Journal of Statistics: Advances in Theory and Applications
Volume 1, Number 2, 2009, Pages 159-168

A SAMPLING-BASED APPROACH FOR A HYBRID SCALE INTENSITY MODEL

JULIANA COBRE and FRANCISCO LOUZADA-NETO

DEs
Universidade Federal de São Carlos
Via Washington Luís
Km 235, CP. 676
13565-905 São Carlos-SP
Brasil
e-mail: dfln@ufscar.br

Abstract

Several areas such as biomedical, criminology, financial, among others, present recurrent events for the same individual. Louzada Neto [9] proposes a hybrid scale intensity model, which allows multiple scales time, as well as the counting of the events. In this paper, a sampling-based inference procedure based on Markov chain Monte Carlo Methods is developed. We analyzed the proposal through two artificial and a well known real dataset.

1. Introduction

Lifetime data where some event occurs more than one time to the same subjective arise in areas such as biomedical studies, criminology, demography, manufacturing and industrial reliability. The data on each individual consist of the total number of events (lifetimes) observed, the ordered lifetimes and, additionally, covariates information. In such studies, interest may lie in explaining the nature of variation between subjects in terms of treatments and fixed covariates.

2000 Mathematics Subject Classification: 62XX.

Keywords and phrases: intensity models, recurrent events, Bayesian approach, MCMC.

Received April 17, 2009

There is an extensive literature on recurrent event data. The representation of a Poisson process (Lawless [7]) can be used for modelling the total time on study. A renewal process (Prentice et al. [13]) is usually used for modelling the interval time, that is, the time from previous event, implicating that the risk of the next event does not begin until after the previous event has occurred. Another model considers modulated renewal and Poisson processes (Cox [2]), where both scales, total time and interval time, are considered in the study. Besides, Wei et al. [15] propose separate proportional hazards models for the total lifetimes, Pepe and Cai [12] consider rate functions, Nelson [10] presents graphical methods while Lawless and Nadeau [8] consider a more general model, which include prognostic covariates.

Louzada Neto [9] proposes a parametric hybrid scale intensity model for analyzing multiple survival data for use in studies, where individuals can experience recurrent events. The idea is to combine the two time scales, total time and interval time, and the event counts. The model is very flexible and accommodates a broad class of intensity models including the Poisson and renewal process models as very special cases. To date inference for the hybrid scale intensity model has been conducted, wholly, in the classical framework. However, in this paper, we develop for the first time, a sampling-based approach for analyzing the parametric hybrid scale intensity model, in which inference for the model parameters are based on Markov Chain Monte Carlo (MCMC) methods.

The paper is organized as following. In Section 2, we presented the model. In Section 3, we present our sampling-based inference procedure for the model parameters. The developed methodology is applied to two artificial data set and to a real set of medical data in Section 4. In Section 5, we make some final remarks.

2. Model Formulation

The intensity function at time t represents the instantaneous rate of failure at time t given the history of the processes up to time t, mathematically, it is defined as

$$h(t|\mathcal{M}(t)) = \lim_{\Delta t \to 0} \frac{P(m(t, t + \Delta t) = 1|\mathcal{M}(t))}{\Delta t},$$

where $m(t, t + \Delta t)$ denotes the number of events over the interval $[t, t + \Delta t)$, and $\mathcal{M}(t) = \{m(s) : s < t\}$ denotes the history of the processes up to time t. The failure processes are assumed orderly, and this means that the limiting probability of two or more failures in the interval $[t, t + \Delta t)$, given that at least one failure occurs in it, tends to zero as $\Delta t \to 0$.

Following Cox [2], Louzada Neto [9] modelled $h(t|\mathcal{M})$ as a multiplicative function of components for total time, interval time, number of events and covariates, and proposed the hybrid scale intensity model given by

$$h(t|z, \mathcal{M}(t)) = q_1(\nu_t; \theta_1)q_2(t; \theta_2)q_3(j_t; \theta_3)g(\beta^T z), \tag{1}$$

where $g(\cdot)$ is a known positive function that equals one when its argument is zero, β is a vector of unknown regression parameters, and $q_1(\cdot), q_2(\cdot)$ and $q_3(\cdot)$ are positive functions denoting the parametric baseline intensity functions on the interval time, $\nu_t = t - t_{ji}$, total time, t, and event count, $j_t = m(t-)$, respectively, with unknown parameter vectors θ_1, θ_2 and θ_3. The covariates are assumed to be fixed and therefore not affected by the event process.

Particular cases with a reduced number of parameters can be obtained of the wide spectrum of intensity-based models given in (1). We obtain the non homogeneous Poisson process taking $q_1(\cdot) = q_3(\cdot) = 1$ (Lawless [7]). If $q_2(\cdot) = q_3(\cdot) = 1$, we have the renewal process model (Prentice et al. [13]). And if only $q_3(\cdot) = 1$, a hybrid Poisson/renewal intensity model is obtained (Cox [2]).

We studied a special model, where the renewal component is given by a Weibull-type process $q_1(\nu_t; \theta_1) = q_1(\nu_t; \alpha, \gamma) = \alpha\gamma(\alpha\nu_t)^{\gamma-1}$. The Poisson component works as a time dependent Poisson process part, $q_2(t; \theta_2) = q_2(t; \alpha, \phi) = 1 + \alpha\phi t$. The event count function is taking to penalize the occurrence of large numbers of events and it is given by $q_3(j_t; \theta_3) = q_3(j_t; \psi) = \psi^{j_t - 1}$. An exponentially proportional covariate, $g(\beta^T z) = \exp(\beta^T z)$ effect completes deformulation. Then from (1), the intensity function at time t of the considered model is given by

$$h(t|z, \mathcal{M}(t)) = \alpha\gamma(\alpha\nu_t)^{\gamma-1}(1 + \alpha\phi t)\psi^{j_t - 1}e^{\beta^T z}, \qquad (2)$$

where $\alpha, \gamma, \phi, \psi > 0$ and $\beta^T z = \beta_1 z_1 + \beta_2 z_2 + \cdots$ has no intercept term, which is absorbed by α. An advantage of this parametrization is its relatively easy interpretation. The parameters γ, ϕ, ψ are dimensionless numbers, and α denotes a combination between the scale parameter of the renewal part and the intercept term.

The special model (2) also includes several particular models. For instance, the ordinary Weibull renewal model for the interval times is obtained, if $\phi = 0$ and $\psi = 1$. To obtain a special non homogeneous Poisson intensity model, we consider $\gamma = 1$ and $\psi = 1$. If, beyond these last ones, we consider $\phi = 0$, we obtain an ordinary homogeneous Poisson intensity model.

3. Inference

For inference, we adopt a fully Bayesian approach. The likelihood function, prior distributions for the parameters in the model and details of the MCMC algorithm are described below.

3.1. Likelihood function

Suppose that n individuals may experience a certain recurrent event. The dataset of the ith individual is given by: m_i denotes the total numbers of events occurred over the studied time; $t_{i1} \leq t_{i2} \leq \cdots$ are the continuous failure times; $x_{ij} = t_{ij} - t_{ij-1}$ denote the time intervals between successive events; z_i is the covariates vector; and δ_i is indicator variable defined by $\delta_i = 1$, if the failure time is observed, and $\delta_i = 0$, if the failure time is a right-censored observation.

The likelihood function from an individual's interval time x_j is obtained from the intensity function Lawless [7]. The log-likelihood from each individual's interval time x_j is given by

$$l_{j|j-1,\ldots,1} = \delta_j\{\beta^T z + \log\gamma + \gamma\log\alpha + (\gamma-1)\log x_j + \log(1+\alpha\phi t_j) +$$

$$(j-1)\log\psi\} - \left\{1 + \alpha\phi t_{j-1} + \alpha\phi\frac{\gamma}{\gamma+1}x_j\right\}(\alpha x_j)^{\gamma}\psi^{j-1}e^{\beta^T z}. \tag{3}$$

And the log-likelihood is given by the sum regarding all the recurrences and all the individuals, that is,

$$l = \sum_{i=1}^{n}\sum_{j=1}^{m_i} l_{i,\,j|j-1,\ldots,1}. \tag{4}$$

3.2. Sampling-based inference

The target distribution for inference is the posterior of the parameters of interesting $\alpha, \beta, \phi, \gamma$ and ψ. For that, we need to obtain the marginal posterior densities of each parameter, which are obtained by integrating the joint posterior density with respect to each parameter. The posterior distribution is proper considering proper prior distribution (Ibrahim et al. [6]). However, irrespective of the prior distribution chosen, the joint posterior distribution for the proposed model is analytically intractable. As an alternative, we use one of Markov Chain Monte Carlo methods (MCMC), particularly, in our case, the Metropolis-Hastings algorithm (see, e.g., Chib and Greenberg [1]).

Although, in principle, it is not required for our development, we considered $\alpha, \beta, \phi, \gamma$ and ψ to be independent a priori, and we assume a priori density given by $\pi(\alpha, \beta, \phi, \gamma, \psi) = \pi(\alpha)\pi(\beta)\pi(\phi)\pi(\gamma)\pi(\psi)$. These priors shall be properly specified in the Section 4.

Then from (4) the posteriori distribution of the parameter vector, $(\alpha, \beta, \phi, \gamma, \psi)$, based on the observed data, $\boldsymbol{D}$, is given by

$$\pi(\alpha, \beta, \phi, \gamma, \psi | D) \propto \pi(\alpha, \beta, \phi, \gamma, \psi) \exp\left(\sum_{i=1}^{n} \sum_{j=1}^{m_i} l_{i, j | j-1, \ldots, 1} \right). \tag{5}$$

In a Bayesian analysis of a model, the goal is to determine the summary statistics of the conditional posteriori distribution of each parameter, which are not analytically obtainable in our case. For obtaining its numerical approximation, we used a well known Markov Chain Monte Carlo (MCMC) method, the Metropolis-Hastings, see Chib and Greenberg [1]. The complete conditional densities of each parameter that are needed by the algorithm are presented in Appendix A.

4. Application

In this section, the methodology is illustrated on two artificial data set and on a real data set.

4.1. Artificial dataset

In this section, the methodology is applied to two data set, generated from two groups, control $(z = 0)$ and treatment $(z = 1)$, with $\alpha = 1, 5$, $\beta = -1, 2$, $\phi = 0, 8$, $\gamma = 2$ and $\psi = 1, 8$. We analyzed cases with ten recurrent event per individual and different sample sizes (30 and 100).

We use in the simulation study the following prior densities, $\log \alpha \sim \mathcal{N}(0, 3; \sigma^2)$, $\beta \sim \mathcal{N}(-1; \sigma_\beta^2)$, $\log \phi \sim \mathcal{N}(-0, 4; \sigma^2)$, $\log \gamma \sim \mathcal{N}(0, 8; \sigma^2)$ and $\log \psi \sim \mathcal{N}(0, 4; \sigma^2)$, for $\sigma^2 = 0, 2$, $\sigma_\beta^2 = 0, 9$ and $\sigma^2 = 0, 3$, $\sigma_\beta^2 = 1, 3$. In all simulations were made 10,000 iterations, and the first 1,000 iterations were discarded, and the remaining ones were selected with thinning by 50. All simulations were implemented on R system (R Development Core Team [14]). Table 1 summarizes the results of our study, as well as the acceptance rates (AR) of the Metropolis-Hastings algorithm Hastings [5]. The convergence of the chain was monitored through the method proposed by Geweke [4].

Table 1. Posterior summaries with their acceptance rates (AR) of the Metropolis-Hastings algorithm

n	Parameters	$\alpha = 1, 5$	$\beta = -1, 2$	$\phi = 0, 8$	$\gamma = 2, 0$	$\psi = 1, 8$
30	Estimatives	1,878	−1,049	0,722	2,405	1,617
	AR	0,469	0,992	0,861	0,454	0,479
100	Estimatives	1,757	−1,015	0,761	2,809	1,928
	AR	0,584	0,950	0,890	0,463	0,510

4.2. The mammary tumour data

We consider the data extracted from Table 1 of Gail et al. [3], which presents the times to development of mammary tumours for 48 rats in a carcinogenicity experiment. Twenty-three rats were assigned randomly to a treatment group and the remaining 25 to a control group. The rats were induced to remain tumourfree during the first 60 days. And, then they were observed during more 122 days. The hybrid intensity scale model was fitted, on Bayesian approach, to the data considering a transformation from diary scale to annual scale.

The prior densities considered are $\log \alpha \sim \mathcal{N}(-0, 7; 0, 3)$, $\beta \sim \mathcal{N}(2; 1, 5)$, $\log \phi \sim \mathcal{N}(-0, 4; 0, 3)$, $\log \gamma \sim \mathcal{N}(0, 7; 0, 3)$ e $\log \psi \sim \mathcal{N}(1; 0, 3)$. The estimative for the parameters of the model, $\alpha, \beta, \phi, \gamma$ and ψ, and also the credibility interval (between parentheses) are, respectively, 0,463 (0,175;0,938); 2,037 (0,124;4,027); 0,805 (0,294;1,734); 3,463 (1,433;7,124); 3,511 (1,089;6,971). The estimative for ψ is significantly different from one (its nullity point), and it allows to conclude that the count event parameter is important on the dataset analyze, in spite of it is not considered in previous analysis of such data set.

5. Final Remarks

The hybrid intensity scale model accommodates two time scales (interval time and total time), the event counts and covariates. The model provides several particular cases which can be verified straightforwardly via credibility interval. The sampling-based approach of the hybrid intensity scale model allows, besides of the incorporation of priori information, a small computational effort to the obtaining of the estimates of the parameters. The results obtained with the real data show us the importance of the event counting parameter in the analysis.

Appendix A

The full conditionals of the parameters are described down.

$$\pi(\alpha|\cdot) \propto \exp\left\{ \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\gamma \log\alpha + \log(1+\alpha\phi) - (1+\alpha\phi t_{i,j-1} + \alpha\phi\frac{\gamma}{\gamma+1}x_{i,j})\times \right.\right.$$

$$\left.\left.(\alpha x_{i,j})^{\gamma}\psi^{j-1}e^{\beta z_i}\right]\right\}\frac{1}{\alpha}\exp\left\{-\frac{1}{2\sigma_\alpha^2}(\log\alpha-\mu_\alpha)^2\right\},$$

$$\pi(\beta|\cdot) \propto \exp\left\{ \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\beta z_i - (1+\alpha\phi t_{i,j-1}\alpha\phi\frac{\gamma}{\gamma+1}x_{i,j})(\alpha xi,j)^{\gamma}\psi^{j-1}e^{\beta z_i}\right]\right\}\pi(\beta),$$

$$\pi(\phi|\cdot) \propto \exp\left\{ \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\log(1+\alpha\phi) - (1+\alpha\phi t_{i,j} + \alpha\phi\frac{\gamma}{\gamma+1}x_{i,j})\times \right.\right.$$

$$\left.\left.(\alpha x_{i,j})^{\gamma}\psi^{j-1}e^{\beta z_i}\right]\right\}\frac{1}{\phi}\exp\left\{-\frac{1}{2\sigma_\phi^2}(\log\phi-\mu_\phi)^2\right\},$$

$$\pi(\gamma|\cdot) \propto \exp\left\{ \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\log\gamma + \gamma\log\alpha + (\gamma-1)\log x_{i,j} - (1+\alpha\phi t_{i,j-1} + \alpha\phi\frac{\gamma}{\gamma+1}x_{i,j})\times \right.\right.$$

$$\left.\left.(\alpha x_{i,j})^{\gamma}\psi^{j-1}e^{\beta z_i}\right]\right\}\frac{1}{\gamma}\exp\left\{-\frac{1}{2\sigma_\gamma^2}(\log\gamma-\mu_\gamma)^2\right\},$$

and

$$\pi(\psi|\cdot) \propto \exp\left\{ \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[(j-1)\log\psi - \left(1 + \alpha\phi t_{i,j-1} + \alpha\phi\frac{\gamma}{\gamma+1}x_{i,j}\right)\times \right.\right.$$

$$\left.\left.(\alpha x_{i,j})^{\gamma}\psi^{j-1}e^{\beta z_i}\right]\right\}\frac{1}{\psi}\exp\left\{-\frac{1}{2\sigma_{\psi}^{2}}(\log\psi - \mu_{\psi})^{2}\right\},$$

where $\mu_\alpha, \mu_\phi, \mu_\gamma$ and μ_ψ and $\sigma_\alpha, \sigma_\phi, \sigma_\gamma$ and σ_ψ are, respectively, prioris means and prioris deviance of $\log\alpha$, $\log\phi$, $\log\gamma$ and $\log\psi$.

Acknowledgment

The research was partially supported by the Brazilian Organizations CNPQ and CAPES.

References

[1] S. Chib and E. Greenberg, Understanding the Metropolis-Hastings algorithm, The American Statistician 49 (1995), 327-335.

[2] D. R. Cox, The statistical analysis of dependencies in point process, Stochastic Point Processes, pp. 55-66, 1972.

[3] M. H. Gail, T. J. Santner and C. C. Brown, An analysis of comparative carcinogenesis experiments based on multiple times to tumor, Biometrics 36 (1980), 255-266.

[4] J. Geweke, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, Bayesian Statistics 4 (1992), 169-193.

[5] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their application, Biometrika 57 (1970), 247-264.

[6] J. G. Ibrahim, M. H. Chen and D. Sinha, Bayesian Survival Analysis, Springer, New York, (2001).

[7] J. F. Lawless, Statistical Models and Methods for Lifetime Data, Wiley, New York, (1982).

[8] J. F. Lawless and C. Nadeau, Some simple robust methods for the analysis of recurrent events, Technometrics 37 (1995), 158-168.

[9] F. Louzada Neto, A hybrid scale intensity model for recurrent event data, Communication in Statistics 33 (2004), 119-133.

[10] W. Nelson, Graphical analysis of the system repair data, Journal of Quality Technology 20 (1988), 24-35.

[11] W. Nelson, Confidence limits for recurrent data-applied to cost or number of product repairs, Technometrics 37 (1995), 147-157.

[12] M. S. Pepe and J. Cai, Some graphical displays marginal regression analysis for recurrent failure times and time dependent covariates, J. Amer. Stat. Assoc. 88 (1993), 811-820.

[13] R. L. Prentice, B. J. Willians and A. V. Peterson, On the regression analysis of multivariate failure time data, Biometrika 68 (1981), 373-379.

[14] R Development Core Team, R, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, (2006).

[15] L. J. Wei, D. Y. Lin and L. Weissefeld, Regression analysis of multivariate incomplete failure time data by modelling marginal distributions, J. Amer. Stat. Assoc. 84 (1989), 100-116.

■