Computational Statistics and Data Analysis 98 (2016) 31–45

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Destructive weighted Poisson cure rate models with bivariate random effects: Classical and Bayesian approaches

Diego I. Gallardo [a,*], Heleno Bolfarine [b], Antonio Carlos Pedroso-de-Lima [b]

[a] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile

[b] Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 29 May 2014
Received in revised form 5 December 2015
Accepted 6 December 2015
Available online 28 December 2015

Keywords:
REML
Dirichlet process
Competing risks

ABSTRACT

In this paper, random effects are included in the destructive weighted Poisson cure rate model. For parameter estimation we implemented a classical approach based on the restricted maximum likelihood (REML) methodology and a Bayesian approach based on Dirichlet process priors. A small scale simulation study is conducted to discuss parameter recovery and the performance of the proposed methodology is illustrated with a real data example.

1. Introduction

Cure rate models were developed to account for the presence of cured or immune individuals in a population where the main objective is the study of survival (in a broad sense) times. Berkson and Gage (1952) is frequently cited as the pioneering approach in this setting. Their proposed model, often called the *mixture model* (MM), assumes the existence of two types of subjects in the population: susceptible and cured individuals. Maller and Zhou (1996) presents a concise and complete study on the MM methodology from a classical point of view. Alternatively, Yakovlev and Tsodikov (1996) considered the so-called *promotion time cure rate model* (PTCRM), having in mind cancer patients. Specifically, they assume the existence of a latent quantity, M, representing the number of cells that may develop a cancerous tumour for a given individual. In spite of its medical genesis, this model is frequently considered in non-medical settings, where cells are replaced by latent causes of the event of interest. Susceptible and cured individuals are characterized by $M \geq 1$ and $M = 0$, respectively. When choosing between the two approaches, one may rely on the fact that the PTCRM allows inference for both, the probability of an specific individual being cured and the initial number of carcinogenic cells. In contrast, the MM allows inference only on the cure probability. In an attempt to generalize the PTCRM for a broader class of applications, Rodrigues et al. (2011) introduced the so-called *destructive weighted Poisson cure rate model* where it is assumed that each one of the initial causes have a probability p of generating the event of interest. Therefore, out of M, only $D \leq M$ causes would remain in effect. The interpretation of M and D depends on the particular situation: in cancer trials, M may represent the initial number of potentially cancerous cells, whereas D denotes the number of such cells that are kept active after a given treatment, possibly leading to a relapse. Obviously, when $p = 1$ the original PTCRM is obtained.

In this paper we deal with a slightly more general situation, where patients are naturally grouped into clusters, such as clinics of families. If those clinics are considered as a random sample from the population of clinics, the use of mixed models

* Corresponding author.
E-mail addresses: diego.gallardo@uantof.cl (D.I. Gallardo), hbolfar@ime.usp.br (H. Bolfarine), acarlos@ime.usp.br (A.C. Pedroso-de-Lima).

http://dx.doi.org/10.1016/j.csda.2015.12.006
0167-9473/

Table 1
Characteristics of some destructive weighted Poisson cure rate models discussed in Rodrigues et al. (2011).

	Destructive length biased Poisson model (DLBPM)	Destructive exponentially weighted Poisson model (DEWPM)	Destructive negative binomial model (DNBM)
$w(m;\phi)$	m	$e^{\phi m}$	$\Gamma(\phi^{-1}+m)$
Parameter space for ϕ	–	$\phi \in \mathbb{R}$	$\phi > 0$
Distribution of M	$\text{Po}(\theta)+1$	$\text{Po}(\theta e^{\phi})$	$\text{NB}\left(\phi, \frac{\phi\theta}{1+\phi\theta}\right)$
Distribution of D	$\text{Po}(\theta p)+\text{Bern}(p)$	$\text{Po}(\theta p e^{\phi})$	$\text{NB}\left(\phi, \frac{\phi\theta p}{1+\phi\theta p}\right)$
$S_{pop}(t;\theta,p,\phi)$	$(1-pF(t \mid \boldsymbol{\lambda}))e^{-\theta pF(t\mid\boldsymbol{\lambda})}$	$\exp\{-\theta p e^{\phi}F(t \mid \boldsymbol{\lambda})\}$	$\{1+\phi\theta pF(t \mid \boldsymbol{\lambda})\}^{-\phi^{-1}}$
$h_{pop}(t;\theta,p,\phi)$	$pf(t \mid \boldsymbol{\lambda})\left\{\theta+[1-pF(t \mid \boldsymbol{\lambda})]^{-1}\right\}$	$\theta p e^{\phi} f(t \mid \boldsymbol{\lambda})$	$\frac{\theta p f(t\mid\boldsymbol{\lambda})}{1+\phi\theta pF(t\mid\boldsymbol{\lambda})}$
Cure rate	$(1-p)e^{-\theta p}$	$\exp\{-\theta p e^{\phi}\}$	$\{1+\phi\theta p\}^{-\phi^{-1}}$

is a natural choice. We consider two random effects related to clinics: U related to relapse times of the disease caused by the non-destroyed cells and V associated with the clinic cure rate. Although several authors considered this approach in the context of non-destructive models (see, for example, Yau and Ng (2001), Lai and Yau (2008), Lopes and Bolfarine (2012) and Gallardo et al. (2013)) to the best of our knowledge there are no studies related to destructive models. In addition to the inclusion of random effects, we propose classical and Bayesian approaches in the estimation process. The random effects vector (U, V) is supposed to be bivariate normally distributed for the classical approach and a non-parametric framework based on Dirichlet processes priors is considered for the Bayesian approach.

The paper is organized as follows. Section 2 presents the ordinary destructive weighted Poisson model. In Section 3, we extend the model incorporating the bivariate random effects. In Section 4 we develop classical and Bayesian approaches for parameter estimation. Section 5 presents a simulation study to evaluate parameter recovery for the classical approach. Section 6 deals with an application of the proposed model and approaches to a real data set related to a study of the Oropharynx carcinoma. Section 7 presents a final discussion on the performance of the proposed methodologies.

2. Destructive weighted Poisson cure rate models

The model introduced in Rodrigues et al. (2011), considers M as a (unobservable) random variable denoting the initial number of carcinogenic cells of an individual, with probability mass function

$$P(M=m;\theta,\phi)=\frac{w(m;\phi)p^*(m;\theta)}{E_\theta[w(M;\phi)]},\quad m=0,1,2,\ldots, \tag{1}$$

where $w(\cdot;\phi)$ is a non-negative weight function with parameter ϕ, $p^*(\cdot;\theta)$ is the probability mass function (pmf) of the Poisson distribution with mean $\theta>0$. $E_\theta[\cdot]$ indicates that the expectation is taken with respect to the variable M following a Poisson distribution with mean θ. Given $M=m$, let ϱ_j, $j=1,2,\ldots,n$, be independent and identically distributed Bernoulli random variables. If the jth potentially cancerous cell is still alive after a procedure or treatment, $\varrho_j=1$; otherwise, $\varrho_j=0$. Therefore, for $P(\varrho_j=1)=p$, MM and PTCRM can be seen as particular cases of $p=1$ (which implies $M=D$).

The unobserved quantity

$$D=\begin{cases}\varrho_1+\cdots+\varrho_M, & \text{if } M>0,\\ 0, & \text{if } M=0,\end{cases}$$

$D\le M$, is the total number of carcinogenic cells not destroyed by the treatment. Clearly, $D \mid M=m \sim \text{Bin}(m,p)$ if $m>0$ and $P(D=0 \mid M=0)=1$. Also,

$$P(D=d;\theta,p,\phi)=\frac{e^{-\theta p}(\theta p)^d}{d!E_\theta[w(M;\phi)]}E_{\theta(1-p)}[w(M;\phi)].$$

Let W_a be a random variable expressing the time at which the ath non-destroyed cell produces a tumour (also known as the *promotion time*). For uncured patients, $D>0$ and we assume that W_a, $a=1,2,\ldots,D$, are conditionally independent given D, with common distribution function $F(t \mid \boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a set of unknown parameters. For cured patients, $D=0$ and we set $P(W_0=\infty)=1$. The distribution F is a proper distribution function.

The time until the occurrence of the event of interest can be represented by $T=\min\{W_a,\ 0\le a\le D\}$. The corresponding survival function, also called the *population survival function*, is given by

$$S_{pop}(t;\theta,p,\phi)=P(T>t)=\exp\{-\theta pF(t \mid \boldsymbol{\lambda})\}\frac{E_{\theta(1-pF(t\mid\boldsymbol{\lambda}))}[w(M;\phi)]}{E_\theta[w(M;\phi)]}. \tag{2}$$

Note that $\lim_{t\to+\infty}S_{pop}(t)=\exp\{-\theta p\}\{E_{\theta(1-p)}[w(M;\phi)]/E_\theta[w(M;\phi)]\}$; therefore (2) is an improper function and the limiting value corresponds to the *cure fraction* or the probability of cure. Table 1 summarizes some features corresponding to the three models considered by Rodrigues et al. (2011).