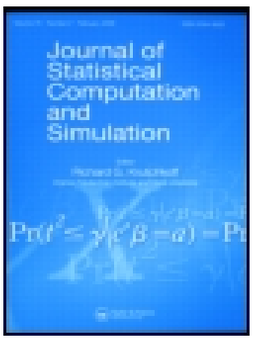

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

A note on the EM algorithm for estimation in the destructive negative binomial cure rate model

Diego I. Gallardo, Yolanda M. Gómez & Mário de Castro

To cite this article: Diego I. Gallardo, Yolanda M. Gómez & Mário de Castro (2017): A note on the EM algorithm for estimation in the destructive negative binomial cure rate model, Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2017.1327589

To link to this article: http://dx.doi.org/10.1080/00949655.2017.1327589

Published online: 17 May 2017.

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2017
https://doi.org/10.1080/00949655.2017.1327589

A note on the EM algorithm for estimation in the destructive negative binomial cure rate model

Diego I. Gallardo [a*], Yolanda M. Gómez [b*] and Mário de Castro [c]

[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; [c]Instituto de Ciẽncias Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brazil

ABSTRACT

In this note we present a modification in the EM algorithm for the destructive negative binomial cure rate model. This alteration enables us to obtain the estimates of the whole parameter vector from the complete log-likelihood function, avoiding the corresponding observed log-likelihood function, which is more involved. To achieve this goal, we resort to the mixture representation of the negative binomial distribution in terms of the Poisson and gamma distributions.

ARTICLE HISTORY
Received 6 April 2016
Accepted 3 May 2017

KEYWORDS
Cure rate model; negative binomial distribution; survival analysis

1. Introduction

The destructive negative binomial model (DNB) proposed in [1] can be described in the following way. Let M be a random variable denoting the initial number of carcinogenic cells of an individual. In their work, the authors used the weighted Poisson distribution for M, but for our purposes we only present the negative binomial model (a distribution in that class) with probability mass function given by

$$P(M=m;\theta,\phi)=\frac{\Gamma(\phi^{-1}+m)}{m!\Gamma(\phi^{-1})}\left(\frac{\phi\theta}{1+\phi\theta}\right)^{m}(1+\phi\theta)^{-\phi^{-1}},\quad m=0,1,2,\ldots. \tag{1}$$

We denote the distribution in Equation (1) as $NB(\phi,\phi\theta/[1+\phi\theta])$. Suppose that only D ($\leq M$) cells remain active to result in a new tumour (also called damaged cells) with probability p. We may associate to the jth cell a Bernoulli random variable ζ_j, independent of M, so that $P(\zeta_j=1)=p$ and

$$D=\begin{cases}\zeta_1+\cdots+\zeta_M & \text{if } M>0,\\ 0 & \text{if } M=0.\end{cases}$$

Notice that if $D=0$ or $D>0$, the individual is termed cured or non-cured, respectively. For a cured individual, the event of interest (e.g. death due to cancer) will never occur. Assuming independence among the random variables ζ_j, $j=1,2,\ldots$, it is clear that the distribution of D conditioning on $M=m$ is Bin(m,p) if $m>0$ and $P(D=0\mid M=0)=1$. For non-cured individuals, define W_a as the random variable denoting the time taking for the ath cell to produce a detectable cancer. Assume W_a, $a=1,2,\ldots,D$, are conditionally independent given $D=d$ and identically distributed with common cumulative distribution function (cdf)$F(t;\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a set of unknown parameters. The cdf F is

CONTACT Diego I. Gallardo diego.gallardo@uda.cl
*Present address: Departamento de matemáticas, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile.

a proper function in the sense that $\lim_{t\to\infty} F(t;\boldsymbol{\lambda}) = 1$. The time until the event of interest can be represented by $T = \min\{W_a,\ 0 \le a \le D\}$. Under such conditions, the (improper) survival function for T is given by

$$S_{\text{pop}}(t;\theta,p,\phi) = [1+\phi\theta pF(t;\boldsymbol{\lambda})]^{-\phi^{-1}}.$$

For cured individuals, we define $P(W_0 = \infty) = 1$. The cure fraction of the model is given by $S_{\text{pop}}(\infty;\theta,p,\phi) = (1+\phi\theta)^{-\phi^{-1}}$.

Assume that data where obtained subject to right censoring. Specifically, assume that the observable data for the ith individual can be represented by random variables $T_i = \min(T_i^*, C_i)$ and $\delta_i = I(T_i^* \le C_i)$, $i = 1,\dots,n$, with T_i^* and C_i denoting failure and censoring times, respectively. Moreover, assume that for each individual we associate a set of covariates $\boldsymbol{z}_{1i}$ (a $r_1 \times 1$ vector) related to the initial number of cells and $\boldsymbol{z}_{2i}$ (a $r_2 \times 1$ vector) associated with the parameters in Equation (1), in such way that

$$\log(\theta_i) = \boldsymbol{z}_{1i}^\top\boldsymbol{\beta}_1 \quad \text{and} \quad \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{z}_{2i}^\top\boldsymbol{\beta}_2, \quad i = 1,\dots,n. \tag{2}$$

Note that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ do not share common elements and $\boldsymbol{\beta}_1$ does not include an intercept to circumvent identifiability problems in the sense of Li et al. [2].

This work is organized as follows. In Section 2, we present a new way to perform the parameter estimation for the DNB model. In Section 3, we present in detail the modified EM algorithm for this model. Section 4 presents a comparison with other proposals to obtain estimates in a real data set. Finally, in Section 5 we present some remarks. For the sake of space, we concentrate on the EM algorithm and omit details such as the likelihood function for the model. Interested readers can refer to Cancho et al. [3] and Gallardo et al. [4].

2. The proposal

In the work by Gallardo et al. [4] dealing with the destructive negative binomial cure rate model, an EM algorithm is proposed to estimate $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\lambda}$ separately instead of performing the maximization with respect to all components, say $\boldsymbol{\psi} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \phi, \boldsymbol{\lambda})$. However, the solution in that proposal to estimate ϕ in the destructive negative binomial cure rate model (DNB) is to construct a set of values for ϕ and choose the value that maximizes the profile log-likelihood function (see Section 4). A simpler solution is to assume that ϕ is known (see, for instance, Cancho et al. [3]). The negative binomial distribution of M_i in Equation (1) is equivalent to consider

$$M_i \mid Y_i = y_i \sim \text{Poisson}(y_i) \quad \text{and} \quad Y_i \sim \text{Gamma}(\phi^{-1}, \phi\theta_i), \quad i = 1,\dots,n,$$

where $Y \sim \text{Gamma}(a,b)$ denotes the gamma distribution with density function $f(y;a,b) = b^a y^{a-1}\,\mathrm{e}^{-by}/\,\Gamma(a)$, for $y > 0$. The observed data is denoted by $\boldsymbol{D}_{\text{obs}} = (\boldsymbol{t}, \boldsymbol{\delta})$. On the other hand, now the complete data are $\boldsymbol{D}_{\text{comp}} = (\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{D}, \boldsymbol{M}, \boldsymbol{Y})$, where $\boldsymbol{M} = (M_1,\dots,M_n)$, $\boldsymbol{D} = (D_1,\dots,D_n)$ and $\boldsymbol{Y} = (Y_1,\dots,Y_n)$ denote the latent variables. The density function for the complete data is given by

$$\begin{aligned} f(t_i,\delta_i,d_i,m_i,y_i;\theta_i,p_i,\phi,\boldsymbol{\lambda}) &= f(t_i,\delta_i \mid D_i = d_i)P(D_i = d_i \mid M_i = m_i;p_i) \\ &\quad \times P(M_i = m_i \mid Y_i = y_i)f_{Y_i}(y_i;\theta_i,\phi) \\ &= S(t_i;\boldsymbol{\lambda})^{d_i-\delta_i}[d_i f(t_i;\boldsymbol{\lambda})]^{\delta_i}\binom{m_i}{d_i}p_i^{d_i}(1-p_i)^{m_i-d_i}\frac{y_i^{m_i}}{m_i!}\mathrm{e}^{-y_i} \\ &\quad \times \frac{(\phi\theta_i)^{\phi^{-1}}}{\Gamma(\phi^{-1})}y_i^{\phi^{-1}-1}\mathrm{e}^{-\phi\theta_i y_i}, \quad \delta_i \le d_i \le m_i,\ \ y_i > 0, \end{aligned} \tag{3}$$

for $i = 1,\dots,n$. Based on (3), we establish the following proposition.

Proposition 2.1. *For the DNB model, the conditional distributions of* (i) $M_i - \delta_i$ *given* $Y_i = y_i$ *and* $\boldsymbol{D}_{\text{obs}}$ *and (ii)* Y_i *given* $\boldsymbol{D}_{\text{obs}}$ *are*

$$M_i - \delta_i \mid Y_i = y_i, \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{Poisson}(y_i[1 - p_i F(t_i; \boldsymbol{\lambda})])$$

and

$$Y_i \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi} \sim \text{Gamma}\left(\phi^{-1} + \delta_i, \frac{1 + \phi\theta_i p_i F(t_i; \boldsymbol{\lambda})}{\phi\theta_i}\right),$$

for $i = 1, \ldots, n$.

The proof of Proposition 2.1 is given in the appendix. The following corollary is immediate from Proposition 2.1.

Corollary 2.1. *The expected values of* (i) Y_i *given* $\boldsymbol{D}_{\text{obs}}$ *and* (ii) $\log(Y_i)$ *given* $\boldsymbol{D}_{\text{obs}}$ *are*

$$(i) \quad \mathbb{E}(Y_i \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) = \frac{(1 + \phi\delta_i)\theta_i}{1 + \phi\theta_i p_i F(t_i; \boldsymbol{\lambda})}$$

and

$$(ii) \quad \mathbb{E}(\log(Y_i) \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) = \eta(\phi^{-1} + \delta_i) + \log(\phi) + \log(\theta_i) - \log(1 + \phi\theta_i p_i F(t_i; \boldsymbol{\lambda})),$$

for $i = 1, \ldots, n$, *where* $\eta(\cdot)$ *denotes the digamma function.*

Moreover, note that

$$\begin{aligned}
\mathbb{E}(M_i - \delta_i \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) &= \mathbb{E}(\mathbb{E}(M_i - \delta_i \mid Y_i, \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) \\
&= [1 - p_i F(t_i; \boldsymbol{\lambda})]\mathbb{E}(Y_i \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}) \\
&= \frac{(1 + \phi\delta_i)\theta_i[1 - p_i F(t_i; \boldsymbol{\lambda})]}{1 + \phi\theta_i p_i F(t_i; \boldsymbol{\lambda})},
\end{aligned}$$

for $i = 1, \ldots, n$, agreeing with Corollary 1 in [4]. In the sequel we present a modification in the EM algorithm for the DNB model.

3. A modification in the EM algorithm

From (3), up to a constant that does not depend on $\boldsymbol{\psi}$, the complete log-likelihood for $\boldsymbol{\psi}$ can be written as

$$\begin{aligned}
\ell_c(\boldsymbol{\psi}; \boldsymbol{D}_{\text{comp}}) = \sum_{i=1}^{n} \Bigg\{ & D_i \log(S(t_i; \boldsymbol{\lambda})) + \delta_i \log(h(t_i; \boldsymbol{\lambda})) + D_i \log(p_i) + (M_i - D_i) \log(1 - p_i) \\
& -\phi^{-1}[\log(\theta_i) + \log(\phi) - \log(Y_i)] - \frac{Y_i}{\phi\theta_i} \Bigg\} - n \log(\Gamma(\phi^{-1})),
\end{aligned}$$

where $h(t; \boldsymbol{\lambda}) = f(t; \boldsymbol{\lambda})/S(t; \boldsymbol{\lambda})$ is the hazard function. Let $\boldsymbol{\psi}^{(k)} = (\boldsymbol{\beta}_1^{(k)}, \boldsymbol{\beta}_2^{(k)}, \phi^{(k)}, \boldsymbol{\lambda}^{(k)})$ be the estimate of $\boldsymbol{\psi}$ at the kth iteration of the EM algorithm and denote the conditional expectation of $\ell_c(\boldsymbol{\psi} \mid \boldsymbol{D}_{\text{comp}})$

given the observed data by $Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)})$. With these notations, we have that

$$Q(\boldsymbol{\psi} \mid \boldsymbol{\psi}^{(k)}) = Q_1(\boldsymbol{\xi} \mid \boldsymbol{\psi}^{(k)}) + Q_2(\boldsymbol{\beta}_2 \mid \boldsymbol{\psi}^{(k)}) + Q_3(\boldsymbol{\lambda} \mid \boldsymbol{\psi}^{(k)}),$$

with $\boldsymbol{\xi} = (\boldsymbol{\beta}_1, \phi)$,

$$Q_1(\boldsymbol{\xi} \mid \boldsymbol{\psi}^{(k)}) = \sum_{i=1}^{n} \left\{ \phi^{-1}(\widetilde{\log}(Y_i^{(k)}) - \log(\theta_i) - \log(\phi)) - \frac{\tilde{Y}_i^{(k)}}{\phi\theta_i} \right\} - n\log(\Gamma(\phi^{-1})) + C_1, \quad (4)$$

$$Q_2(\boldsymbol{\beta}_2 \mid \boldsymbol{\psi}^{(k)}) = \sum_{i=1}^{n} \{\tilde{D}_i^{(k)} \log(p_i) + (\tilde{M}_i^{(k)} - \tilde{D}_i^{(k)}) \log(1 - p_i)\} + C_2 \quad \text{and} \quad (5)$$

$$Q_3(\boldsymbol{\lambda} \mid \boldsymbol{\psi}^{(k)}) = \sum_{i=1}^{n} \{\tilde{D}_i^{(k)} \log(S(t_i; \boldsymbol{\lambda})) + \delta_i \log(h(t_i; \boldsymbol{\lambda}))\} + C_3, \quad (6)$$

where $\tilde{D}_i^{(k)} = \mathbb{E}(D_i \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)})$, $\tilde{M}_i^{(k)} = \mathbb{E}(M_i \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)})$, $\tilde{Y}_i^{(k)} = \mathbb{E}(Y_i \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)})$, $\widetilde{\log}(Y_i^{(k)}) = \mathbb{E}(\log(Y_i) \mid \boldsymbol{D}_{\text{obs}}; \boldsymbol{\psi}^{(k)})$ and C_1, C_2 and C_3 do not depend on $\boldsymbol{\xi}$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\lambda}$, respectively. The expressions of $\tilde{D}_i^{(k)}$ and $\tilde{M}_i^{(k)}$ are provided in [4], whereas $\tilde{Y}_i^{(k)}$ and $\widetilde{\log}(Y_i^{(k)})$ can be computed using Corollary 2.1. The advantage of this approach is that we do not need to fix ϕ and build the profile log-likelihood for it. Instead, ϕ can be introduced in the estimation procedure and all parameters are estimated from the complete log-likelihood function, so that the more involved observed log-likelihood function, as in [4], is avoided.

In short, the kth iteration of the algorithm comprises the following steps:

- *E-step*: For $i = 1, \ldots, n$, compute

$$\tilde{D}_i^{(k)} = \delta_i + \frac{(1 + \phi^{(k-1)}\delta_i)\theta_i^{(k-1)} p_i^{(k-1)} S(t_i; \boldsymbol{\lambda}^{(k-1)})}{1 + \phi^{(k-1)}\theta_i^{(k-1)} p_i^{(k-1)} F(t_i; \boldsymbol{\lambda}^{(k-1)})},$$

$$\tilde{M}_i^{(k)} = \delta_i + \frac{(1 + \phi^{(k-1)}\delta_i)\theta_i^{(k-1)} [1 - p_i^{(k-1)} F(t_i; \boldsymbol{\lambda}^{(k-1)})]}{1 + \phi^{(k-1)}\theta_i^{(k-1)} p_i^{(k-1)} F(t_i; \boldsymbol{\lambda}^{(k-1)})},$$

$$\tilde{Y}_i^{(k)} = \frac{(1 + \phi^{(k-1)}\delta_i)\theta_i^{(k-1)}}{1 + \phi^{(k-1)}\theta_i^{(k-1)} p_i^{(k-1)} F(t_i; \boldsymbol{\lambda}^{(k-1)})} \quad \text{and}$$

$$\widetilde{\log}(Y_i^{(k)}) = \eta[\phi^{-1(k-1)} + \delta_i] + \log(\phi^{(k-1)}) + \log(\theta_i^{(k-1)}) - \log(1 + \phi^{(k-1)}\theta_i^{(k-1)} p_i^{(k-1)} F(t_i; \boldsymbol{\lambda}^{(k-1)})).$$

- *M-step*: Given $\tilde{\boldsymbol{D}}^{(k)} = (\tilde{D}_1^{(k)}, \ldots, \tilde{D}_n^{(k)})$, $\tilde{\boldsymbol{M}}^{(k)} = (\tilde{M}_1^{(k)}, \ldots, \tilde{M}_n^{(k)})$, $\tilde{\boldsymbol{Y}}^{(k)} = (\tilde{Y}_1^{(k)}, \ldots, \tilde{Y}_n^{(k)})$ and $\widetilde{\log}(\boldsymbol{Y}^{(k)}) = (\widetilde{\log}(Y_1^{(k)}), \ldots, \widetilde{\log}(Y_n^{(k)}))$, find $\boldsymbol{\xi}^{(k)}$, $\boldsymbol{\beta}_2^{(k)}$ and $\boldsymbol{\lambda}^{(k)}$ that maximize (4), (5) and (6) with respect to $\boldsymbol{\xi}$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\lambda}$, respectively.

Maximization of Equations (4)–(6) can be performed using extant software; e.g. with the optim function in R [5]. Finally, the covariance matrix of the maximum likelihood estimator $\hat{\boldsymbol{\psi}}$ can be estimated based on the Hessian matrix of the log-likelihood function. The package *numDeriv* [6] in R provides a good numerical approximation to this matrix.

4. Application

In order to illustrate the efficiency of our proposal, we present an application to a well-known data set on a phase III cutaneous melanoma clinical trial available at http://merlot.stat.uconn.edu/ ∼ mhchen/survbook/, labelled as E1690 data. The clinical trial was conducted by the Eastern Cooperative Oncology Group (see [7], for details). The main goal of the study was to assess a postoperative treatment performance with a high dose of the drug Interferon alpha-2b, in order to prevent recurrence. The study included patients between 1991 and 1995 and follow-up was conducted until 1998.

A characteristic of this disease is the presence of a proportion of patients that can lead a normal life, comparable (in longevity) to patients without the disease. In other words, a proportion commonly known as 'cured'. After deleting patients with incomplete data and missing observation times, the data set is composed by $n = 408$ individuals. The collected variables were (SD stands for standard deviation): observed time (in years, average $= 2.31$ and SD $= 1.93$), treatment (0: control and 1: interferon alfa-2b with 198 and 210 patients, respectively), age (in years, average $= 48.1$ and SD $= 13.1$), nodal category (categorical variable with levels 1–4 with 110, 131, 86 and 81 patients in each group, respectively, where 1 indicates the lower risk patients and 4 the higher risk patients) and tumour thickness (in mm, average $= 3.98$ and SD $= 3.22$). Figure 1 displays the Kaplan–Meier curves by nodule category, confirming a well pronounced plateau in all nodule categories.

For this data set we fit the DNB applying the EM algorithm in the way presented in [4] (i.e. building a profile log-likelihood approach for ϕ) and our proposal, where ϕ is included in the estimation procedure. For the time-to-event of the cells, we consider the Weibull distribution that is a suitable model in this biological context. Specifically, we consider the parametrization such as $S(t;\boldsymbol{\lambda}) = \exp\{-e^{\alpha}t^{\nu}\}$. Following Cancho et al. [3]in Equation (1), we link the covariates nodule and thickness to θ and treatment to p. A routine was developed in R language and is available from the authors upon request. We ran the program in a computer equipped with an AMD A6-6310 APU processor with 1.80 GHz and a RAM memory of 8 GB. Table 1 shows the estimates using both methods. Note that all estimates from the two models are close. However, the runtime using our proposal is much lower. The reduction in the runtime is explained by the fact that the EM algorithm based on the profile log-likelihood function should be applied to a set of values for ϕ. The results in Table 1 correspond to a search beginning with the set $\{1, 2, \ldots, 10\}$. Let ϕ^* be the value of ϕ in this set that maximizes the observed log-likelihood function. With this value, a new set $\{\phi^* - 9/10, \phi^* - 8/10, \ldots, \phi^* + 9/10\}$ is formed. From this set we update ϕ^*. The process is iterated taking sets with increments divided by 10 (1, 1/10, 1/100 and so

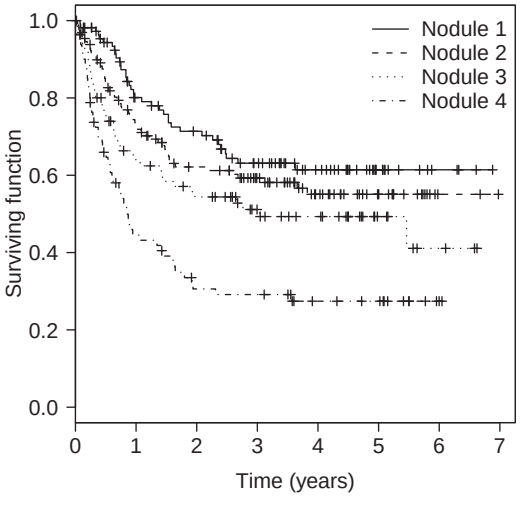

Figure 1. Kaplan–Meier curves stratified by nodule category.

Table 1. Parameter estimates (Est.) and standard errors (SE) for the DNB model.

	EM algorithm			
	Profiling ϕ		Our proposal	
Parameter	Est.	SE	Est.	SE
$\beta_{1,\text{nodule1}}$	0.468	0.619	0.459	0.618
$\beta_{1,\text{nodule2}}$	1.515	0.771	1.514	0.769
$\beta_{1,\text{nodule3}}$	2.154	0.895	2.153	0.893
$\beta_{1,\text{nodule4}}$	3.071	0.943	3.070	0.941
$\beta_{1,\text{thickness}}$	0.086	0.049	0.086	0.049
$\beta_{2,\text{treatment}}$	−0.796	0.455	−0.796	0.454
α	−1.315	0.441	−1.314	0.440
ν	1.537	0.196	1.537	0.196
ϕ	3.181	1.320	3.177	1.317
Runtime (in minutes)	63.86		0.89	

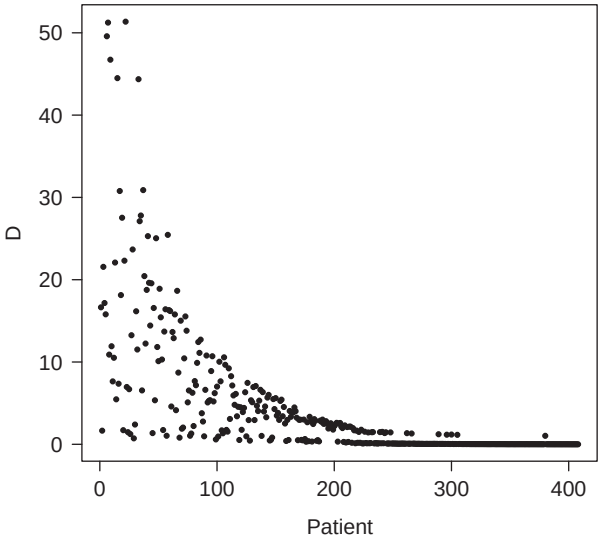

Figure 2. Index plot of the predicted number of activated carcinogenic cells (D).

on). The refinements stop when the difference between two successive estimates of ϕ attains a fixed tolerance. As an alternative, a fixed set can be chosen. However, often this choice is not obvious and can involve trial and error steps, as well as refinements of a coarser set. Therefore, both approaches are potentially time-consuming. On the other hand, under our proposal ϕ is included in the estimation process through the maximization of the function in Equation (4). In Table 1 the two versions of the EM algorithm were run with tolerance for ϕ equal to 0.0001.

Finally, Figure 2 presents the predictions of the number of activated carcinogenic cells (D) obtained from the EM algorithm. As the patients are sorted according to the observed times (in increasing order), it is expected that the first patients presented more activated carcinogenic cells than the last patients, as shown in the figure. The predicted values of D are a by-product of the EM algorithm and can be useful for understanding the carcinogenesis process.

5. Conclusion

In this work, we developed a modification in the EM algorithm described by Gallardo et al. [4] leading us to a more efficient estimation procedure. Differently from that work, the precision parameter ϕ is not estimated by profiling the observed log-likelihood function, but instead it is directly included in

the estimation procedure via the maximization of the complete log-likelihood function. The efficiency of the method is illustrated with a real data set. In our example, adopting the EM algorithm in this work, the runtime was reduced by a factor of about 72 relative to the proposal in [4]. We envision that a similar strategy could be applied to other cure models built on the negative binomial distribution.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The first author acknowledges partial support from Programa de iniciación en investigación para investigadores jenes de la Universidad de Antofagasta, INI 16-17-01. The work of the first author is partially supported by grant Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) 11160670, Chile. The work of the third author is partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

ORCID

D.I. Gallardo http://orcid.org/0000-0001-8184-7403

References

[1] Rodrigues J, de Castro M, Balakrishnan N, et al. Destructive weighted Poisson cure rate models. Lifetime Data Anal. 2011;17:333–346.
[2] Li C-S, Taylor JMG, Sy JP. Identifiability of cure models. Stat Probab Lett. 2001;54:389–395.
[3] Cancho V, Bandyopadhyay D, Louzada F, et al. The destructive negative binomial cure rate model with a latent activation scheme. Stat Methodol. 2013;13:48–68.
[4] Gallardo DI, Bolfarine H, Pedroso-de-Lima AC. An EM algorithm for estimating the destructive weighted Poisson cure rate model. J Stat Comput Simul. 2016;86:1497–1515.
[5] R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. 2015. ISBN: 3-900051-07-0, version 3.0.3.
[6] Gilbert P, Varadhan R. numDeriv: accurate numerical derivatives. R package version 2014.2-1. 2015. Available from: https://CRAN.R-project.org/package = numDeriv.
[7] Ibrahim JG, Chen MH, Sinha D. Bayesian survival analysis. New York: Springer; 2001.

Appendix. Proof of Proposition 2.1

From Equation (3), we obtain

$$\begin{aligned} f(t_i,\delta_i,m_i,y_i;\boldsymbol{\psi}) &= \sum_{d_i=0}^{m_i} f(t_i,\delta_i,d_i,m_i,y_i) \\ &= \left[\frac{f(t_i;\boldsymbol{\lambda})}{S(t_i;\boldsymbol{\lambda})}\right]^{\delta_i}(1-p_i)^{m_i} f_{Y_i}(y_i;\theta_i,\phi)\sum_{d_i=\delta_i}^{m_i} d_i^{\delta_i}\binom{m_i}{d_i}\left[\frac{p_i S(t_i;\boldsymbol{\lambda})}{1-p_i}\right]^{d_i} \\ &= [p_i f(t_i;\boldsymbol{\lambda})]^{\delta_i}[1-p_i F(t_i;\boldsymbol{\lambda})]^{m_i-\delta_i} f_{Y_i}(y_i;\theta_i,\phi)\,\mathrm{e}^{-y_i}\frac{y_i^{m_i}}{(m_i-\delta_i)!}, \end{aligned}$$

and $f(t_i,\delta_i;\boldsymbol{\psi})$ is provided in the appendix in [4]. Therefore,

$$\begin{aligned} f(m_i,y_i \mid \boldsymbol{D}_{\text{obs}};\boldsymbol{\psi}) &= \frac{f(t_i,\delta_i,m_i,y_i;\boldsymbol{\psi})}{f(t_i,\delta_i;\boldsymbol{\psi})} \\ &= \frac{[p_i f(t_i;\boldsymbol{\lambda})]^{\delta_i}[1-p_i F(t_i;\boldsymbol{\lambda})]^{m_i-\delta_i}\,\mathrm{e}^{-y_i}\frac{y_i^{m_i}}{(m_i-\delta_i)!}\frac{(\phi\theta_i)^{\phi^{-1}}}{\Gamma(\phi^{-1})}y_i^{\phi^{-1}-1}\,\mathrm{e}^{-\phi\theta_i y_i}}{[\theta_i p_i f(t_i;\boldsymbol{\lambda})]^{\delta_i}[1+\phi\theta_i p_i F(t_i;\boldsymbol{\lambda})]^{-(\phi^{-1}+\delta_i)}} \\ &= \underbrace{\frac{(\alpha_i^*)^{m_i-\delta_i}}{(m_i-\delta_i)!}\mathrm{e}^{-\alpha_i^*}}_{M_i-\delta_i\mid Y_i=y_i,D_{\text{obs}};\psi\sim\text{Poisson}(\alpha_i^*)} \times \underbrace{\frac{(\theta_i^*)^{\phi_i^*}}{\Gamma(\phi_i^*)}y_i^{\phi_i^*-1}\mathrm{e}^{-\theta_i^* y_i}}_{Y_i\mid D_{\text{obs}};\psi\sim\text{Gamma}(\phi_i^*,\theta_i^*)}, \end{aligned}$$

where $\alpha_i^* = y_i[1-p_iF(t_i;\boldsymbol{\lambda})]$, $\theta_i^* = [1+\phi\theta_i p_i F(t_i;\boldsymbol{\lambda})]/\phi\theta_i$ and $\phi_i^* = \phi^{-1}+\delta_i$, for $i=1,\ldots,n$.