Computational Statistics and Data Analysis 56 (2012) 75–87

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Random effects in promotion time cure rate models

Celia Mendes Carvalho Lopes [a,b,*], Heleno Bolfarine [b]

[a] Universidade Presbiteriana Mackenzie, Brazil

[b] Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP), Brazil

ARTICLE INFO

Article history:
Received 8 August 2008
Received in revised form 16 May 2011
Accepted 17 May 2011
Available online 25 May 2011

Keywords:
Long-term survivors
Random effects
REML
Metropolis–Hastings

ABSTRACT

In this paper, a survival model with long-term survivors and random effects, based on the promotion time cure rate model formulation for models with a surviving fraction is investigated. We present Bayesian and classical estimation approaches. The Bayesian approach is implemented using a Markov chain Monte Carlo (MCMC) based on the Metropolis–Hastings algorithms. For the second one, we use restricted maximum likelihood (REML) estimators. A simulation study is performed to evaluate the accuracy of the applied techniques for the estimates and their standard deviations. An example on an oropharynx cancer study is used to illustrate the model and the estimation approaches considered in the study.

1. Introduction

In survival analysis studies, the time until an event occurs is evaluated assuming that it is supposed to occur for all subjects under study. However, there are situations where some subjects will never experience the event. For example, if a group of patients with a certain type of cancer were assigned to a treatment and followed until the resurgence of the disease, a portion of them may have relapses and another portion may not present the recurrence of the event. Hence, there is great interest in studying survival analysis with a cure rate fraction, which includes the presence of immunity in the study.

When analyzing some survival data, it is sometimes noticed that there are a large number of subjects with right censored survival times. This fact may be evidence that, to some subjects under investigation, the event of interest may never happen. In other words, a fraction of the population might be immune (or cured).

Many researchers have been studying survival models with a cure fraction. More details can be found in Maller and Zhou (1996). These models have been called cure rate models or long-term survival models. In 1952, the researchers Berkson and Gage published an article (Berkson and Gage, 1952) in which they present a mixture type model to study the cases where there is a proportion p of cured subjects. This model has been studied by many different authors, such as Goldman (1984), Greenhouse and Wolfe (1984), Sposto et al. (1992) and Mizoi et al. (2007), among others.

Another approach for models with long-term survivors appears in Yakovlev and Tsodikov (1996) and Chen et al. (1999), in which the long-term survival times are modeled based upon the non-observable number of susceptible cells that a subject has that might present a detectable cancer. This model, presented in Section 2, will be referred to as the promotion time cure rate model. Many articles have been presented using this model approach. For example, we can mention Asselain et al. (1996), Tsodikov (1998a,b), Tsodikov et al. (1998), Chen et al. (2002a,b), Ibrahim et al. (2001a,b) and Yin (2005).

Even more recently, some authors are proposing approaches that unify the mixture and the promotion time models for long-term survivors. It can be found in Yin and Ibrahim (2005) and Rodrigues et al. (2009).

* Corresponding author. Tel.: +55 11 30229395; fax: +55 11 30229395.
E-mail addresses: celiagiz@yahoo.com.br, celia.lopes@mackenzie.br, celiagiz@mackenzie.br, celiagiz@ime.usp.br (C.M. Carvalho Lopes), hbolfar@ime.usp.br (H. Bolfarine).

0167-9473/$ – see front matter © 2011 Published by Elsevier B.V.
doi:10.1016/j.csda.2011.05.008

It is quite common in survival analysis for the data to come from different groups, for example, several clinics. Often, the results in each of the clinics can be distinct. In such cases, it might be appropriate to study models with random effects, so that we can evaluate the effect of the clinic in the treatment, as it is the case, for example, with cancer studies. Yau and Ng (2001) incorporate random effects in mixture models for long-term survivors and they use only the classical approach.

In this present work, it is proposed to include random effects in the promotion time cure rate model. The estimation is made by classical and Bayesian methods. Also presented is a simulation study for both approaches. To illustrate an application of the estimation approaches, an analysis of the data presented in Kalbfleisch and Prentice (1980, pp. 225–229) is shown, which was also studied in Yau and Ng (2001).

The paper is organized as follows. In Section 2, the ordinary promotion time cure rate model is defined and its main properties are studied. In Section 3, the model defined in Section 2 is extended to incorporate random effects. In Section 3.1, the Bayesian estimation approach is developed for the model defined in the beginning of Section 3. Since a full Bayesian approach is not possible, a Markov chain Monte Carlo approach for sampling from the joint posterior distribution is implemented. In Section 3.2, a classical estimation approach is developed, with the use of BLUP and REML estimators. In Section 4, a simulation study is conducted to illustrate the performance of the estimators presented in the previous section. In Section 5, a real data analysis is presented to illustrate the use of the model proposed in Section 2. The paper is concluded with a main discussion section.

2. Promotion time cure rate model

The model proposed in Chen et al. (1999) is defined as follows. Let N be the number of cells susceptible to develop a cancer (called carcinogenic cells). Suppose that N has a Poisson distribution with mean θ. Let Z_c be the random variable for the time until the c-th carcinogenic cell produces a detectable cancer. The variables Z_c, $c = 1, 2, \ldots, N$, are assumed to be independent and identically distributed (iid) with distribution function $F(t) = 1-S(t)$ and are also independent of N. Thus, the detectable time for the cancer can be defined by the random variable $T = \min\{Z_c, 0 \leq c \leq N\}$ where $P(Z_0 = \infty) = 1$ and N is independent of the sequence $Z_1, Z_2, \ldots$. Consequently, it can be shown that the survival function for T is given by the limit

$$S_p(t) = P(T > t) = \mathrm{e}^{-\theta F(t)}.$$

Note that the cure fraction is calculated by the limit

$$S_p(\infty) = \lim_{t\to\infty} S_p(t) = \lim_{t\to\infty} \mathrm{e}^{-\theta F(t)} = \mathrm{e}^{-\theta}.$$

For the subject j, with $j = 1, 2, \ldots, n$, consider $y_j = \min\{T_j, C_j\}$ in which $T_j = \min\{Z_{j0}, Z_{j1}, \ldots, Z_{jN_j}\}$, where $N_j \sim$ Poisson (θ_j) and C_j is a right censored variable. Also, consider that the noninformative censoring variables C_j are independent. And, let the failure indicator function be ν_j. For $N_j = n_j$, Chen et al. (1999), Mizoi (2004) and Mizoi et al. (2007) show that

$$f(y_j, \nu_j|n_j) = S(y_j)^{n_j-\nu_j}[n_j f(y_j)]^{\nu_j}. \tag{1}$$

We can further extend the model above by using covariates $\mathbf{x}_j$ to model θ_j. Since θ_j is positive, we can consider the transformation $\theta_j = \mathrm{e}^{\mathbf{x}_j^\top \boldsymbol{\gamma}}$, for example.

Moreover, we can specify a parametric model for the survival times. In this case, we have $f(y_j) = f(y_j|\boldsymbol{\lambda}_j)$. Often, according to Chen et al. (1999), the chosen distribution is the Weibull or gamma distribution. Note that in such cases, $\boldsymbol{\lambda}_j = (\lambda, \alpha)$. The Weibull distribution used in Chen et al. (1999) has a density function which can be written as

$$f(y) = \alpha y^{\alpha-1} \exp\{\lambda - y^\alpha \exp(\lambda)\},$$

in which $\alpha > 0$ and $\lambda \in R$.

To model $\boldsymbol{\lambda}_j$ we can also include covariates $\mathbf{x}_j$. For example, $\boldsymbol{\lambda}_j = \mathbf{x}_j^\top \boldsymbol{\beta}$.

3. Promotion time cure rate model with random effects

Suppose that cases were studied in M clinics and that for the k-th clinic there were n_k subjects, $k = 1, 2, \ldots, M$. Consider also N_{jk} as the number of carcinogenic cells that are present in the subject j from the clinic k, for $j = 1, 2, \ldots, n_k$. Further, suppose that N_{jk} are independent and follow a Poisson distribution with mean θ_{jk}. For each $k = 1, 2, \ldots, M$, define the vector $\mathbf{N}_k = [N_{1k}, N_{2k}, \ldots, N_{n_k k}]^\top$.

Let Z_{cjk} be a random variable for the time until the c-th carcinogenic cell of subject j from clinic k produces a detectable cancer. Suppose that the variables Z_{cjk} are iid with distribution function $F(t|\boldsymbol{\lambda}) = 1 - S(t|\boldsymbol{\lambda})$ and are also independent of N_{jk}. Thus, the time for the cancer to be detectable can be defined by the random variable $T_{jk} = \min\{Z_{cjk}, 1 \leq c \leq N_{jk}\}$ but $T_{jk} = \infty$ with probability 1, if $N_{jk} = 0$.

In this article, we will adopt parametric forms for $F(\cdot)$, such as, for example, $y \sim$ Weibull distribution. In this case, $\boldsymbol{\lambda} = (\lambda, \alpha)$, $f(y|\boldsymbol{\lambda}) = \alpha y^{\alpha-1}\mathrm{e}^{\lambda - y^\alpha \mathrm{e}^\lambda}$ and $S(y|\boldsymbol{\lambda}) = \mathrm{e}^{-y^\alpha \mathrm{e}^\lambda}$, with $\alpha > 0$ and $\lambda \in R$.