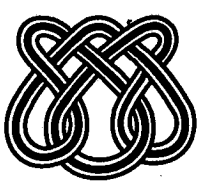
UNIVERSIDADE DE SÃO PAULO

A NOTE ON BAYESIAN EXPONENTIAL REGRESSION MODEL WITH CENSORED DATA

JOSEMAR RODRIGUES
JOSENILDO DE SOUZA CHAVES

Nº 56

NOTAS



Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas e de Computação

ISSN - 0103-2577

A NOTE ON BAYESIAN EXPONENTIAL REGRESSION MODEL WITH CENSORED DATA

JOSEMAR RODRIGUES
JOSENILDO DE SOUZA CHAVES

Nº 56

NOTAS DO ICMC
Série Estatística

São Carlos
Dez./1998

Modelo de Regressão Bayesiana com dado censurados

Resumo

Neste artigo, nós introduzimos uma análise Bayesiana para o modelo exponencial com dados censurados com ênfase na priori via média condicional introduzida por Bedrick, Christensen and Jonhson (1996). Uma ilustração para analisar a sensibilidade do método Bayesiano em relação a essa priori informativa e aos dados censurados, é realizada via algoritmo Metropolis-Hastings utilizando dados reais de pacientes com leucemia.

A note on Bayesian Exponential Regression model with censored data

Josemar Rodrigues and Josenildo de Souza Chaves
ICMC-USP,C.P.668,13560-970, São Carlos ,SP, Brazil

Abstract

In this paper, we consider a Bayesian analysis of the exponential censored survivals time with emphasis on the the conditional means priors introduced by Bedrick, Christensen and Johnson (1996). An illustration to examine the sensitivity of the Bayesian method to this informative prior and censored data is given via Metropolis-Hasting algorithm using acute leukemia survival data.

Some Key words: Conditional mean prior; Metropolis-Hasting algorithm; Predictive de density.

1 Introduction

Feigl and Zelen (1965) have described a statistical model for the study of uncensored survival data in which survival time for a chronic desease pacient is assumed an exponential distribution with linear relation between the failure mean time and some characteristics of the desease or explanatory variables. However, some authors in the literature (Kaplan and Meier, 1958; Prentice, 1973) have been frequently suggested an exponential relation between the failure rate and the explanatory variables. So, we consider in this paper the following hazard function :

$$h(t_i) = \exp\{-(\alpha + \beta Z_i)\}, \quad i = 1, \ldots, n, \tag{1}$$

where t_i is the survival time of the ith individual, $Z_i = X_i - \overline{X}$, X_i the explanatory variable and α and β the unknown parameters. The explanatory

variables have been standardized about the mean in order that α may be interpreted as the average of the logarithm of the failure rates. The hazard function in (1) with a single explanatory variable was suggested by Glasser (1967) for use in application very similar to those that motivated the present work. Also, Cox and Snell (1968) considered this model to illustrate the analysis of residuals. Many authors like Fiegl & Zelen, 1965; Cox, 1964; Zippen& Armitage, 1996 and others suggested a linear relation between the failure rate and the explanatory variables. The disadvantage of this relation is that the range of the values for the coefficient regressions is restrict to the values of the mean failure time.

The purpose of this paper is to develop a Bayesian analysis for the parameters (α, β) and to estimate the predictive survival curves based on censored data . To develop a Bayesian procedure we used a more realistic prior introduced by Bedrick et al. (1996) which consists in eliciting prior information for the mean responses corresponding to responses with two conveniently fixed explanatory variables to induce a prior on (α, β). The sensibility of this informative prior and the effect of the censoring procedure to the posterior distributions are illustrated via Metropolis-Hasting algorithm in the final section.

2 Bayesian procedure for an exponential regression model with a single explanatory variable.

Consider that n pacients are entered into a study, and that r of these have died and $n-r$ are still alive at the end of the study. Then, $t_1, t_2, \ldots, t_r$, are observed survival times while $t_{r+1}, \ldots, t_n$, are censored at $t^o_{r+1}, \ldots, t^o_n$, respectively. Assuming this censored data is generated by the model in (1), the likelihood function is

$$L[\alpha, \beta] = \Pi_{i=1}^{r} e^{-(\alpha+\beta Z i)} \exp\{-t_i e^{-(\alpha+\beta Z_i)}\} \exp\{-\sum_{j=r+1}^{n} t^o_j e^{-(\alpha+\beta Z_j)}\}. \qquad (2)$$

Our purposes from the Bayesian point of view are to estimate the survival curve

$$P[T > t \mid \alpha, \beta, X) = \exp\{-te^{-(\alpha+\beta(X-\overline{X}))}\},$$

the marginal posteriors of α and β and the posterior quantiles.

2.1 The induced prior on (α, β).

Following the conditional mean prior procedure (CMP) introduced by Bedrick et al. (1996), let

$$\tilde{m}_i = E[T_i \mid \alpha, \beta, \tilde{X}_i] = e^{-(\alpha+\beta(\tilde{X}_i-\overline{X}))}, \quad i = 1, 2,$$

where $\{(\tilde{t}_i, \tilde{x}_i), i = 1, 2, \tilde{X}_1 \neq \tilde{X}_2\}$ are the "prior data", then

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 & (\tilde{X}_1 - \overline{X}) \\ 1 & (\tilde{X}_2 - \overline{X}) \end{pmatrix}^{-1} \begin{pmatrix} -ln(\tilde{m}_1) \\ -ln(\tilde{m}_2) \end{pmatrix}. \tag{3}$$

Assuming that $\tilde{m}_i \sim InvGamma(\tilde{w}_i, \tilde{w}_i\tilde{t}_i), i = 1, 2,$, the values $\tilde{X}_1, \tilde{X}_2$ are well separated to guarantee some independence between $\tilde{m}_1$ and $\tilde{m}_2$ and $\tilde{w}_i$ as a weight parameter associated to $\tilde{t}_i$, the induced prior on (α, β), from the result in (3), is

$$\pi(\alpha, \beta) \propto \Pi_{i=1}^{2} e^{-(\alpha+\beta\tilde{Z}_i)\tilde{w}_i} \exp\{-\tilde{t}_i\tilde{w}_i e^{-(\alpha+\beta\tilde{Z}_i)}\}. \tag{4}$$

Now, our problem consist in choosing an appropriated value for the weight parameter $\tilde{w}_i, i = 1, 2$ and assessing the "prior information" $\tilde{t}_i$.

- **Choosing the weight $\tilde{w}_i$:**
  If we make a comparison between the uncensored part of the likelihood function in (2) and the induced prior on (α, β), it is reasonable to think $\tilde{w}_i = 1$, that is, the prior information, $\tilde{t}_i$, having the same "weight" as the information in a single observation t_i. It is interesting to observe that this particular induced prior is the same prior obtained by Bedrick et al. called "Data Augmentation Prior"(DAP) for an uncensored sample.

- **Assessing the hyperparameter $\tilde{t}_i$, (Bedrick et al. , 1996)**
  Bedrick et al. suggested the following interesting way to assess the hyperparameter $\tilde{t}_i$: Let $\tilde{p}_i = e^{t_i/\tilde{m}_i}$, for an appropriated choice of $t_i, i = 1, 2$. Given specified values $\tau_i, i = 1, 2$, we elicit γ_i such that

  $$Pr[\tilde{p}_i \geq \gamma_i] = \tau_i, i = 1, 2,$$

  Thus, we have that

  $$\tau_i = Pr[\frac{\tilde{t}_i}{\tilde{m}_i} \leq -\frac{\tilde{t}_i ln(\gamma_i)}{t_i}],$$

where $\frac{\tilde{t}_i}{\tilde{m}_i} \sim Gamma(1,1)$ and an explict solution for $\tilde{t}_i$ is

$$\tilde{t}_i = \frac{t_i ln(1-\tau_i)}{ln(\gamma_i)}. \tag{5}$$

2.2 The joint posterior distribution

Based on (2) and (4). we have for $\tilde{w}_1 = 1$, the following joint posterior distribution for (α, β):

$$\pi(\alpha, \beta \mid data) \propto e^{(r+2)\alpha+\beta(\tilde{Z}_1+\tilde{Z}_2+\sum_{i=1}^{r} Z_i)} \exp\{-\sum_{i=1}^{2} \tilde{t}_i e^{-(\alpha+\beta\tilde{Z}_i)} +$$

$$+\sum_{i=1}^{r} t_i e^{-(\alpha+\beta Z_i)} + \sum_{i=r+1}^{n} t_i^o e^{-(\alpha+\beta Z_i)}\}. \tag{6}$$

Now, we list the conditional densities used in the transition measure of the Markov chain,

$$\pi(\alpha \mid \beta, data) \propto \exp\{-(r+2)\alpha\} A(\alpha, \beta), \tag{7}$$

$$\pi(\beta \mid \alpha, data) \propto \exp\{-(\sum_{i=1}^{r} Z_i + \tilde{Z}_1 + \tilde{Z}_2)\beta\} A(\alpha, \beta),$$

where,

$$A(\alpha, \beta) = \exp\{-(\sum_{i=1}^{r} t_i e^{-(\alpha+\beta Z_i)} + \sum_{i=r+1}^{n} t_i^o e^{-(\alpha+\beta Z_i)} + \sum_{i=1}^{2} \tilde{t}_i e^{-(\alpha+\beta\tilde{Z}_i)})\}.$$

The variables α and β should be generated using the Metropolis-Hastings algorithm (see for example, Chib and Greemberg, 1994).

3 An illustrative example

As an illustration, consider the data given by Zippen and Armitage (1966). Table 1 shows the white blood count values, follow up times and the status of each of pacients. From the conditional densities given in (7) and the "prior

Table 1: White Blood Count (WBC)-Dead (D)- Alive (A)

WBC:thousands	$log_{10}WBC$(X)	t weeks	Status
0.75	2.8751	156	D
2.3	3.3617	65	D
2.6	3.4150	13	A
4.3	3.6335	100	D
5.4	3.7324	39	D
6.0	3.7782	16	D
7.0	3.8451	143	D
9.4	3.9731	34	A
10.0	4.0000	121	D
10.5	4.0212	108	D
17.0	4.3204	4	D
32.0	4.5052	5	A
35.0	4.5441	13	A
812.0	4.7160	5	D
100.0	5.0000	1	A
100.0	5.0000	1	D
100.0	5.0000	65	D

data" in Table 2, we generated 5 separated Gibbs chains each of which ran for 2000 interations, and we monitored the convergence of the Gibbs samplers using the Gelman and Rubin (1992) method that uses the analysis of variance technique to determine if further iterations are needed. For each parameter we consider the 515th, 530th,..., 2000th iteration , which for 5 chains yelds a sample of size 500. Figure 1 presents contour plots of the joint posterior densities for different "prior sample" given by Table 2. In Table 3, we have the posterior quantiles for $\tilde{t}_1 = 86.37$ and $\tilde{t}_2 = 19.24$.

Table 2: The prior samples for τ=0.90 and $\tau_2 = 0.10$.

t_1	t_2	t_1	t_2	γ_1	γ_2	$\tilde{t}_1$	$\tilde{t}_2$
3	5	52	52	0.25	0.75	86.37	19.04
3	5	52	52	0.45	0.55	149.95	9.16
3	5	52	52	0.65	0.35	277.95	5.22
3	5	52	52	0.80	0.20	536.58	3.40

Table 3: The posterior quantiles for β with the censored data

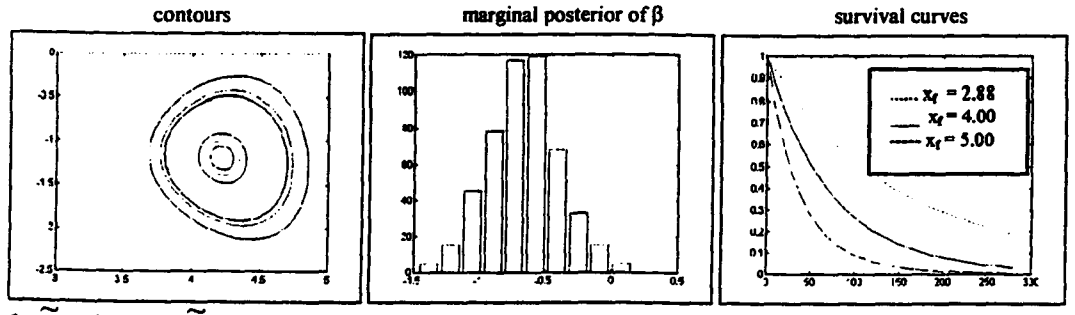
$\tilde{t}_1$	$\tilde{t}_2$	0.05	0.25	Median	0.75	0.95
86.37	19.04	-1.10	-8.45	-6.72	-4.92	-2.00
149.95	9.16	-9.30	-6.79	-5.05	-3.45	-1.16
277.95	5.22	-8.04	-5.22	-3.76	-1.71	1.15
536.58	3.40	-5.90	-2.95	-1.22	5.89	3.43

Figure 1 calls our attention for some interesting points about the influence of the hyperparametrs stated in Table 2:

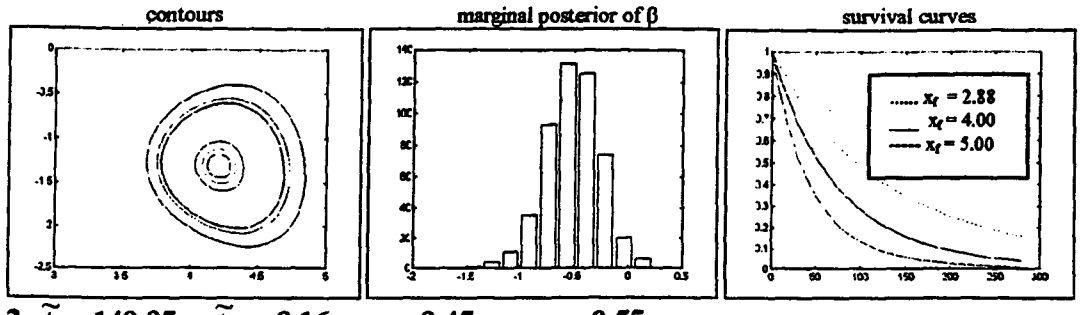
- The contours show some sensitivity on the parameter β but can see a stability with respect to the parameter α. This remark is very important bacause the coefficient β measures the influence of the explanatory variable X.

- The marginal posteriors of β confirm this influence of the prior distribution.

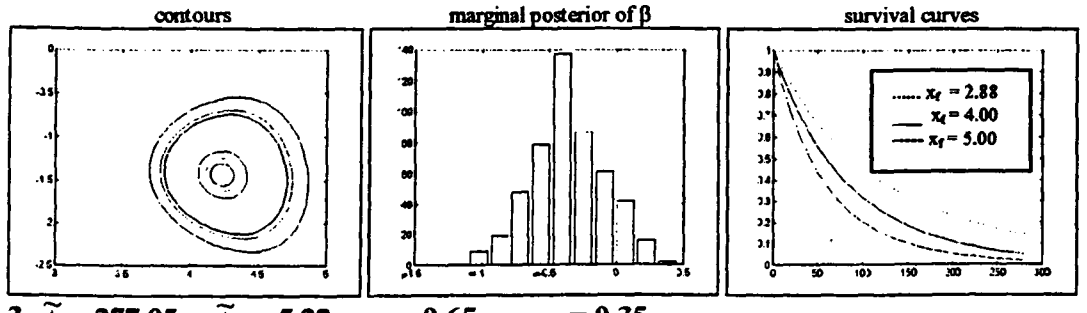
- The predictive curves are very sensitity to the CMP for different values of $\tilde{t}_1$ and $\tilde{t}_2$. If an expert gives a non realistic ilicitation for γ_1, for example, $\gamma_1 = 0.65$ for $\tau = 0.90$, the diagnostic will be same for a terminal pacient or not.



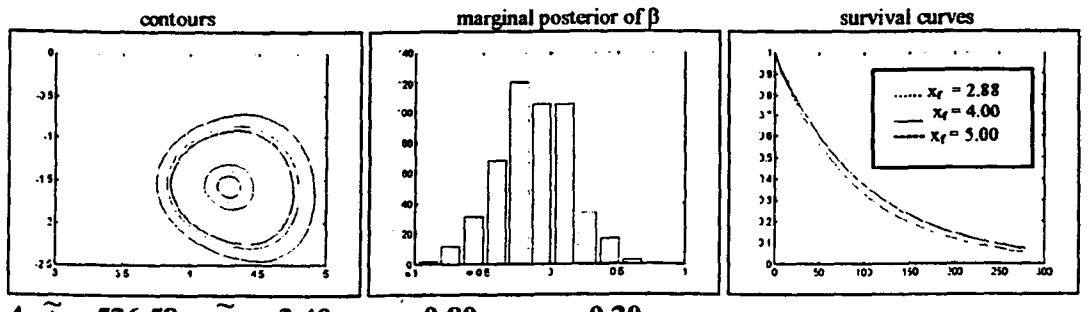
1. $\tilde{t}_1 = 86.37$; $\tilde{t}_2 = 19.04$; $\gamma_1 = 0.25$; $\gamma_2 = 0.75$



2. $\tilde{t}_1 = 149.95$; $\tilde{t}_2 = 9.16$; $\gamma_1 = 0.45$; $\gamma_2 = 0.55$



3. $\tilde{t}_1 = 277.95$; $\tilde{t}_2 = 5.22$; $\gamma_1 = 0.65$; $\gamma_2 = 0.35$



4. $\tilde{t}_1 = 536.58$; $\tilde{t}_2 = 3.40$; $\gamma_1 = 0.80$; $\gamma_2 = 0.20$

Figure 1: Contours of the joint posterior density of (α, β), the marginal posterior density of β and the predictive survival curves for the censored data.

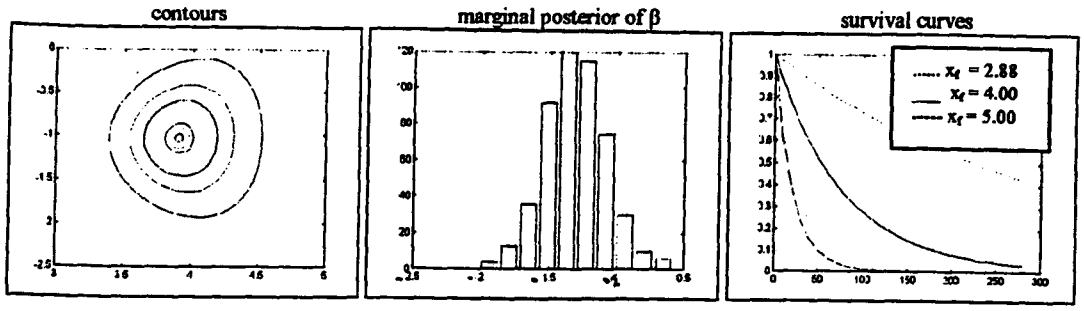


Figure 2: Contours of the joint posterior density of, (α, β), the marginal posterior density of β and the predictive survival curves for the uncensored data, $\tilde{t}_1 = 86.37$; $\tilde{t}_2 = 19.04$; $\gamma_1 = 0.25$; $\gamma_2 = 0.75$.

Figure 1 and 2 present the estimates of predictive suvival curves;

$$Pr[T_f \geq t \mid x_f, data) = \int \int \exp\{-te^{-(\alpha+\beta x_f)}\}\pi(\alpha, \beta \mid data)d\alpha d\beta.$$

Comparing Figure 1.1 (censored data) with Figure 2. (uncensored data), we see clearly a very strong effect of the censored data on the predictive density.

References

Chib, S. and Greenberg, E. (1994), Understand the Metropolis-HAstings algorithm. Manuscript. Bedrick, E.J., Christensen, R. and Johnson, W. (1996), A new perspective on priors for generalized linear models, Journal of the American Statistical Association, 91,1450-1460.

Cox, D.R. (1964), Some applications of exponential ordered scores, Journal Royall Statistical Society,B,26,103-10.

Cox,D.R.,and Snell, E.J. (1968), A general definition of residuals (with discussion), Journal Royall od Statistical Society, B,30,248-75.

Feigl,P., and Zelen M. (1995), Estimation of exponential survival probabilities with concomitant information, Biometrics , 21, 826-838.

Gelman, A.E. and Rubin, D. (1992), Inference from iterative simulation using multiple sequences, Statistical Science, 7,457-472.

Glasser, M. (1967), Exponential survival with covariance, Journal of American Statistical Association, 62,561-8.
Kaplan, E.L. and Meier, P. (1958), Nonparametric estimation from incomplete observations, Journal of American Statistical Association, 53,457-81.
Prentice, R.L. (1973), Exponential survivals with censoring and explanatory variables, Biometrika, 60,2, 279-288.
Zippen,C. and Armitage, P. (1996), Use of concomitant variables and exponential survival parameter, Biometrics, 22, 665-72

NOTAS DO ICMC

SÉRIE ESTATÍSTICA

055/98 RODRIGUES, J.; CHAVES, J.S. – A note on bayesian exponential regression model with censored data.

054/98 ACHCAR, J.A.; FORTULAN, V.C. – Relação entre o uso de hormônio e câncer em mulheres: uma aplicação de meta-análise sob um enfoque bayesiano.

053/98 CID, J. E.R.; ACHCAR, J..A. – Bayesian inference for nonhomogeneous poisson processes in software reliability models assuming nonmonotonic intensity functions.

052/98 ANDRADE FILHO, M.G; MIZOI, M.F. – Aplicação de MCMC na estimação de máxima verossimilhança para processos AR(p) e MA(q).

051/98 RODRIGUES, J. – Bayesian analysis for the accelerated life tests with informative prior distributions obtained from fixed stresses.

050/98 ACHCAR, J.A.; ANDRADE, .M.G.; LOIBEL, S. - A bayesian analysis for homogeneous poisson processes with change-points.

049/98 ACHCAR, J.A; PEREIRA, G.A. - Use of mixture of exponential power distributions for interval-censored survival data in presence of covariates.

048/98 CID, J.E.R.; ACHCAR, J.A. - Software reliability considering the super position of non-homogeneous Poisson processes in the presence of a covariate.

047/98 ACHCAR, J.A.; PEREIRA, G.A. - Use of exponential power distributions for mixture models in the presence of covariates.

046/98 ANDRADE, M.G.; HUTTER, C.F.F. - Teste de sazonalidade para função de autocorrelação de processos auto-regressivos periódicos - PAR (pm)