Then the gamma process $\Lambda(t)$ with parameters $(\Lambda_0(t), c(t))$ is defined to be a Lévy process such that

$$\log E(\exp(-\theta\Lambda_0(t))) = \int_0^t \int_0^\infty (e^{-\theta x} - 1)\frac{c(s)}{x} \exp\{-c(s)x\} dx d\Lambda_{0c}(s) + \sum_{t_k \le t} \log E\{\exp(-\theta S_k)\},$$

where S_k are independent random variables distributed according to $G(c(t_k) \times \Delta\Lambda_0(t_k), c(t_k))$. Here $G(a, b)$ represents a gamma distribution with mean a/b. If $c(t)$ is a positive constant (*i.e.* $c(t) \equiv c > 0$), then $\Lambda(t) - \Lambda(s)$ is distributed according to $G(c(\Lambda_0(t) - \Lambda_0(s)), c)$. More details about gamma processes can be found in Basawa and Prakasa Rao (1980) and Lo (1982).

3. Posterior Distribution : Gibbs Sampling

This section describes the Gibbs sampling algorithm for generating samples from the joint posterior distribution of the parameters in the model. The Gibbs sampling is a Markov chain Monte Carlo method, which generates samples from the conditional distributions successively. See Brook (1998) and references therein for details of Markov chain Monte Carlo methods.

In our model, three parameters exist: Λ, β and α. Also we consider $\mathbf{w} = (w_1, \ldots, w_n)$ as a parameter becasue doing so makes the Gibbs samling algorithm much simpler. Hence our Gibbs sampling consists of generating samples successively from the four conditional distributions: $\mathcal{L}(\Lambda|\beta, \alpha, \mathbf{w}, \mathbf{N})$, $\mathcal{L}(\beta|\alpha, \Lambda, \mathbf{w}, \mathbf{N})$, $\mathcal{L}(\alpha|\beta, \Lambda, \mathbf{w}, \mathbf{N})$ and $\mathcal{L}(\mathbf{w}|\beta, \alpha, \Lambda, \mathbf{N})$ where $\mathbf{N} = (N_1, \ldots, N_n)$. Here, $\mathcal{L}(X|Y)$ denotes the conditional distribution of X given Y.

Assume that Λ_0 is continuous and $c(t) \equiv c > 0$, *a priori*. Then, the likelihood function of $(\beta, \alpha, \Lambda, \mathbf{w})$ given $\mathbf{N}$ is given by

$$\begin{aligned} L(\beta, \alpha, \Lambda, \mathbf{w}|\mathbf{N}) \propto & \prod_{i=1}^{n} \prod_{t\in[0,\tau^*]} \left\{ w_i Y_i(t) \exp(\beta Z_i) d\Lambda(t) \right\}^{dN_i(t)} \\ & \times \exp\left(\int_0^{\tau^*} w_i Y_i(s) \exp(\beta Z_i) d\Lambda(s) \right) \\ & \times \prod_{i=1}^{n} \frac{\alpha^\alpha}{\Gamma(\alpha)} w_i^{\alpha-1} \exp(-\alpha w_i), \end{aligned} \tag{3.1}$$

where $\tau^* = \max\{\tau_i\}$.

First, consider $\mathcal{L}(\Lambda|\beta, \alpha, \mathbf{w}, \mathbf{N})$. In Appendix, we prove that the posterior distribution of Λ given $\beta, \alpha, \mathbf{w}$ and $\mathbf{N}$ is again a gamma process with mean Λ^p and scale parameter $c^p(t)$, where

$$\Lambda^p(t) = \int_0^t \frac{d\Lambda_0(s) + dN_\cdot(s)}{c + R(s:\beta)} \tag{3.2}$$

and

$$c^p(t) = c + R(t:\beta). \tag{3.3}$$

Here $N_\cdot = \sum_{i=1}^n N_i$ and $R(t:\beta) = \sum_{i=1}^n w_i Y_i(t) \exp(\beta Z_i)$. Generating a sample path from the gamma process is not straightforward since it has infinitely many jumps with probability one. However, it will turn out soon that it suffices to generate at most $n + N_\cdot(\tau^*)$ many gamma variables to implement our Gibbs sampling algorithm.

For $\mathcal{L}(\beta|\alpha, \Lambda, \mathbf{w}, \mathbf{N})$, we can write

$$\begin{aligned}&\mathcal{L}(\beta|\alpha, \Lambda, \mathbf{w}, \mathbf{N})\\ &\propto \prod_{i=1}^{n} \prod_{t\in[0,\tau^*]} \exp(\beta Z_i)^{dN_i(t)} \exp\Big\{w_i Y_i(t) \exp(\beta Z_i) d\Lambda(t)\Big\}\pi(\beta).\end{aligned} \tag{3.4}$$

Let $0 < t_1 < t_2 < \ldots < t_k$ be ordered times at which $N.$ makes jump and let $0 < s_1 < s_2 < \ldots < s_m$ be ordered distinct times among $\{t_1, \ldots, t_k\} \cup \{\tau_1, \ldots, \tau_n\}$. Then (3.4) can be rewritten by

$$\begin{aligned}\mathcal{L}(\beta|\alpha, \Lambda, \mathbf{w}, \mathbf{N}) \propto \prod_{i=1}^{m} \left(\prod_{j=1}^{n} I(dN_j(s_i) > 0) \exp(\beta Z_j) \right)^{dN.(s_i)}\\ \times \exp\Big\{(\Lambda(s_i) - \Lambda(s_{i-1}))R(s_i : \beta)\Big\}\pi(\beta).\end{aligned} \tag{3.5}$$

β can be generated from (3.5) by use of the Metropolis-Hastings algorithm.

Next, it is easy to see that conditional on β, α, Λ and $\mathbf{N}$, $w_1, \ldots, w_n$ are independent random variables distributed according to $G(N_i(\tau_i) + \alpha^{-1}, E_i + \alpha^{-1})$, where

$$E_i = \sum_{i=1}^{m} \Big\{\Lambda(s_i) - \Lambda(s_{i-1})\Big\} \exp(\beta Z_i) Y_i(s_i).$$

Finally, we have

$$\mathcal{L}(\alpha|\beta, \Lambda, \mathbf{w}, \mathbf{N}) \propto \frac{\nu^{n\nu}}{\Gamma(\nu)^n} \exp\Big\{ -\nu \sum_{i=1}^{n} (w_i - \log w_i)\Big\}\kappa(1/\nu), \tag{3.6}$$

where $\nu = 1/\alpha$. Now, α can be generated from its conditional posterior distribution (3.6) via the Metropolis-Hastings algorithm.

It should be noted that we only need to generate $\Lambda(s_i) - \Lambda(s_{i-1})$ and $d\Lambda(s_i)$ only if $dN.(s_i) > 0$ for $i = 1, \ldots, m$. Here, $\Lambda(s_i) - \Lambda(s_{i-1})$ are independent random variables distributed according to $G(c(\Lambda^p(s_i) - \Lambda^p(s_{i-1})), c^p(s_i))$, and $d\Lambda(s_i)$ are independent random variables distributed according to $G(dN.(s_i), c^p(s_i))$ when $dN.(s_i) > 0$. Therefore, for Λ in the the Gibbs sampling algorithm we only need to generate m many gamma random variables which is less than or equal to $n + N.(\tau^*)$.

We generate random variables from the above four conditional distributions successively, and after a sufficient number of iterations or after satisfying some convergence criterion, stop the iteration and regard the final values of $(\alpha, \beta, \Lambda, \mathbf{w})$ as a sample from the joint posterior distribution.

4. Illustration

In this section, we illustrate our method to a data set of bladder tumor recurrent times (Byar, 1980), whose raw data is available in Wei *et al.* (1989). Observations are the recurrence times of tumor of 85 patients. One of the main interest of this study is to evaluate the effectiveness of thiotepa. There are 47 patients who are in Group 1 (placebo) to which we assign $Z = 0$, and for the remaining 38 patients who are in group 2 (thiotepa) we assign $Z = 1$. For prior distributions, improper flat priors are used for $\exp(\beta) = \gamma$ and α. For a gamma process prior, two positive constants c and θ are chosen such that $c(t) \equiv c$ and $\Lambda_0(t) = \theta t$. It can be easy to show that the resulting posterior distribution with the improper priors for γ and α is proper.

For numerical computation, we choose c = 1.0 and $\theta = 0.1$. In the Gibbs sampling, the conditional posterior distribution of γ given $\Lambda, \mathbf{w}, \alpha$ and $\mathbf{N}$ becomes a gamma distribution with shape parameter $\sum_{i=1}^{n} N_i(\tau) I(Z_i = 1)$ and scale parameter $\sum_{i=1}^{n} w_i I(Z_i = 1)\Lambda(\tau_i)$, so we use Metropolis-Hastings algorithm only for α. For the proposal density, simply the exponential distribution with mean 1 is used. That is, we generate α^* from the exponential distribution with mean 1 and compare this value and the old value to decide whether we update α by α^* or not based on the transition probability. Apparently, this simple Metropolis-Hastings algorithm does not produce any trouble and the whole Gibbs sampling algorithm converges to its stationary distribution fairly fast.

Figure 4.1 shows the plots of density estimates of the posterior distributions of β and α. These are obtained by using the kernel density estimation based on 30000 samples generated from the Gibbs sampling algorithm. Also, basic statistics of β and α are summarized in Table 4.1. Since 95% posterior interval of β covers the origin, we conclude that thiotepa does not reduce the risk of recurrence of bladder tumor significantly. Also, the results about α provides an evidence of heterogeneity among patients. For comparison, we add the result of the estimation of the regression coefficient based on the marginal likelihood (Wei *et al.*, 1989). Our point estimation of the regression coefficient is more or less different from that of the marginal model. This discrepancy is due to that the frailty model is not marginally proportional intensity model. However, the both results confirm that the effect of thiotepa is not statistically significant.

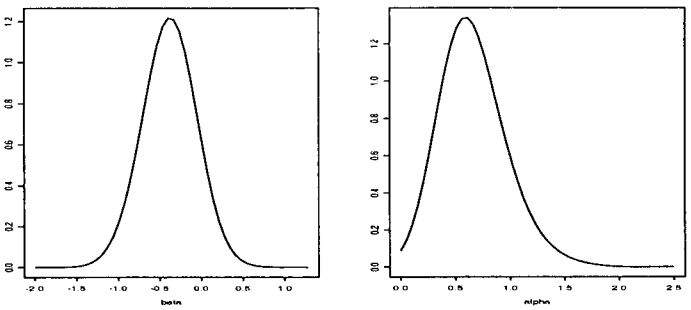

FIGURE 4.1 *Posterior densities of β and α*

TABLE 4.1 *Summary statistics of the posterior distributions of β and α*

	mean	*standard deviation*	*95% posterior interval*
β	−0.385	0.273	(−0.949, 0.119)
$\hat{\beta}$	−0.549	0.285	(−1.119, 0.021)
α	0.660	0.265	(0.240, 1.283)

NOTE : *$\hat{\beta}$ is the estimator based on the marginal model (Wei et al., 1989)*

5. Discussion

Our Bayesian model can easily accommodate time dependent covariates when the covariates are step functions in time. In this case, only thing we should do more is to generate additional gamma variates for Λ whenever the covariates change their values. Also, complicated censoring mechanisms such as on-off system can be dealt with in our model as long as the censoring mechanism is noninformative. Furthermore, it is true, although the proof is not easy, that our Gibbs sampling provides the posterior distribution when the covariates and the censoring mechanism are predictable with respect to the Poisson process as long as they are noninformative. The proof can be done using techniques of Kim (1999). A predictably censored Poisson process is used for the counter model where the process becomes off right after an event occurs and get back to on after certain period of time whose duration is independent with the process. See Karlin and Taylor (1975) for details.

In frequentist's settings, E-M algorithm can be used to calculate the maximum likelihood estimators (Nielsen *et al.*, 1992) in the frailty model. However, within the author's knowledge, no method is available to calculate the variance

of the estimators (in particular, about the baseline cumulative hazard function). In this view, the Bayesian approach would be a good alternative method of estimation from the frequentist's points of view if the large sample properties of the posterior distributions are similar to those for the maximum likelihood estimators. This point should be addressed in future.

Appendix : Conditional Posterior Distribution of Λ

In this appendix, we prove that the conditional posterior distribution of Λ given $\beta, \alpha, \mathbf{w}$ and $\mathbf{N}$ is again a gamma process when Λ is a gamma process *a priori*. Suppose that Λ is a gamma process defined on $[0, \tau]$ with parameters $\big(\Lambda_0(\cdot),\, c(\cdot)\big)$. Let $\eta(t)$ is a positive measurable function on $[0, \tau]$.

Lemma A.1. *Define a process Λ_η by*

$$\Lambda_\eta(t) = \int_0^t \eta(s) d\Lambda(s).$$

Then Λ_η is again a gamma process with parameters $\big(\Lambda_{0\eta}(\cdot), c_\eta(\cdot)\big)$, where

$$\Lambda_{0\eta}(t) = \int_0^t \eta(s) d\Lambda_0(s)$$

and

$$c_\eta(t) = \frac{c(t)}{\eta(t)}.$$

Proof. First, let $c(t) \equiv c > 0$. Then this lemma can be proved easily when η is an indicator function. And then the general conclusion follows from a limiting arguments.

If $c(t)$ is not a constant function, we can write

$$\Lambda(t) = \int_0^t \frac{1}{c(s)} d\tilde{\Lambda}(s)$$

where $\tilde{\Lambda}$ is a gamma process with parameters $\big(\int_0^t c(s) d\Lambda_0(s), 1\big)$. Hence,

$$\Lambda_\eta(t) = \int_0^t \frac{\eta(s)}{c(s)} d\tilde{\Lambda}(s)$$

and the proof is done. □

Theorem A.1. *For a given positive measurable function η, let N is a Poisson process with mean Λ_η where*

$$\Lambda_\eta(t) = \int_0^t \eta(s) d\Lambda(s).$$

Suppose that a priori $\Lambda(t)$ is a gamma process with parameters $\big(\Lambda_0(\cdot), c(\cdot)\big)$. Then the posterior distribution of Λ given N is a gamma process with parameters $\big(\Lambda^p(\cdot), c^p(\cdot)\big)$ where

$$c^p(t) = c(t) + \eta(t)$$

and

$$\Lambda^p(t) = \int_0^t \frac{c(s)d\Lambda_0(s) + dN(s)}{c(s) + \eta(s)}.$$

Proof. By Lemma A.1, Λ_η is a gamma process with parameters $\big(\Lambda_\eta(\cdot), c_\eta(\cdot)\big)$ where

$$\Lambda_\eta(t) = \int_0^t \eta(s) d\Lambda_0(s)$$

and

$$c_\eta(t) = \frac{c(t)}{\eta(t)}.$$

Hence, the conclusion follows from Theorem 3.1 in Lo (1982). □

Now, we can prove that the posterior distribution of Λ given $\beta, \alpha, \mathbf{w}$ and $\mathbf{N}$ is a gamma process with parameters (3.2) and (3.3) by repeated application of Theorem A.1, conditioning first on N_1 with $\eta(t) = Y_1(t)\exp(\beta Z_1)$, then on N_2 with $\eta(t) = Y_2(t)\exp(\beta Z_2)$ and so on.

REFERENCES

Andersen, P. K. and Gill, R. D. (1982). "Cox's regression models for counting processes : A large sample study", *The Annals of Statistics,* **10**, 1100–1120.

Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes*, Academic Press.

Brooks, S. P. (1998). "Markov chain Monte Carlo method and its application", *The Statistician*, **47**, 69–100.

Byar, D. P. (1980). "The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors", In *Bladder Tumors and Other Topics in Urological Oncology* (M. Pavone-Macaluso, P. H. Smith and F. Edsmyn, eds.), Plenum, New York, 363–370.

Clayton, D. (1991). "A model for association in bivariate life-tables and its application in epidemiological studies of chronic disease incidence", *Biometrika*, **65**, 141–151.

Kalbfleisch, J. D. (1978). "Nonparametric Bayesian analysis of survival time data", *Journal of the Royal Statistical Society*, **B40**, 214–221.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York.

Kim, Y. (1999). "Nonparametric Bayesian estimators for counting processes", *The Annals of Statistics*, **27**, 562–588.

Lagakos, S. W., Sommer, C. J. and Zelen, M. (1978). "Semi-Markov models for partially censored data", *Biometrika*, **65**, 311–317.

Lo, A. Y. (1982). "Bayesian nonparametric statistical inference for Poisson point processes", *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **59**, 55–66.

Murphy, S. A. and van der Vaart, A. W. (1997). "Semiparametric likelihood ratio inference", *The Annals of Statistics*, **25**, 1471–1509.

Oakes, D. (1982). "Frailty models for multiple event times", In *Survival Analysis : State of the Art* (J. P. Klein, P. K. Goel and Dordrecht, eds.), Kluwer Academic Publishers, 371–379.

Sinha, E. (1993). "Semiparametric Bayesian analysis of multiple event time data", *Journal of the American Statistical Association*, **88**, 979–983.

Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality", *Demography*, **16**, 439–454.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). "Regression analysis of multivariate incomplete failure time data by modeling marginal distribution", *Journal of the American Statistical Association*, **84**, 1065–1073.