Computational Statistics and Data Analysis 91 (2015) 51–63

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Long-term survival models with overdispersed number of competing causes

Wagner Barreto-Souza

Departamento de Estatística, Universidade Federal de Minas Gerais, Pampulha, 31270-901, Belo Horizonte, MG, Brazil

ARTICLE INFO

Article history:
Received 20 April 2014
Received in revised form 27 April 2015
Accepted 6 May 2015
Available online 18 May 2015

Keywords:
Cure rate models
Lifetime
Exponential family
Mixed Poisson distributions
Heterogeneous promotion models

ABSTRACT

We introduce a new class of long-term survival models by assuming that the number of competing causes, say N, belongs to a class of mixed Poisson distributions, which are overdispersed. More specifically, we suppose that $N|Z$ follows a Poisson distribution with mean λZ, with $\lambda > 0$, and Z is a positive continuous random variable belonging to the exponential family. With this, we obtain a general class for N, which includes, for example: negative binomial, Poisson-inverse gaussian and Poisson generalized hyperbolic secant distributions. Therefore, our long-term survival models can be viewed as heterogeneous promotion models. We present some statistical properties of our models and show that the promotion model is obtained as a limiting case. Some special models of the proposed class are discussed in details. We consider the expected number of competing causes depending on covariates, so allowing to a direct modeling of the cure rate through covariates. Estimation by maximum likelihood and inference for the parameters of models are discussed. In particular, we state sufficient conditions for the maximum likelihood estimators to be consistent and asymptotically normally distributed. A small simulation study is presented in order to check the finite-sample behavior of the maximum likelihood estimators and to illustrate the importance of our models when significant covariates are non-observed. We analyze a real data set from a melanoma clinical trial to illustrate the potential for practice of our proposed models.

1. Introduction

Survival models taking into account a cure fraction are a frequent theme considered in several theoretical and applied statistical articles. Such models are known as *cure rate models*. The interest in these models is increasing mainly due to the recent medical advances, where a considerable fraction of the patients has been cured of, for instance, some type of cancer. Boag (1949) and Berkson and Gage (1952) proposed a cure rate model with population survival function given by

$$S_{pop}(t) = p + (1-p)S(t), \quad t > 0,$$

where $p \in (0, 1)$ is the fraction of cured of the population and $S(t)$ is a proper survival function of the noncured group. In this model, the number of competing causes (which it will be referred as N from now on) is assumed to be a random variable following a Bernoulli distribution. This model is known as two-component mixture model with long-term survivals (or standard cure rate model). Asymptotic properties of the maximum likelihood estimators (MLEs) of the two-component mixture model have been established by Ghitany and Maller (1992) and Ghitany et al. (1994).

E-mail addresses: wagnerbs85@gmail.com, wagnerbs@est.ufmg.br.

http://dx.doi.org/10.1016/j.csda.2015.05.003
0167-9473/

Works dealing with the two-component mixture models have appeared in the literature. For instance, see Gray and Tsiatis (1989), Maller and Zhou (1996), Sposto et al. (1992), Kuk and Chen (1992), Lu and Ying (2004) and Yu et al. (2004), among others.

An alternative model to the two-component mixture model was proposed and studied by Yakovlev and Tsodikov (1996), Chen et al. (1999) and Tsodikov et al. (2003). This alternative model is defined under the first-activation scheme (minimum mechanism) and by assuming that the number of competing causes follows a Poisson distribution. This model is known in the last years as promotion model. Chen et al. (1999) introduced the promotion model through a biological context, which has motivated recently several works in this subject. A unified approach of the two-component mixture and promotion models was proposed by Yin and Ibrahim (2005). This unification is based on the Box–Cox transformation of the population survival function. A novel biological interpretation for a Box–Cox transformation cure rate model was given recently by Peng and Xu (2012). Under the assumption that N follows a Conway–Maxwell–Poisson (in short COM-Poisson) distribution, Rodrigues et al. (2009) introduced a model that also includes the two-component mixture and promotion models as particular cases.

More recently, Cooner et al. (2007) proposed a class of cure rate models based on a hierarchical-activation scheme, so obtaining the first and last activation schemes as particular cases. Further, they discuss the Poisson, Bernoulli, binomial and geometric specifications for the random number of competing causes.

For the readers interested in more details and issues about cure rate models we recommend the following two important books: Maller and Zhou (1996) and Ibrahim et al. (2001).

Zeng et al. (2006) and Tournoud and Ecochard (2008) introduced some alternatives to the promotion model. In particular, they proposed a cure rate model with number of causes following a negative binomial distribution. This model is known as negative binomial (NB) cure rate model. More specifically, the NB model arises by considering a multiplicative random effect (following a gamma distribution) in the mean of the number of causes in a promotion model.

Our aim in this present article is to introduce a general class of cure rate model by assuming that the number of competing causes N belongs to a class of mixed Poisson distributions, which are overdispersed. More specifically, we suppose that $N|Z$ follows a Poisson distribution with mean λZ, with $\lambda > 0$, and Z is a positive continuous random variable belonging to the exponential family. Therefore, our long-term survival models can be viewed as heterogeneous promotion models. The general mixed Poisson distributions we consider in this article includes the NB, Poisson-inverse gaussian and Poisson generalized hyperbolic secant distributions as particular cases and the Poisson distribution as a limiting case. Therefore, our general class of long-term survivals includes the promotion and NB cure rate modes as special cases. Further, our class introduces new models that were not yet considered in the literature, such as Poisson-inverse gaussian and Poisson-generalized hyperbolic secant cure rate models. These models also permit us to take into account the heterogeneity caused by the non-observation of significant covariates. This will be illustrated through a simulation study in Section 4.

The article is organized as follows. In Section 2, we introduce our class of long-term survival models with overdispersed number of competing causes and discuss some of its properties. Section 3 deals with the inferential aspects of the proposed class such as point estimation using maximum likelihood and its asymptotic properties. We state conditions that ensure the asymptotic normality of the maximum likelihood estimators. A small Monte Carlo simulation is presented in Section 4 in order to evaluate the finite-sample behavior of the maximum likelihood estimators and to illustrate the importance of heterogeneous promotion models when significant covariates are non-observed. An empirical illustration about the clinical trial E1673 is presented in Section 5. A brief discussion and future researches are addressed in Section 6. Proofs of the results stated along the article are given in the Appendix.

2. Model

In order to define our model let us to introduce some objects and notations of interest. First of all, we will define the mixed Poisson class of distributions which will be considered here. Let Z be a continuous positive random variable with distribution belonging to the exponential family with mean 1 and dispersion parameter $\phi > 0$. We denote this by $Z \sim \text{EF}(\phi)$. More specifically, we consider the density of Z assuming the form

$$f(z;\xi_0,\phi) = \exp\{\phi[z\xi_0 - b(\xi_0)] + c(z;\phi)\}, \quad z > 0, \tag{1}$$

where ξ_0 satisfies $E(Z) = b'(\xi_0) = 1$, with $b'(\cdot)$ denoting the first derivative of $b(\cdot)$, $c(z;\phi)$ is a function depending only on z and ϕ, and $b(\cdot)$ is a three-times continuously differentiable function. The variance of Z is $\text{var}(Z) = \phi^{-1}b''(\xi_0)$, where $b''(\cdot)$ is the second derivative of $b(\cdot)$. We define that a random variable N belongs to our mixed Poisson class of distributions if it satisfies the following stochastic representation: $N|Z \sim \text{Poisson}(\lambda Z)$ with $Z \sim \text{EF}(\phi)$. In this case, we denote $N \sim \text{MP}(\lambda,\phi)$. Note that we have taken $E(Z) = 1$ in order to avoid identifiability problems. The marginal probability generating function of N becomes

$$g_N(t) \equiv E(t^N) = \exp\big\{-\phi\big[b(\xi_0) - b\big(\xi_0 + \lambda\phi^{-1}(t-1)\big)\big]\big\}, \quad t \in \mathbb{R}. \tag{2}$$

It can be checked that $E(N) = \lambda$ and $\text{var}(N) = \lambda[1 + \lambda\phi^{-1}b''(\xi_0)]$. Therefore, N is overdispersed.

Let $\{W_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. positive random variable with survival function $S(t) = P(W_1 > t)$ for $t > 0$, $W_0 = \infty$ with probability 1 and $N \sim \text{MP}(\lambda,\phi)$. We are supposing all previous variables independent among them. With this, we define the random variable $Y = \min(W_0, W_1, \ldots, W_N)$. In several biological applications, the variable Y denotes the lifetime