UNIVERSIDADE DE SÃO PAULO

A BAYESIAN ANALYSIS FOR HOMOGENEOUS POISSON PROCESSES WITH CHANGE-POINTS

JORGE ALBERTO ACHCAR
MARINHO GOMES ANDRADE
SELENE LOIBEL

Nº 50

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

A BAYESIAN ANALYSIS FOR HOMOGENEOUS POISSON PROCESSES WITH CHANGE-POINTS

JORGE ALBERTO ACHCAR
MARINHO GOMES ANDRADE
SELENE LOIBEL

Nº 50

NOTAS DO ICMSC
Série Estatística

São Carlos
Mai./1998

A Bayesian analysis for homogeneous Poisson processes with change-points

Jorge Alberto Achcar
Marinho Gomes Andrade
Selene Loibel
University of São Paulo, ICMC
C.P. 668, 13560-970, São Carlos, S.P., Brazil

ABSTRACT

"Gibbs sampling with Metropolis-Hastings algorithms are proposed to perform a Bayesian analysis for homogeneous Poisson processes with one or more change-points. We also present some Bayesian criteria to discriminate different models. The methodology is illustrated with a data set introduced by Maguire, Pearson and Wynn (1952)."

Key words: Poisson processes, change-points, Gibbs sampling.

1. Introduction

Consider a homogeneous Poisson process with one or more change-points at unknown times. With a single change-point, the rate of occurrence at time s is given by,

$$\lambda(s) = \begin{cases} \lambda_1 & 0 \leq s \leq \tau \\ \lambda_2 & s > \tau \end{cases} \tag{1}$$

The analysis of the Poisson process is based on the observation period $[0, T]$, during which $N(T) = n$ events occur at times $t_1, t_2, \ldots, t_n$.

With two change-points at unknown times τ_1 and τ_2, the rate of occurrences are given by

$$\lambda(s) = \begin{cases} \lambda_1, & 0 < s \leq \tau_1 \\ \lambda_2, & \tau_1 < s \leq \tau_2 \\ \lambda_3, & \tau_2 < s \leq T \end{cases} \tag{2}$$

We also could have homogeneous Poisson processes with more than two change-points.

Observe that times between failures for homogeneous Poisson process follow a exponential distribution.

Applications of change-points models are given in many areas of interest. For example, medical researchers usually have interest to know if a new therapy of leukemia produces a departure from the usual experience of a constant relapse rate after the induction of a remission (see for example, Matthews and Farewell, 1982; or Matthews, Farewell and Pyke, 1985). Bayesian analysis for change-point models are introduced by many authors (see for example, Achcar and Bolfarine, 1989). A Bayesian analysis of a Poisson process with a change-point is introduced by Raftery and Akman (1986).

In this paper, we present Bayesian inferences for homogeneous Poisson processes with one or more change-points using Metropolis-with-Gibbs algorithm (see for example, Gelfand and Smith, 1990; Chib and Greenberg, 1995; or Smith and Roberts, 1993).

2. The likelihood function

Let $x_i = t_i - t_{i-1}$, $i = 1, 2, \ldots, n$ where $t_0 = 0$, be the interfailure times and assume a single-change-point model (1). Assuming τ is taking one value t_i, the likelihood function for λ_1, λ_2 and τ is given by

$$L(\lambda_1, \lambda_2, \tau) = \prod_{i=1}^{N(T)} (\lambda_1 e^{-\lambda_1 x_i})^{\epsilon_i} (\lambda_2 e^{-\lambda_2 x_i})^{1-\epsilon_i} \tag{3}$$

where $\epsilon_i = 1$ if $\sum_{j=1}^{i} x_j \leq \tau$ and $\epsilon_i = 0$ if $\sum_{j=1}^{i} x_j \geq \tau$.

That is,

$$L(\lambda_1, \lambda_2, \tau) = \lambda_1^{N(\tau)} e^{-\lambda_1 \tau} \lambda_2^{N(T)-N(\tau)} e^{-\lambda_2 (T-\tau)} \tag{4}$$

where $N(\tau) = \sum_{i=1}^{N(T)} \epsilon_i$, $N(T) = n$, $\tau = \sum_{i=1}^{N(T)} x_i \epsilon_i$ and $T - \tau = \sum_{i=1}^{N(T)} x_i (1 - \epsilon_i)$.

Assuming a two-change-point model (2) with the change-points τ_1 and τ_2 taking discrete values ($\tau_1 = t_i$, $\tau_2 = t_j$, $i \neq j$), the likelihood function for λ_1, λ_2, λ_3, τ_1 and τ_2 is given by,

$$L(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2) = \prod_{j=1}^{3} \prod_{i=1}^{n} (\lambda_j e^{-\lambda_j x_i})^{\epsilon_{ji}} \tag{5}$$

where $\epsilon_{ji} = 1$ if $\tau_{j-1} < \sum_{j=1}^{i} x_j \leq \tau_j$, $\epsilon_{ji} = 0$ in other case , $j = 1, 2, 3$; $\tau_0 = 0$ and $\tau_3 = T$.

That is,

$$L(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2) = \prod_{j=1}^{3} \lambda_j^{N(\tau_j)} e^{-\lambda_j T_j} \tag{6}$$

where $N(\tau_j) = \sum_{i=1}^{n} \epsilon_{ji}$ e $T_j = \sum_{i=1}^{n} x_j \epsilon_{ji}$.

Observe that $T_1 = \tau_1$, $T_2 = \tau_2 - \tau_1$ and $T_3 = T - \tau_2$ $(\tau_1 < \tau_2)$.

3. A Bayesian analysis

Assume the change-point model (1) with a single change-point τ.

Given $\tau = t_i$ and assuming prior independence among the parameters λ_1 and λ_2, a non-informative prior density for λ_1 and λ_2 (see for example, Box and Tiao, 1973) is given by

$$\pi(\lambda_1, \lambda_2 | \tau = t_i) \propto \frac{1}{\lambda_1 \lambda_2} \tag{7}$$

where $\lambda_1, \lambda_2 > 0$.

Assuming an uniform prior distribution $\pi_0(\tau = t_i) = 1/n$, the joint posterior distribution for λ_1, λ_2 and τ is given by,

$$\pi(\lambda_1, \lambda_2, \tau | \mathcal{D}) \propto \lambda_1^{N(\tau)-1} e^{-\lambda_1 \tau} \lambda_2^{n-N(\tau)-1} e^{-\lambda_2 (T-\tau)} \tag{8}$$

where $\mathcal{D}$ denotes the data set.

The marginal posterior distribution for τ is , from (8), given by,

$$\pi(\tau | \mathcal{D}) \propto \pi_0(\tau) \frac{\Gamma[N(\tau)]\Gamma[n-N(\tau)]}{\tau^{N(\tau)}(T-\tau)^{n-N(\tau)}} \tag{9}$$

Assuming τ known, the marginal posterior distribution for λ_1 and λ_2 are given by,

$$\begin{aligned} &i)\ \lambda_1 | \tau, \mathcal{D} \sim \Gamma[N(\tau), \tau] \\ &ii)\ \lambda_2 | \tau, \mathcal{D} \sim \Gamma[n - N(\tau), T - \tau] \end{aligned} \tag{10}$$

where $\Gamma[a, b]$ denotes a gamma distribution with mean a/b and variance a/b^2.

Assuming τ unknown, the conditional posterior distributions for the Gibbs algorithm are given by,

$$i)\ \lambda_1 | \lambda_2, \tau, \mathcal{D} \sim \Gamma[N(\tau), \tau]$$

$$ii)\ \lambda_2 | \lambda_1, \tau, \mathcal{D} \sim \Gamma[n - N(\tau), T - \tau] \quad (11)$$

$$iii)\ \pi(\tau | \lambda_1, \lambda_2, \mathcal{D}) \propto \left(\frac{\lambda_1}{\lambda_2}\right)^{N(\tau)} exp\{-\tau(\lambda_1 - \lambda_2)\}$$

Observe that, we need to use the Metropolis-Hastings algorithm to generate the variable τ.

We could monitor the convergence of the Gibbs samples using the Gelman and Rubin (1992) method that uses the analysis of variance technique to determine if further iterations are needed.

Consider now, the change-point model (2) with two change-points τ_1 and τ_2. The prior density for $\lambda_1, \lambda_2, \lambda_3, \tau_1$ and τ_2 is given by,

$$\pi(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2) = \pi(\lambda_1, \lambda_2, \lambda_3 | \tau_1 = t_i, \tau_2 = t_j)\, \pi_0(\tau_1 = t_i, \tau_2 = t_j) \quad (12)$$

Given $\tau_1 = t_i, \tau_2 = t_j$, $i \neq j$, and assuming prior independence among the parameters λ_1, λ_2 and λ_3 , a non-informative prior density for λ_1, λ_2 and λ_3 is given by,

$$\pi(\lambda_1, \lambda_2, \lambda_3 | \tau_1, \tau_2) \propto \frac{1}{\lambda_1 \lambda_2 \lambda_3} \quad (13)$$

where $\lambda_1, \lambda_2, \lambda_3 > 0$.

Assuming uniform prior distribution for the discrete variables τ_1 and τ_2, the joint posterior distribution for $\lambda_1, \lambda_2, \lambda_3, \tau_1$ and τ_2 is given by,

$$\pi(\lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2 | \mathcal{D}) \propto \prod_{j=1}^{3} \lambda_j^{N(\tau_j)-1} e^{-\lambda_j T_j} \quad (14)$$

where T_j is given in (6).

The conditional posterior distributions for the Gibbs algorithm are given by,

$$i)\ \pi(\lambda_j | \underset{\sim}{\lambda}_{(j)}, \tau_1, \tau_2, \mathcal{D}) \propto \Gamma[N(\tau_j), \tau_j],\ \ j = 1, 2, 3.$$

$$ii)\ \pi(\tau_1 | \lambda_1, \lambda_2, \lambda_3, \tau_2, \mathcal{D}) \propto \lambda_1^{N(\tau_1)} e^{-\tau_1(\lambda_1 - \lambda_2)} \quad (15)$$

$$iii)\ \pi(\tau_2 | \lambda_1, \lambda_2, \lambda_3, \tau_1, \mathcal{D}) \propto \lambda_2^{N(\tau_2)} e^{-\tau_2(\lambda_2 - \lambda_3)}$$

where $\underset{\sim}{\lambda}_{(j)}$ is the vector of λ_i, $i = 1, 2, 3$ not including λ_j.

Observe that we need to use the Metropolis-Hastings algorithm to generate the variables τ_1 and τ_2.

4. Some considerations on model selection

For model selection, we could use the predictive density for the interfailure time x_i given $\underset{\sim}{x}_{(i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

The predictive density for x_i given $\underset{\sim}{x}_{(i)}$ is,

$$c_i = f(x_i | \underset{\sim}{x}_{(i)}) = \int f(x_i | \underset{\sim}{\theta}) \pi(\underset{\sim}{\theta} | \underset{\sim}{x}_{(i)})\, d\underset{\sim}{\theta} \tag{16}$$

where $\pi(\underset{\sim}{\theta} | \underset{\sim}{x}_{(i)})$ is the posterior density for a vector of parameters $\underset{\sim}{\theta}$ given the data $\underset{\sim}{x}_{(i)}$.

Using the Gibbs samples, (16) can be approximated by its Monte Carlo estimates,

$$\widehat{f}(x_i | \underset{\sim}{x}_{(i)}) = \frac{1}{M} \sum_{j=1}^{M} f(x_i \backslash \underset{\sim}{\theta}^{(j)}) \tag{17}$$

where $\underset{\sim}{\theta}^{(j)}$ are the generated Gibbs samples, $j = 1, 2, \ldots, M$.

We can use $c_i = \widehat{f}(x_i | \underset{\sim}{x}_{(i)})$ in model selection. In this way, we consider plots of c_i versus i $(i = 1, 2, \ldots, n)$ for different models; large values of c_i (in average) indicates the better model. We also could choose the model such that $P_l = \prod_{i=1}^{n} c_i(l)$ is maximum (l indexes models).

5. An example

In table 1, we have the time intervals (in days) between explosions in mines, involving more than 10 men killed, from 6 December 1875 to 29 May 1951 (data introduced by Maguire, Pearson and Wynn , 1952).

378	36	15	31	215	11	137	4	15	72	96	124	50	120
203	176	55	93	59	315	59	61	1	13	189	345	20	81
286	114	108	188	233	28	22	61	78	99	326	275	54	217
113	32	23	151	361	312	354	58	275	78	17	1205	644	467
871	48	123	457	498	49	131	182	255	195	224	566	390	72
228	271	208	517	1613	54	326	1312	348	745	217	120	275	20
66	291	4	369	338	336	19	329	330	312	171	145	75	364
37	19	156	47	129	1630	29	217	7	18	1357			

Table 1 - Time intervals in days between explosions in mines

From a plot of $N(t_i)$ versus t_i, $i = 1, 2, \ldots, 109$ (see figure 1), we observe that the two change-points model (2) seems more appropriate for the data set of table I. These change-points are approximately $\hat{\tau}_1 = t_{45} = 5231$ and $\hat{\tau}_2 = t_{81} = 19053$ (from figure 1).

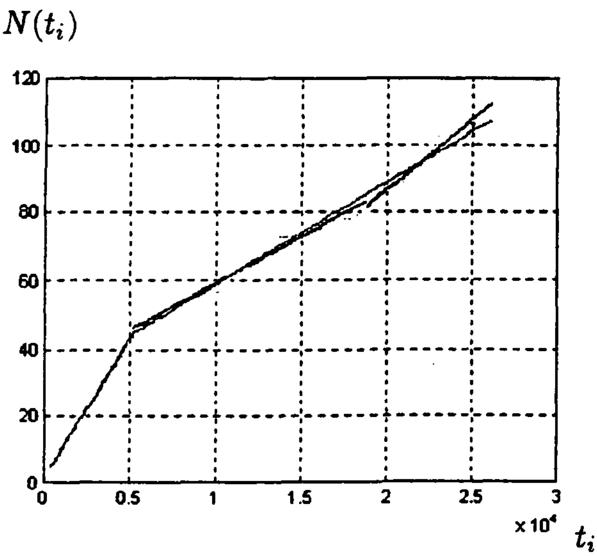

Figure 1 - Plot of $N(t_i)$ versus t_i

If we assume the change-point model (1) with a single change-point τ, the mode of the marginal posterior distribution for τ (see (9)) is given by $\tilde{\tau} = 5376$ (see figure 2). Assuming $\tau = 5376$ known, the modes of the marginal posterior distributions (10) are given by $\tilde{\lambda}_1 = 0.008352$ and $\tilde{\lambda}_2 = 0.002942$.

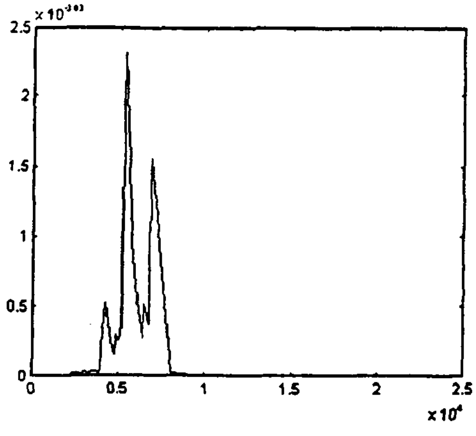

Figure 2 - Marginal posterior distribution for τ

Assuming τ unknown, we generated 5 separate Gibbs chains from the conditional posterior distributions (11) each of which ran for 2000 iterations, and we monitored the convergence of the Gibbs samples using the Gelman and Rubin (1992) method. For each parameter, we considered the 515^{th}, 530^{th}, ..., 2000^{th} iterations, which for 5 chains yields a sample of size 500.

In table 2, we have the obtained posterior summaries for the parameters λ_1, λ_2 and τ, and in figure 3, we have the approximate marginal posterior densities considering the 500 Gibbs samples. We also have in table 2, the estimated potential scale reductions $\widehat{R}$ (see Gelman and Rubin, 1992) for the parameters. In this case, the considered number of iterations were sufficient for approximate convergence. ($\sqrt{\widehat{R}} < 1.1$ for all parameters).

	MEAN	S.D.	MODE	$\widehat{R}$
τ	5674	762	5311	0.9989
λ_1	0.0081	0.0013	0.0076	1.0023
λ_2	0.0030	0.0004	0.0031	1.0025

Table 2 - Posterior summaries (change-point model 1)

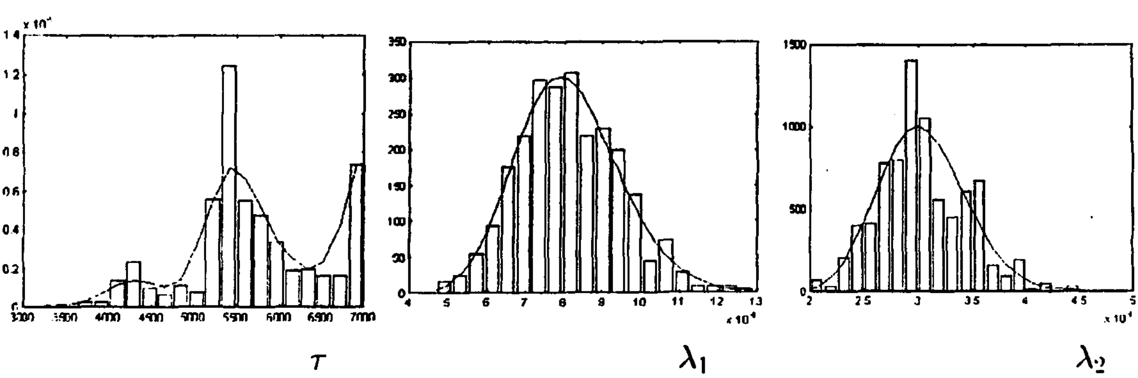

Figure 3 - Marginal posterior distribution (change-point model 1)

Assuming now the two change-point model (2), we also generated 5 separate Gibbs chains from the conditional posterior distributions (15) each of which ran for 2000 iterations. For each parameter, we considered the 515^{th}, 530^{th}, ...,2000^{th} iterations, which for 5 chains yields a sample of size 500.

In table 3, we have the obtained posterior summaries for parameters $\lambda_1, \lambda_2, \lambda_3, \tau_1$ and τ_2, and in figure 4 we have the approximate marginal posterior densities considering the 500 Gibbs samples. In this case, we have approximate convergence ($\sqrt{\widehat{R}} < 1.1$ for all parameters).

	MEAN	S.D.	MODE	$\widehat{R}$
τ_1	5030	18.3	5044	1.0072
τ_2	18681	99.0	18796	1.0084
λ_1	0.0081	0.0014	0.0084	1.0015
λ_2	0.0024	0.0004	0.0022	1.0050
λ_3	0.0039	0.0007	0.0037	0.9980

Table 3 - Posterior summaries (change-point model 2)

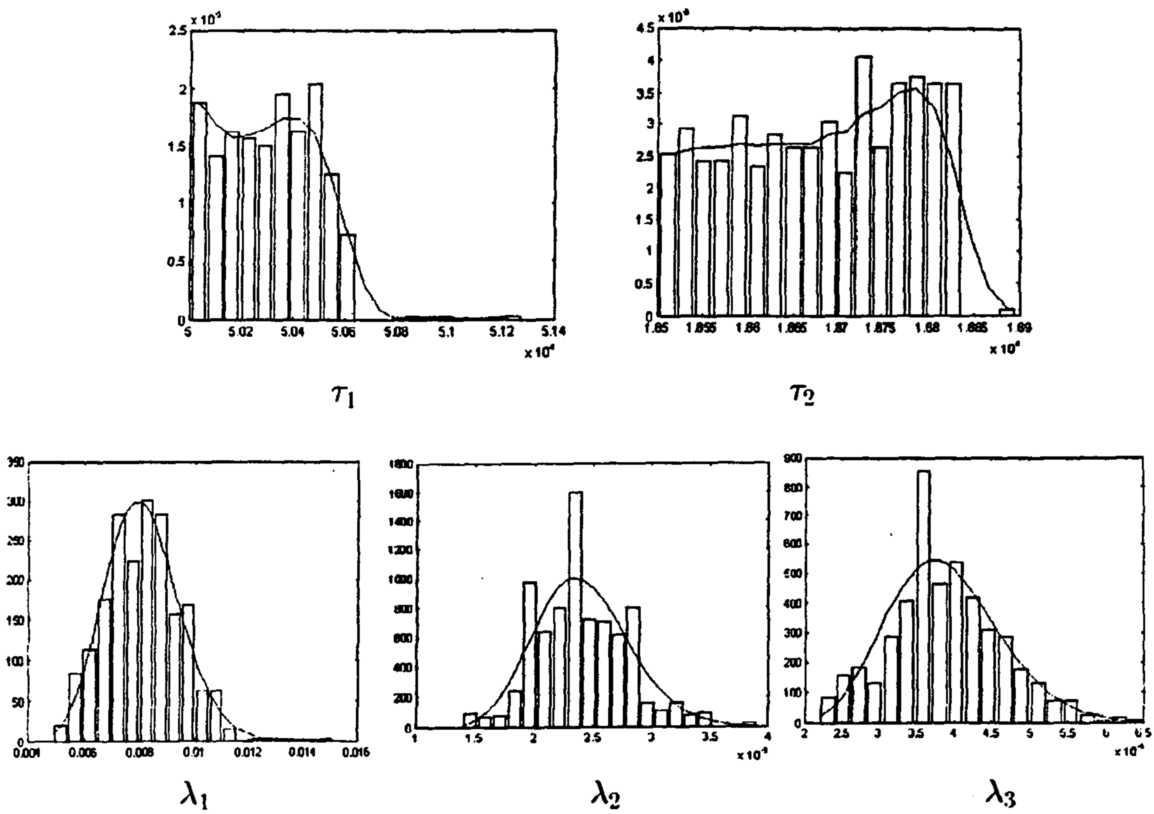

Figure 4 - Marginal posterior distributions (change-point model 2)

In figure 5, we have plots of the predictive densities $c_i = f(x_i|\underline{x}_{(i)})$, $i = 1, 2, \ldots, n$ approximated by the Monte Carlo estimates (17) for both models M_1(a single change-point model) and M_2 (two change-points model). We observe better fit for the data set of table 1 considering the two change-point model (2). For model M_1, we have

$P_1 = \prod_{i=1}^{n} \hat{c}_{1i} = 3.8649 \times 10^{-302}$ and for model M_2, we have $P_2 = \prod_{i=1}^{n} \hat{c}_{2i}$ $= 1.6319 \times 10^{-301}$.

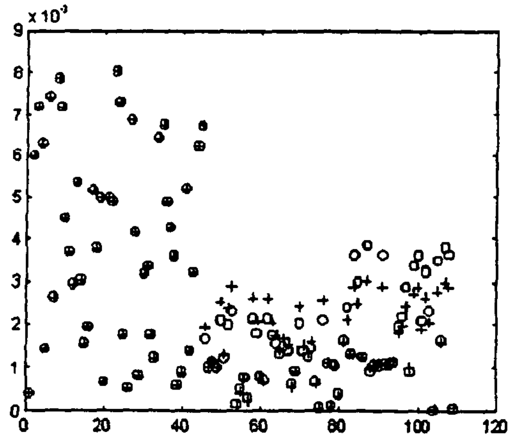

Figure 5 - Plot of c_i versus i (M_1+, M_2 o)

6. Overall conclusions

The use of Markov Chain Monte Carlo methods for a Bayesian analysis of homogeneous Poisson processes, in the presence of one or more change-points, is a suitable way to get accurate inferences for the parameters of the model.

Similar results could be obtained for more than two change-points.

The use of Monte Carlo estimates for the predictive densities $f(x_i|\underline{x}_{(i)})$, $i = 1, 2, \ldots, n$ based on the obtained Gibbs samples, give a simple way to discriminate the different change-point models, a problem of great practical interest.

REFERENCES

Achcar, J.A. ; Bolfarine, H. (1989), Constant hazard against a change-point alternative: a Bayesian approach with censored data. *Communications in Statistics - Theory and methods*, 18(10), 3801-3819.

Box, G.E. ; Tiao, G.C. (1973). *Bayesian inference in Statistical Analysis*, New York: Addison-Wesley.

Chib,S. ; Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, 49,4,327-335.

Chib,S. ; Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, 49,4,327-335.

Gelfand, A. E. ; Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

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

Maguire, B. A. ; Pearson, E. S. ; Wynn, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika*, 39, 168-180.

Matthews, D. E. ; Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics*, 38, 463-468.

Matthews, D. E. ; Farewell, V. T. ; Pyke, R. (1985). Asymptotic score statistic processes and tests for constant hazard against a change-point alternative. *Annals of Statistics*, 13, 2, 583-591.

Raftery, A. E. ; Akman, V. E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika*, 73, 1, 85-89.

Smith, A. F. M. ; Roberts, G. O. (1993). Bayesian computations via Gibbs Sampler and related Markov Chain Monte Carlo methods (with discussion). *Journal of Royal Statistical Society*, B, 55, 3-23.

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

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)

045/98 ACHCAR, J.A.; ANDRADE, M.G.; LOIBEL, S. - Weibull hazard function with a change-point: a bayesian approach using Markov chain Monte carlo methods.

044/97 ACHCAR, J.A.; PEREIRA, G.A. - Bayesian analysis of mixture models for survival data: some computational aspects.

043/97 ACHCAR, J.A.; PEREIRA, G.A. - Mixture models for type II censored survival data in the presence of covariates.

042/97 MOALA, F.A.; RODRIGUES, J. - A note on the prior distributions for the Weibull reliability function.

041/97 RODRIGUES, J. - Diagnostic of convergence of a Rao-Black Wellised estimate of the marginal density via calibrated divergence measures.

040/97 BARATELA, D.S.; RODRIGUES, J. - Uma caracterização da existência da posteriori marginal do parâmetro N do modelo de Jelinski-Moranda.