Test (2011) 20: 204–221
DOI 10.1007/s11749-010-0195-1

ORIGINAL PAPER

Objective Bayesian reference analysis for the Poisson process model in presence of recurrent events data

José Miguel Bernardo · Vera Lucia Tomazella

Received: 5 May 2009 / Accepted: 27 April 2010 / Published online: 16 June 2010

Abstract This article applies Bayesian reference analysis, widely considered as the most successful method to produce objective, model-based, posterior distributions, to a problem of inference in survival analysis. A formulation is considered where individuals are expected to experience repeated events, along with concomitant variables. The sampling distribution of the observations is modeled through a proportional intensity homogeneous Poisson process.

Keywords Reference analysis · Recurrent events · Intensity function · Poisson process · Bayesian inference

Mathematics Subject Classification (2000) 63F15 · 62N01

1 Introduction

The theory for survival data has been sufficiently developed to analyze the function of risk or survival of a patient. The methodology is designed to determine which variables affects the form of the risk function and to obtain estimates of these functions for each individual. This study involves following units (individuals) until the occurrence of some event of interest, for example, the fault (death) of the unit. Frequently, this event does occur for some units during the period of observation, thus producing

Communicated by Domingo Morales.

J.M. Bernardo
Departmento de Estadística, Universidad de Valencia, Valencia, Spain
e-mail: jose.m.bernardo@uv.es

V.L. Tomazella (✉)
Departmento de Estatística, Universidade Federal de São Carlos—UFSCar, São Paulo, Brazil
e-mail: vera@power.ufscar.br

censured data. Another characteristic of survival data is that some events of interest are not terminal, events which are able to occur more than once for the same individual, producing the recurrent events. Lifetime data where more than one event is observed on each subject arise in areas such as, biomedical studies, criminology, demography, manufacturing, and industrial reliability. For example, an offender may be convicted several times, several tumors may be observed for an individual; recurrent pneumonia episodes arise in patients with human immunodeficiency syndrome; a piece of equipment may experience repeated failures or warranty claims.

The data on the ith individual consists of the total number, m_i of the events observed over the time period $(0, T_i]$ and the ordered epoch of the m_i events, $0 \leq t_{i1} < t_{i2} < \cdots < t_{im_i} \leq T_i$. Additionally, we may have covariate information on each subject defined by a vector of censoring indicators. In some studies, interest may lie in understanding and characterizing the event which defines the process for individual subject or may focus on treatment comparisons based on the time to each distinct event, the number of events, the type of events, and interdependence between events. The idea is to explain the nature of the variation between subjects in terms of treatments, fixed covariates, or other factors as unobservable factors.

In this work the methodology is illustrated using the animals carcinogenicity data described by Gail et al. (1980) (Table 1). The experiment used 48 female rats mammary tumors. There were 23 rats in Group 1 (Treatment) and 25 rats in Group 2 (Control), and the data are the days on which new tumors occurred for each animal; a given animal may experience any number of tumors. The main objective of analysis was to assess the difference between treatment Groups 1 and 2 with regard to the development of tumors. The rats in study were all induced to remain tumor-free for 60 days and were observed over the period 60 to 182 days.

Several methodologies have been proposed to analyze the problem of recurrent events. Lawless and Nadeu (1995) apply the Poisson process to develop models that focus on the expected number of events occurred in a determined time interval. The development of statistical models based on counting process data were originally introduced by Aalen (1978). There is an extensive literature about point process models (e.g., Cox and Isham 1980); this approach offers tools powerful enough able to generalize several situations. In this article the problem is treated under the focus of punctual counting process. The Poisson process has been well studied, and many recent discussions about lifetime and stochastic process transition data have focused on modeling and analyzing the effects of so-called unobserved heterogeneity (e.g., Flinn and Heckman 1982).

From a Bayesian perspective there is an extensive literature. Main pointers are Ibrahim and Laud (1991), Ibrahim et al. (2001), Ibrahim and Chen (1998), Kim and Ibrahim (2001), and Tomazella (2003). It is well known that under a Bayesian perspective the outcome of any problem of inference is the posterior distribution of the quantity of interest, which combines the information provided by the data with available prior information; it has been often recognized that there is a pragmatically important need for a form of prior to posterior analysis which captures, in a well-defined sense, the notion that the prior should have a minimal effect, relative to the data, on the posterior inference.

The use of a prior function that somehow represents lack of prior knowledge about the quantity of interest, has been a constant in the history of the Bayesian inference.

Table 1 Days where new tumors occurred for 48 rats

N° of tumors	Treatment (23 rats)	N° of tumors	Control (25 rats)
1	182	7	63,102,119,161,161,172,179,182[a]
0	182[a]	11	88,91,95,105,112,119,119,137,145,167,172,182[a]
2	63,68,182[a]	9	91,98,108,112,134,137,161,179,182[a]
1	152,182[a]	2	71,174,182[a]
4	130,134,145,152,182[a]	9	95,105,134,134,137,140,145,150,150,182[a]
3	98,152,182	4	68,68,130,137,182[a]
6	98,95,105,130,137,167,182[a]	6	77,95,112,137,161,174,182[a]
1	152,182[a]	7	81,84,126,134,161,161,174,182[a]
1	181,182[a]	6	68,77,98,102,102,102,182[a]
5	71,84,126,134,152,182[a]	1	112,182[a]
2	116,130,182[a]	13	88,88,91,98,112,134,134,137,137,140,140,
1	91,182[a]		152,152,182[a]
5	63,68,84,95,152,182[a]	2	77,179,182[a]
2	105,152,182[a]	1	112,182[a]
3	63,102,152,182[a]	10	71,71,74,77,112,116,116,140,140,140,167,182[a]
4	63,77,112,140,182[a]	4	77,95,126,150,182[a]
5	77,119,152,161,167,182[a]	5	88,126,130,130,134,182[a]
5	105,112,145,161,182	11	63,74,84,84,88,91,95,108,134,137,179,182[a]
1	152,182[a]	11	81,88,105,116,123,140,145,152,161,161,179,182[a]
2	81,95,182[a]	9	88,95,112,119,126,126,150,157,179,182[a]
6	84,91,102,108,130,130,134,182[a]	12	68,68,84,84,102,105,119,123,123,137,161,179,182
0	182[a]	1	140,182[a]
1	91,182[a]	3	152,182,182
		1	81,182[a]
		3	63,88,134,182[a]
		3	84,134,182

[a]Indicates that in this time no new tumor was found

Reference analysis, introduced by Bernardo (1979) and futher developed by Berger and Bernardo (1989, 1992a, 1992b, 1992c) and Berger et al. (2009a, 2009b), is widely considered today the most successful algorithm to derive noninformative priors (Bernardo 1997). For a recent review see Bernardo (2005). Notice that reference priors are not proposed as an approximation to the scientist's (unique) personal beliefs, but rather as a collection of formal consensus (not necessarily proper) prior functions which could conveniently be used as standards for scientific communication.

Reference posteriors are obtained by formal use of the Bayes theorem with a reference prior function. If required, they may be used to provide point or region estimates, to test hypothesis (Bernardo 1998), or to predict the value of future observations. This provides a unified set of objective Bayesian solutions to the conventional problems of scientific inference, objective in the precise sense that those solutions only depend on the assumed model and the observed data.

In this paper, reference analysis is developed for a survival model based on proportional intensity Poisson process, where individuals may be expected to experience repeated events, and concomitant variables are observed.

Section 2 contains an overview of reference analysis, where the definition is motivated, and heuristic derivations of explicit expressions for the one-parameter, two-parameter, and multiparameter cases are sequentially presented. In Sect. 3 we describe the survival model. In Sect. 4, the theory is applied to an inference problem, the parameters survival model, for which no objective Bayesian analysis has been previously proposed, and which has been chosen because it combines intrinsic importance and pedagogic value. Some final concluding remarks are presented in Sect. 5.

2 An overview of reference analysis

The notion of a noninformative prior, that is, of a prior which describes lack of prior knowledge about the quantity of interest, has been the object of many debates within the Bayesian community; it is intended to be a prior function which by formal use of the Bayes theorem produces a posterior distribution dominated by the information provided by the data (Bernardo 1979; Berger et al. 2005). Objective reference priors do not depend on the data, but they depend on the probabilistic model that is assumed to have generated the data. The basic idea follows from the fact that the amount of information which may expected from an experiment clearly depends on prior knowledge: the larger the available prior knowledge, the lesser will be the amount of information to be expected form the data. An infinitelty large experiment would provide all information not yet available about the parameter of interest, all *missing information*; it is then natural to define a "noninformative" prior as that which maximizes the missing information about the parameter of interest. However, since "missing information" is defined as a limit which is not necessarily finite, the reference prior is actually defined as the limit of a sequence of priors which maximize increasingly large experiments. We now summarize its formal derivation. For details, see Bernardo (2005).

2.1 One parameter

Definition 1 Consider an experiment ε which consists of one observation $\mathbf{x}$ from $p(\mathbf{x}|\phi)$, $\phi \in \Phi \subset \Re$. Let $\mathbf{z}_k = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be the result of k independent replications of ε. Then, under suitable regularity conditions,

$$\pi_k(\phi) = \exp\left\{ \int_{X^k} p(\mathbf{z}_k|\phi) \log q(\phi|\mathbf{z}_k)\, d\mathbf{z}_k \right\}, \tag{1}$$

where $q(\phi|\mathbf{z}_k)$ is an asymptotic approximation to the posterior distribution $p(\mathbf{z}_k|\phi)$.

The reference posterior distribution is a function $\pi(\phi|\mathbf{x})$ such that

$$\lim_{k\to\infty}\left[\int_{\Phi} \pi_k(\phi|\mathbf{x}) \log\left\{ \frac{\pi_k(\phi|\mathbf{x})}{\pi(\phi|\mathbf{x})} \right\} d\phi \right] = 0,$$

where

$$\pi_k(\phi|\mathbf{x}) = \frac{p(\mathbf{x}|\phi)\pi_k(\phi)}{\int_\Phi p(\mathbf{x}\phi)\pi_k(\phi)\,d\phi}, \quad k = 1, 2, \ldots.$$

A reference prior ϕ is a function which, for any data, provides the reference posterior $\pi(\phi|\mathbf{x})$ by formal use of the Bayes theorem, i.e., a positive function $\pi(\phi)$ such that, for all $\mathbf{x} \in \mathbf{X}$,

$$\pi(\phi|\mathbf{x}) = \frac{p(\mathbf{x}|\phi)\pi(\phi)}{\int_\Phi p(\mathbf{x}|\phi)\pi(\phi)\,d\phi}. \tag{2}$$

Thus the reference prior $\pi(\phi)$ is the limit of the sequence $\{\pi_k(\phi),\ k = 1, 2, \ldots\}$ defined by (1) in the precise sense that the information-type limit of the corresponding sequence of posterior distributions $\{\pi_k(\phi|\mathbf{x}),\ k = 1, 2, \ldots\}$ is the posterior obtained from $\pi(\phi)$ by formal use of the Bayes theorem.

Proposition 1 (Reference priors under asymptotic normality) *Let $p(\mathbf{x}|\phi)$, $\mathbf{x} \in \mathbf{X}$, be a probability model with one real-valued parameter $\phi \in \Phi \subset \Re$. If the asymptotic posterior distribution of ϕ given k replications of the experiment is normal with standard deviation $s(\hat{\phi})$ such that $\hat{\phi}_k$ is a consistent and asymptotically sufficient estimator of the ϕ, then, the reference prior is given by*

$$\pi(\phi) \propto \left\{\frac{1}{s(\phi)}\right\},$$

where, under regularity conditions, $s(\phi) = h(\phi)^{-1/2}$, and $h(\cdot)$ is the Fisher information function.

2.2 One nuisance parameter

Consider now the case where the statistical model $p(\mathbf{x}|\phi,\omega)$, $(\phi,\omega) \in \Phi \times \Omega \subset \Re \times \Re$, contains one nuisance parameter, where the parameter of interest is ϕ, and the nuisance parameter is ω. We shall only consider here the regular case where joint posterior asymptotic normality may be established.

Proposition 2 *Let $p(\mathbf{x}|\phi,\omega)$, $(\phi,\omega) \in \Phi \times \Omega \subseteq \Re \times \Re$, be a probability model with two real-valued parameters ϕ and ω, where ϕ is the quantity of interest, and suppose that the joint posterior distribution of (ϕ,ω) is asymptotically normal with covariance matrix $S(\hat{\phi},\hat{\omega})$, where $(\hat{\phi},\hat{\omega})$ is a consistent estimator of (ϕ,ω). Let $H(\phi,\omega) = S^{-1}(\phi,\omega)$. Tipically, $H(\phi,\omega)$ is Fisher information matrix.*

(i) *The conditional reference prior of ω is*

$$\pi(\omega|\phi) \propto h_{22}(\phi,\omega)^{1/2}, \quad \omega \in \Omega(\phi).$$

(ii) *If $\pi(\omega|\phi)$ is not proper, a compact approximation $\{\Omega_i(\phi),\ i = 1, 2, \ldots\}$ to $\Omega(\phi)$ is required, and the reference prior of ω given ϕ is given by*

$$\pi_i(\omega|\phi) = \frac{h_{22}(\phi,\omega)^{1/2}}{\int_{\Omega_i(\phi)} h_{22}(\phi,\omega)^{1/2}\,d\omega}, \quad \omega \in \Omega_i(\phi).$$

(iii) *Within each* $\Lambda_i(\phi)$, *the marginal reference prior of* ϕ *is obtained as*

$$\pi_i(\phi) \propto \exp\left\{\int_{\Omega_i(\phi)} \pi_i(\omega|\phi)\log\left[s_{11}^{1/2}(\phi,\omega)\right]d\omega\right\},$$

where $s_{11}^{1/2}(\phi,\omega) = h_\phi(\phi,\omega) = h_{11} - h_{12}h_{22}^{-1}h_{21}$.

(iv) *The reference posterior distribution of* ϕ *given data* $\{x_1,\ldots,x_n\}$ *is*

$$\pi(\phi|x_1,\ldots,x_n) \propto \pi(\phi)\left\{\int_{\Omega(\phi)}\left\{\prod_{l=1}^{n} p(x_l|\phi,\omega)\right\}\pi(\omega|\phi)\,d\omega\right\}.$$

Corollary *If the nuisance parameter space* $\Omega(\phi) = \Omega$ *is independent of* ϕ *and if the functions* $s_{11}^{-1/2}(\phi,\omega)$ *and* $h_{22}^{1/2}(\phi,\omega)$ *factorize in the form*

$$\left\{s_{11}(\phi,\omega)\right\}^{-1/2} = f_1(\phi)g_1(\omega), \qquad \left\{h_{22}(\phi,\omega)\right\}^{1/2} = f_2(\phi)g_2(\omega),$$

then

$$\pi(\phi) \propto f_1(\phi), \qquad \pi(\omega|\phi) \propto g_2(\omega);$$

the reference prior relative the parametric value ordered (ϕ,ω) *is given by*

$$\pi(\omega,\phi) = f_1(\phi)g_2(\omega),$$

and in this case, there is no need for compact approximation, even if the conditional reference prior is not proper. For proof and details, see Bernardo (2005).

3 The multiparameter case

The approach to the nuisance parameter considered above was based on the use of an ordered parameterization whose first and second components were ϕ and ω, respectively, referred as the parameter of interest and the nuisance parameter. The reference prior for the ordered parameterization (ϕ,ω), was then sequentially derived to obtain $\pi_\omega(\phi,\omega) = \pi(\omega|\phi)\pi(\phi)$.

When the model parameter vector $\boldsymbol{\theta}$ has more than two components, this sequential conditioning idea can obviously be extended by considering $\boldsymbol{\theta}$ as an ordered parameterization, $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)$, and generating, by successive conditioning, a reference prior, relative to this ordered parameterization, of the form

$$\pi(\boldsymbol{\theta}) = \pi(\theta_m|\theta_1,\ldots,\theta_{m-1})\cdots\pi(\theta_2|\theta_1)\pi(\theta_1). \tag{3}$$

Proposition 3 *Let* $p(x|\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)$ *be a probability model with* m *real-valued parameters, let* θ_1 *be the quantity of interest, and suppose that the joint distribution of* $(\theta_1,\ldots,\theta_m)$ *is asymptotically normal with covariance matrix* $S(\hat{\theta}_1,\ldots,\hat{\theta}_m)$. *Then, if* S_j *is the* $j \times j$ *upper matrix of* S, $H_j = S_j^{-1}$, *and* $h_{jj}(\theta_1,\ldots,\theta_m)$ *is the* (j,j) *element of* H_j.

(i) *The conditional reference priors are*

$\pi(\theta_m|\theta_1,\ldots,\theta_{m-1}) \propto h_{mm}(\theta_1,\ldots,\theta_m)^{1/2}$ *for* $i = m-1, m-2, \ldots, 2,$ *and*

$$\pi(\theta_i|\theta_1,\ldots,\theta_{i-1}) \propto \exp\left[\int_{\Lambda_{i+1}} \cdots \int_{\Lambda_m} \log\left[h_{i+1,i+1}(\theta_1,\ldots,\theta_m)^{1/2}\right] \times \left\{\prod_{j=1}^{m} \pi(\theta_j|\theta_1,\ldots,\theta_{j-1})\right\} d\boldsymbol{\theta}_{i+1}\right],$$

where $d\boldsymbol{\theta}_j = d\theta_j \times \cdots \times d\theta_m$.

(ii) *The marginal reference prior of* θ_1 *is*

$$\pi(\theta_1) \propto \exp\left\{\int_{\Lambda_1} \cdots \int_{\Lambda_m} \log\left[s_{11}(\theta_1,\ldots,\theta_m)^{-1/2}\right] \times \left\{\prod_{j=1}^{m} \pi(\theta_j|\theta_1,\ldots,\theta_{j-1})\right\} d\boldsymbol{\theta}_1\right\}.$$

(iii) *After data* $\{x_1,\ldots,x_n\}$ *have been observed, the reference posterior distribution of the parameter of interest* θ_1 *is*

$$\pi(\theta_1|x_1,\ldots,x_n) \propto \pi(\theta_1) \propto \exp\left\{\int_{\Lambda_1} \cdots \int_{\Lambda_m} \prod_{j=1}^{m} p(x_l|\theta_1,\ldots,\theta_m) \times \left\{\prod_{j=1}^{m} \pi(\theta_j|\theta_1,\ldots,\theta_{j-1})\right\} d\boldsymbol{\theta}_1\right\}.$$

For proof and details, see Berger and Bernardo (1992a, 1992b, 1992c).

A commonly used objective prior in Bayesian analysis is Jeffreys prior (Jeffreys 1946, 1961). It is obtained by applying Jeffreys's rule, which is to take the prior density proportional to the square root of the determinant of the Fisher information matrix ($\sqrt{\det J(\phi,\omega)}$). However, Jeffreys himself was aware that often this does not work well in multiparameter settings.

4 Model formulation

Suppose that n individuals may experience a single type of recurrent event. Let m_i denote the number of events occurring for the ith individual. Assume that the ith individual is observed over the interval $(0, T_i]$, where T_i is determined independently of m_i. Let $0 \le t_{i1} < t_{i2} < \cdots < t_{im_i} \le T_i$, where the variables of interest t_{ij} denote the continuous failure times for the ith individual and the jth occurrence events ($i = 1,\ldots,n$ and $j = 1,\ldots,m_i$). Besides that, we are going to consider that each individual carries a covariate vector represented by $\mathbf{x}$, so data from ith individual consist of the total number of events m_i observed about a time period $(0, T_i]$ in the ordered occurrence, $t_{i1},\ldots,t_{im_i}$, and the covariate vector $\mathbf{x}$.

It is assumed that the repeated events of an individual with $k \times 1$ covariate vector $\mathbf{x}$ occur according to a nonhomogeneous Poisson process with intensity function given by

$$\lambda_{x_i}(t) = \lambda_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta}), \quad t \geqslant 0,\ i = 1, 2, \ldots, n, \tag{4}$$

where $\lambda_0(t)$ is a baseline intensity function, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$, and $\boldsymbol{\beta} = \beta_1, \ldots, \beta_k$ is a vector of unknown parameters.

The corresponding cumulative or integrated intensity function is

$$\Lambda_x(t) = \int_0^t \lambda_x(u)\,du = \Lambda_0(t)e^{\mathbf{x}'\boldsymbol{\beta}}, \tag{5}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)\,du$.

Methods of analysis will be considered semi parametric if $\lambda_0(t)$ is arbitrary and completely parametric if $\lambda_0(t)$ is specified by a parameter vector $\boldsymbol{\theta}$. In the case of the function of baseline hazard to be constant, this is a homogeneous Poisson process (see Cox and Isham 1980). The Poisson process model (4) is often known as the Cox proportional risk model (see Cox 1972).

Consider a parametric Poisson process where $\lambda_0(t) = \lambda_0(t;\boldsymbol{\theta})$. Then, the likelihood function for the model (4) for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ is given by (see Cox and Lewis 1996)

$$L(\boldsymbol{\theta},\boldsymbol{\beta}) = \prod_{i=1}^{n}\left\{\prod_{j=1}^{m_i}\lambda_{x_i}(t_{ij},\boldsymbol{\theta})\right\}\exp\bigl\{-\Lambda_{x_i}(T_i,\boldsymbol{\theta})\bigr\}, \tag{6}$$

which can be decomposed as $L(\boldsymbol{\theta},\boldsymbol{\beta}) = L_1(\boldsymbol{\theta})L_2(\boldsymbol{\theta},\boldsymbol{\beta})$, where

$$L_1(\boldsymbol{\theta}) = \left\{\prod_{i=1}^{n}\prod_{j=1}^{m_i}\frac{\lambda_0(t_{ij};\boldsymbol{\theta})}{\Lambda_0(T_i;\boldsymbol{\theta})}\right\} \quad \text{and}$$

$$L_2(\boldsymbol{\theta},\boldsymbol{\beta}) = \prod_{i=1}^{n}\exp\bigl[\Lambda_0(T_i;\boldsymbol{\theta})e^{\mathbf{x}_i'\boldsymbol{\beta}}\bigr]\bigl[\Lambda_0(T_i;\boldsymbol{\theta})e^{\mathbf{x}_i'\boldsymbol{\beta}}\bigr]^{m_i}.$$

The first likelihood kernel $L_1(\boldsymbol{\theta})$ arises from the conditional distribution of the event times, given the counts, and the second the likelihood kernel $L_2(\boldsymbol{\theta},\boldsymbol{\beta})$ arises form the Poisson distribution of the counts $m_1, \ldots, m_n$.

4.1 Modeling the baseline hazard

The exponential distribution is one of the simplest and more important probability distributions used in the modeling of data that represent the life time. It has been used intensively in the literature of survival and reliability, as, for example, in study areas on lifetime of items manufactured (see Davis 1952), in research involving survival or time of remission of chronic illnesses (see Feigl and Zelen 1965).

The exponential distribution has been extensively used to model the baseline hazard function due to its simplicity and flexibility. This is the particular case where

$$\lambda_0(t) = \lambda_0(t;\boldsymbol{\theta}) = \nu. \tag{7}$$

The corresponding intensity function and integrated intensity function are

$$\lambda_x(t) = \nu e^{\mathbf{x}'\boldsymbol{\beta}} \quad \text{and} \quad \Lambda_x(T) = T\nu e^{\mathbf{x}'\boldsymbol{\beta}}. \tag{8}$$

Considering the decomposition in (6), the likelihood function for ν, β is given by

$$L(\nu, \boldsymbol{\beta}) = \prod_{i=1}^{n} T_i^{-m_i} \exp\left[-\nu T_i\, e^{\mathbf{x}_i'\boldsymbol{\beta}}\right]\left[\nu T_i\, e^{\mathbf{x}_i'\boldsymbol{\beta}}\right]^{m_i}, \tag{9}$$

where $L_2(\nu, \boldsymbol{\beta}) = \prod_{i=1}^{n} \exp[-\nu T_i\, e^{\mathbf{x}_i'\boldsymbol{\beta}}][\nu T_i\, e^{\mathbf{x}_i'\boldsymbol{\beta}}]^{m_i}$ is the nucleus of the regression model for which m_i has a Poisson distribution with average and variance $E(m_i|\mathbf{x}_i) = \mathrm{Var}(m_i|\mathbf{x}_i) = \nu T_i\, e^{\mathbf{x}_i'\boldsymbol{\beta}}$.

The log-likelihood function (9) is given by

$$l(\nu, \boldsymbol{\beta}) \propto \sum_{i=1}^{n} m_i \log(\nu) + \sum_{i=1}^{n} m_i x_i'\boldsymbol{\beta} - \sum_{i=1}^{n} \nu T_i e^{\mathbf{x}'\boldsymbol{\beta}}. \tag{10}$$

Interval estimates and hypothesis tests for the parameters can be performed, in principle, by considering the asymptotic normal distribution of the maximum likelihood estimates (mle) and the asymptotic chi-squared distribution of the likelihood ratio statistics, respectively (Lawless 2002).

4.2 The Fisher information matrix

The posterior distribution of the parameter is often asymptotically normal (see, e.g., Bernardo and Smith 1994, Sect. 5.3). In this case, the reference prior is easily derived.

Considering the log of the likelihood function (10), we have the first and second derivatives given by

$$\begin{aligned}
\frac{\partial l}{\partial \nu} &= \sum_{i=1}^{n} \frac{m_i}{\nu} - \sum_{i=1}^{n} T_i e^{\mathbf{x}_i'\boldsymbol{\beta}}, \\
\frac{\partial l}{\partial \beta_r} &= \sum_{i=1}^{n} m_i x_{ir} - \sum_{i=1}^{n} \nu T_i x_{ir} e^{\mathbf{x}_i'\boldsymbol{\beta}}, \quad r = 0, 1, \ldots, k, \\
\frac{\partial l}{\partial \nu^2} &= -\sum_{i=1}^{n} \frac{m_i}{\nu^2}, \\
\frac{\partial l}{\partial \beta_r \beta_s} &= -\sum_{i=1}^{n} \nu T_i x_{ir} x_{is} e^{\mathbf{x}_i'\boldsymbol{\beta}}, \quad r, s = 0, 1, \ldots, k, \\
\frac{\partial l}{\partial \nu \partial \beta_r} &= -\sum_{i=1}^{n} T_i x_{ir} e^{\mathbf{x}_i'\boldsymbol{\beta}}.
\end{aligned}$$

The elements of the Fisher information matrix are given by

$$I_{\nu\nu} = E\left[-\frac{\partial l(\nu,\beta)}{\partial \nu^2}\right] = \sum_{i=1}^{n} E\left[\sum_{i=1}^{n}\frac{m_i}{\nu^2}\right] = \frac{1}{\nu}\sum_{i=1}^{n} T_i e^{\mathbf{x}_i'\boldsymbol{\beta}},$$

$$I_{\beta_r\beta_s} = E\left[-\frac{\partial l(\nu,\beta)}{\partial \beta_r\beta_s}\right] = \nu\sum_{i=1}^{n} T_i x_{ir}x_{is}e^{\mathbf{x}_i'\boldsymbol{\beta}}, \quad r,s = 0,1,\ldots,k,$$

$$I_{\beta_r\nu} = E\left[-\frac{\partial l(\nu,\beta)}{\partial \beta_r\nu}\right] = \sum_{i=1}^{n} T_i x_{ir}e^{\mathbf{x}_i'\boldsymbol{\beta}}, \quad r = 0,1,\ldots,k.$$

Thus, the Fisher information matrix associated with the model is given by

$$H(\boldsymbol{\theta}) = H(\nu,\beta) = \begin{bmatrix} \frac{1}{\nu}\sum_{i=1}^{n} T_i e^{\mathbf{x}_i'\boldsymbol{\beta}} & \sum_{i=1}^{n} T_i x_{ir} e^{\mathbf{x}_i'\boldsymbol{\beta}} \\ \sum_{i=1}^{n} T_i x_{ir} e^{\mathbf{x}_i'\boldsymbol{\beta}} & \nu\sum_{i=1}^{n} T_i x_{ir}x_{is}e^{\mathbf{x}_i'\boldsymbol{\beta}} \end{bmatrix}. \tag{11}$$

5 Reference analysis for survival model parameters

Following the methodology described in Sect. 2, now we derive the reference prior considering two groups, which corresponds to the ordered partition $\{\nu, \boldsymbol{\beta}\}$, where $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_k\}$, and ν is considered to be the parameter of interest. The reference prior relative to this ordered parameterization is then:

$$\pi(\nu,\boldsymbol{\beta}) = \pi(\boldsymbol{\beta}|\nu)\pi(\nu).$$

From Corollary of Proposition 3 where the nuisance parameter space $\Lambda(\boldsymbol{\beta}) = \Lambda$ is independent of ν, it is easy to see that

$$\pi(\boldsymbol{\beta}|\nu) = |h_{22}|^{1/2} = \nu^{1/2}\left[\sum_{i=1}^{n} T_i x_{ir}x_{is}e^{\mathbf{x}_i'\boldsymbol{\beta}}\right]^{1/2} = f_1(\nu)g_1(\boldsymbol{\beta})$$

and

$$\begin{aligned} h_\nu(\nu,\boldsymbol{\beta}) &= h_{11} - h_{12}h_{22}^{-1}h_{21} = \nu^{-1/2}\left[\sum_{i=1}^{n} T_i e^{\mathbf{x}_i'\boldsymbol{\beta}} - \frac{[\sum_{i=1}^{n} T_i x_{ir}e^{\mathbf{x}_i'\boldsymbol{\beta}}]^2}{\sum_{i=1}^{n} T_i x_{ir}x_{is}e^{\mathbf{x}_i'\boldsymbol{\beta}}}\right]^{1/2} \\ &= f_2(\nu)g_2(\boldsymbol{\beta}). \end{aligned}$$

This implies that the conditional reference prior of the nuisance parameter $\boldsymbol{\beta}$ given the parameter of interest ν is

$$\pi(\boldsymbol{\beta}|\nu) \propto g_1(\boldsymbol{\beta}) = \left[\sum_{i=1}^{n} T_i x_{ir}x_{is}e^{\mathbf{x}_i'\boldsymbol{\beta}}\right]^{1/2}. \tag{12}$$

The reference prior needed to obtain a reference posterior for the parameter of interest ν is

$$\pi(\nu) \propto f_2(\nu) = \nu^{-1/2}. \tag{13}$$

Figure 1 represents the reference prior (13).

It follows that the joint reference prior for parameters ν and $\boldsymbol{\beta}$ is given by

$$\pi(\nu, \boldsymbol{\beta}) = \pi(\boldsymbol{\beta}|\nu)\pi(\nu) \propto \left[\sum_{i=1}^{n} T_i x_{ir} x_{is} e^{\mathbf{x}_i' \boldsymbol{\beta}}\right]^{1/2} \nu^{-1/2}. \tag{14}$$

Figure 2 represents the joint reference prior (14), considering in the particular case where $T = 100$, $n = 30$, $x = 1$.

Fig. 1 Reference prior for the parameter ν

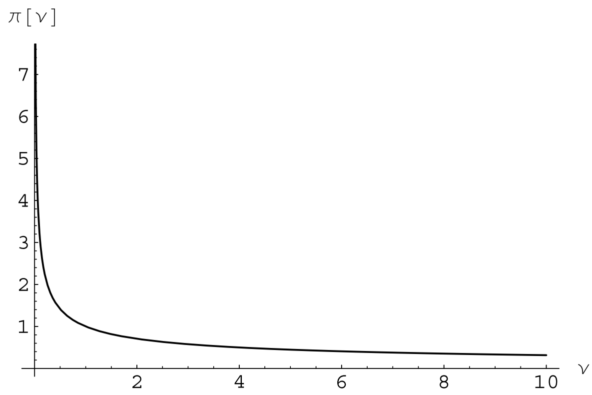
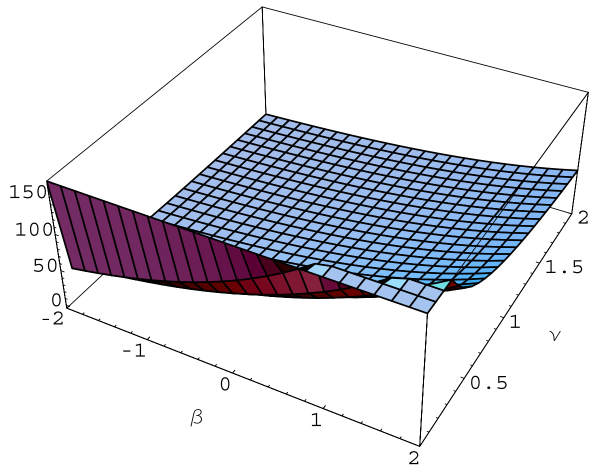


Fig. 2 Joint reference prior for the parameters ν and β



The corresponding reference posterior for ν after data $\mathbf{t} = \{\mathbf{t}_1, \ldots, \mathbf{t}_n\}$ have been observed is

$$\pi(\nu|\mathbf{t}_1, \ldots, \mathbf{t}_n) \propto \pi(\nu) \int\limits_{\Lambda} L(\nu, \boldsymbol{\beta})\pi(\boldsymbol{\beta}|\nu)\, d\boldsymbol{\beta}$$

$$\propto \nu^{-1/2} \int_{\Lambda} \prod_{i=1}^{n} T_i^{-m_i} \exp\left[-\nu T_i\, e^{\mathbf{x}_i'\boldsymbol{\beta}}\right]\left[\nu T_i\, e^{\mathbf{x}_i'\boldsymbol{\beta}}\right]^{m_i}$$

$$\times \left[\sum_{i=1}^{n} T_i x_{ir} x_{is} e^{\mathbf{x}_i'\boldsymbol{\beta}}\right]^{1/2} d\boldsymbol{\beta}. \qquad (15)$$

For this model, the joint Jeffrey's prior for parameters ν and $\boldsymbol{\beta}$ is given by

$$\pi(\nu, \boldsymbol{\beta}) = \left|\det H(\nu, \beta)\right|^{1/2} \propto \left[\sum_{i=1}^{n} T_i^2 x_{ir} x_{is} \left(e^{\mathbf{x}_i'\boldsymbol{\beta}}\right)^2 - \left(\sum_{i=1}^{n} T_i x_{ir} e^{\mathbf{x}_i'\boldsymbol{\beta}}\right)^2\right]^{1/2}. \qquad (16)$$

Note that the joint reference prior depends on ν and β, while the Jeffreys' prior depends only on β.

The marginal reference posterior densities (15) cannot be obtained explicitly. We overcome this difficulty by making use of the Markov Chain Monte Carlo (MCMC) methodology to obtain approximations for such densities. In order to make Bayesian inference for the parameters of interest ν, we implement the MCMC methodology considering the Metropolis–Hastings (see Hastings 1970; Chib and Greenberg 1995).

6 Example with simulated data

To analyze the behavior of the model proposed, a set of recurrent events data was simulated for different samples sizes, $n = 10$, $n = 30$, $n = 50$, and $n = 100$, with dichotomous covariate $\mathbf{x}$ equal to 0 and 1, indexing a control group and treatment group, respectively. In this study we considered the number of recurrent m generated from a Poisson distribution with mean $\lambda(t) = \nu e^{\beta_0 + x\beta}$ where $\nu = 2$, $\beta_0 = -0.15$, and $\beta = -0.5$. The lifetime was generated from a exponential distribution with parameter $\nu = 2$. The total time of observation $T = 2$ was considered for all samples. We present properties of estimators on reference prior and Jeffreys' prior proposed here, where we focus on the coverage probability and the 95% credible intervals.

Considering the simulated data set, a sample of the reference posterior (15) was obtained generated by the Metropolis–Hastings technique, i.e., through Markov chain Monte Carlo methods implemented in software R. The convergence of the chains were assessed according to convergence diagnostics implemented in CODA (Best et al. 1997). Graphical traces of those methods and kernel density estimation for the parameters ν, β_0, and β showed that there were no convergence problems. We generated two chains of 50,000 iterations each for the model parameters. The first 2,000 iterations were ignored (burn-in). A sample size of 4,000 elements was considered.

Table 2 Estimated parameters of the posterior distribution considering the reference prior (14)

n	ν				β_0				β_1			
	Mean	SD	IC(95%)	C	Mean	SD	IC(95%)	C	Mean	SD	IC(95%)	C
10	1.95	0.31	[1.34; 2.60]	0.82	−0.11	0.01	[−0.63; 0.31]	0.88	−0.48	0.23	[−0.50; 0.47]	0.90
30	1.96	0.32	[1.37; 2.64]	0.85	−0.12	0.01	[−0.63; 0.32]	0.90	−0.49	0.24	[−0.52; −0.48]	0.95
50	1.96	0.32	[1.37; 2.64]	0.90	−0.15	0.01	[−0.64; 0.33]	0.96	−0.50	0.24	[−0.53; 0.49]	0.96
100	1.96	0.32	[1.37; 2.64]	0.95	−0.15	0.01	[−0.64; 0.33]	0.96	−0.50	0.24	[−0.53; −0.49]	0.96

Table 3 Estimated parameters of the posterior distribution considering the Jeffreys' prior (16)

n	ν				β_0				β_1			
	Mean	SD	IC(95%)	C	Mean	SD	IC(95%)	C	Mean	SD	IC(95%)	C
10	1.90	0.30	[1.30; 2.60]	0.80	−0.10	0.01	[−0.62; 0.32]	0.80	−0.44	0.21	[−0.49; 0.46]	0.80
30	1.93	0.31	[1.32 ; 2.64]	0.82	−0.11	0.01	[−0.63; 0.32]	0.88	−0.45	0.22	[−0.50; −0.47]	0.85
50	1.94	0.32	[1.35; 2.65]	0.88	−0.10	0.01	[−0.65; 0.33]	0.95	−0.51	0.24	[−0.52; 0.49]	0.86
100	1.96	0.32	[1.37; 2.64]	0.88	−0.15	0.01	[−0.64; 0.33]	0.95	−0.51	0.24	[−0.53; −0.49]	0.90

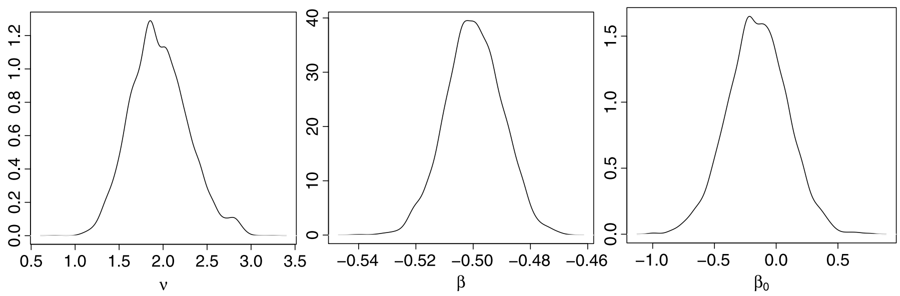


Fig. 3 Density of ν, β and β_0 for $n = 30$

Tables 2 and 3 show the summary posterior of the interest parameters with simulated data, considering reference prior and Jeffreys prior, respectively. We observe that the empirical coverage probability considering a reference a priori is closer to the nominal level than Jeffreys prior; we also observe that the coverage probability approaches its nominal value as sample size increases. Also, the estimate average considering Jeffreys' prior and the true values is almost always smaller than one considering reference prior, but the two prior distributions provide estimated values very close to the true value, which indicates that the results through simulation are adequate.

Figures 3 and 4 show plots of the generated samples and the empirical marginal posteriors for model parameter ν, β, and β_0 considering the sample size $n = 30$ based on the generated chains of the marginal reference posterior (15).

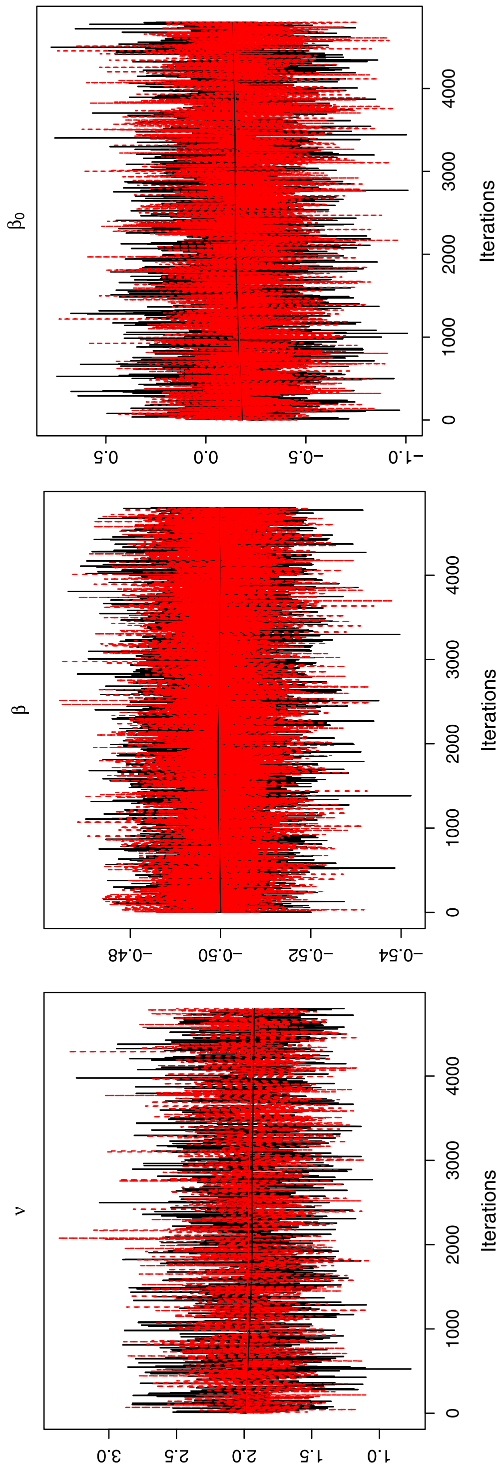


Fig. 4 Trace of ν, β, and β_0 for $n = 30$

7 Example with the animal carcinogenesis data

The methodology is illustrated on animal carcinogenesis data from Gail et al. (1980), Table 1. The lifetimes are days on which new tumors occurred for each animal. Each animal may experience different number of tumors. The main objective of study was to assess the difference between Groups 1 and 2 regarding to the development of tumors. The rats were induced to remain tumor-free for 60 days and were observed over the period from 60 to 182 days.

Suppose that the ith rat has intensity of tumors occurring according to (8). The covariate vector has just one covariate $\mathbf{x}$ indicating whether the individual i is in Group 1 ($x_i = 1$) or in Group 2 ($x_i = 0$), for $i = 1, \ldots, 48$. As in the study the lifetimes were observed after 60 days, the intervals of observation for all animals are indeed $(0, T_i) = (0, 122)$.

The sample for the reference posterior distributions (15) of the parameters ν and the regression parameters β and β_0 were obtained by the Metropolis–Hastings technique, i.e., through Markov Chain Monte Carlo methods implemented in software R. The convergence of the chains were tested by using the Gelman and Rubin method (e.g., Gelman et al. 1995) implemented in CODA (Best et al. 1997). Graphical traces of those methods and kernel density estimation for each parameter showed that there were no convergence problems. We generated two chains of 50,000 iterations, each

Table 4 Posterior summaries for the parameters using reference prior

Parameters	Mean	$D.P$	IC(95)
ν	0.0991	0.0083	[0.0842; 0.1157]
β_0	−0.9991	0.0493	[−1.0945; −0.9012]
β	−0.0517	0.005	[−0.0616; −0.0416]

Table 5 Posterior summaries for the parameters using Jeffreys' prior

Parameters	Mean	$D.P$	IC(95)
ν	0.0992	0.0081	[0.084; 0.116]
β_0	−0.9994	0.0487	[−1.0932; −0.9017]
β	−0.0515	0.0049	[−0.0610; −0.0418]

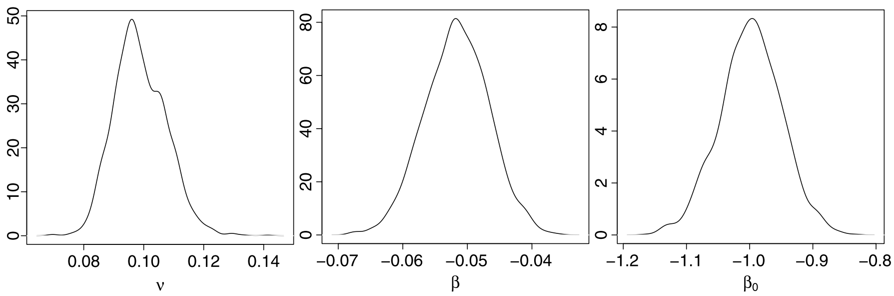


Fig. 5 Density of ν, β, and β_0 for the real data

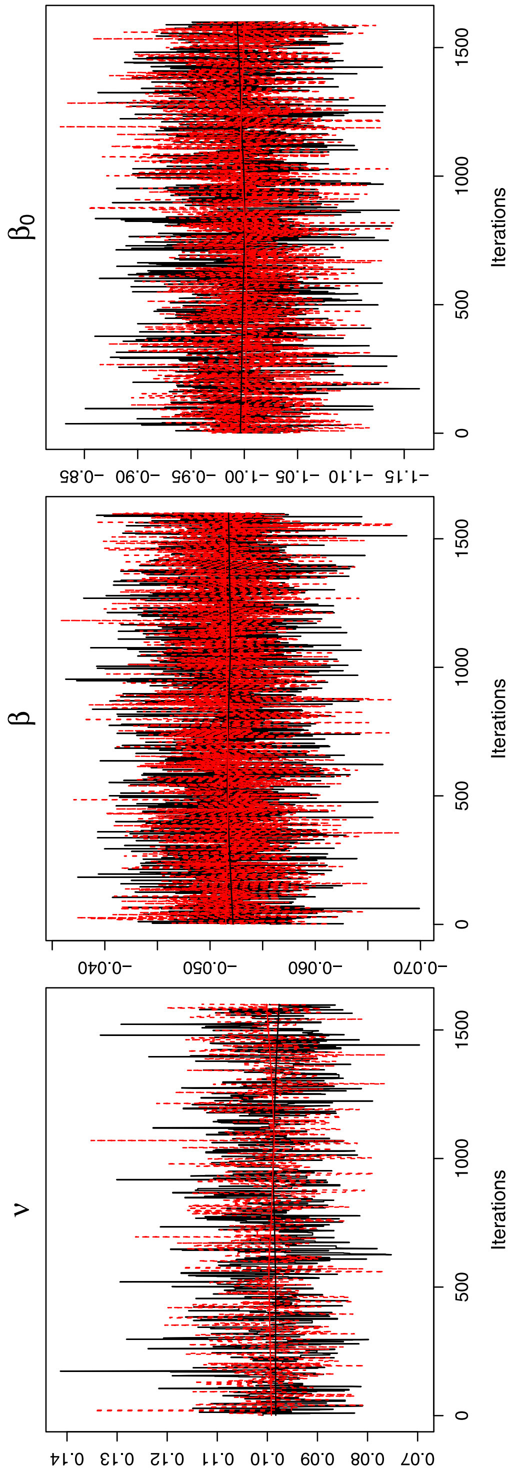


Fig. 6 Trace of ν, β, and β_0 for the real data

for the model parameters. The first 2,000 iterations were ignored (burn-in). For each parameter, we considered a sample size of 4,000 elements.

The posterior results form using both the reference prior and the multivariate Jeffreys prior are shown in Tables 4 and 5, respectively. It may be noticed that the posterior results are all very similar. It is important to note that there is significant difference between Groups 1 and 2 regarding to the development of tumors indicated by the estimated β, which is equal to (-0.0517) using the reference prior and equal to (-0.0515) using Jeffreys prior.

In Figs. 5 and 6, we show plots of the generated samples and the empirical marginal posteriors for model parameter ν, based on the generated chains of the marginal reference posterior (15).

8 Concluding remarks

In this paper, we have summarized the definition and derivation of reference posterior, and we have illustrated the theory with an important example in survival analysis; we have also mentioned some results that may be used to substantiate the claim that they constitute the more promising available method to derive *nonsubjective* prior distributions. We have developed a reference posterior distribution for one of the parameters of the Poisson process model for recurrent events data, so that researchers with a subjective initial information can compare their posterior distributions with the reference distribution. The same technique can be developed for other parameter of interest in the model, for example, the parameter β that measures the effect of treatment. Using both simulated and real data, we have provided evidence that the derived reference prior may well be recommended for objective Bayesian inference in the Possion process model.

References

Aalen OO (1978) Nonparametric inference for a family of counting processes. Ann Stat 6:701–726

Berger JO, Bernardo JM (1989) Estimating a product of means: Bayesian analysis with reference priors. J Am Stat Assoc 84:200–207

Berger JO, Bernardo JM (1992a). Ordered group reference priors with applications to a multinomial problem. Biometrika **79**, 25–37

Berger JO, Bernardo JM (1992b) Reference priors in a variance components problem. In: Goel PK, Iyengar NS (eds) Bayesian analysis in statistics and econometrics. Springer, Berlin, pp 323–340

Berger JO, Bernardo JM (1992c) On the development of reference priors. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics, vol 4. Oxford University Press, London, pp 35–60 (with discussion)

Berger JO, Bernardo JM, Sun D (2009a) Natural induction: an objective Bayesian approach. Rev Acad Sci Madrid A 103:125–159

Berger JO, Bernardo JM, Sun D (2009b) The formal definition of reference priors. Ann Stat 37:905–938

Berger JO, Bernardo JM, Sun D (2005) Reference priors from first principles: a general definition. Tech rep, SAMSI, NC, USA

Bernardo JM (1979) Reference posterior distributions for Bayesian inference. J R Stat Soc B 41:113–147 (with discussion). Reprinted in: Polson NG, Tiao GC (eds) Bayesian inference. Edward Elgar, Brookfield (1995), pp 229–263

Bernardo JM (1997) Noninformative priors that do not exist. J Stat Plan Inference 65:159–189 (with discussion)

Bernardo JM (1998) Nested hypothesis testing: the Bayesian reference criterion. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics, vol 6. Oxford University Press, London, pp 49–64

Bernardo JM (2005) Reference analysis. In: Dey DK, Rao CR (eds) Handbook of statistics, vol 25. Elsevier, Amsterdam, pp 17–90

Bernardo JM, Smith AFM (1994) Bayesian theory. Wiley, Chichester

Best N, Cowles M, Vines K (1997). CODA—convergence diagnosis and output analysis software for Gibbs sampling output: version 0.4. MRC Biostatistics Unit, Cambridge, UK

Chib S, Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. Am Stat 49:327–335

Cox DR (1972) Regression models and life tables. J R Stat Soc B 34:187–220 (with discussion)

Cox DR, Isham V (1980) Point processes. Chapman & Hall, London

Cox DR, Lewis PAW (1996) The statistical analysis of series of events. Methuen, London

Davis DJ (1952) An analysis of some failure data. J Am Stat Assoc 47:113–150

Feigl P, Zelen M (1965) Estimation of exponential survival probabilities with concomitant information. Biometrics 21:826–837

Flinn CJ, Heckman JJ (1982) Models for the analysis of labor force dynamics. Adv Econom 1:35–95

Gail MH, Santner TJ, Brown CC (1980) An analysis of comparative carcinogenesis experiments based on multiple times to tumor. Biometrics 36:255–266

Gelman R, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman & Hall, London

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Ibrahim JG, Chen MH (1998) Prior distributions and Bayesian computation for proportional hazards models. Sankhya B 60:48–64

Ibrahim JG, Laud PW (1991) On Bayesian analysis of generalized linear models using Jeffreys' prior. J Am Stat Assoc 86:981–986

Ibrahim JG, Chen MH, Sinha D (2001) Bayesian survival analysis. Springer, New York

Jeffreys H (1946) An invariant form for the prior probability in estimation problems. Proc R Soc Lond, Ser A 196:453–461

Jeffreys, H (1961) Theory of probability, 3rd edn. Oxford University Press, Oxford

Kim SW, Ibrahim JG (2001) On Bayesian inference for proportional hazards models using noninformative priors. Lifetime Data Anal 6:331–341

Lawless JF (2002) Statistical models and methods for life time data, 2nd edn. Wiley, New York

Lawless JF, Nadeu C (1995) Some simple robust methods for the analysis of recurrent events. Technometrics 37:158–168

Tomazella VLD (2003). Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade. Tese de Doutorado, Universidade de São Paulo-ICMC-São Carlos