Brazilian Journal of Probability and Statistics
2010, Vol. 24, No. 1, 100–103
DOI: 10.1214/08-BJPS015

A note on a unified approach for cure rate models

Mário de Castro[a], Vicente G. Cancho[a] and Josemar Rodrigues[b]

[a]*Universidade de São Paulo*
[b]*Universidade Federal de São Carlos*

Abstract. Yin and Ibrahim [*Canad. J. Statist.* **33** (2005) 559–570] presented a unified class of cure rate models based on a Box–Cox type transformation of the population survival function. Our work provides a probabilistic justification to this transformation by means of the negative binomial distribution.

1 Introduction

Models for survival data with a surviving fraction (also known as cure rate models or long-term survival models) play an important role in reliability and survival analysis. Cure rate models cover the situations where there are sampling units insusceptible to the occurrence of the event of interest. The literature on the subject is by now rich and growing rapidly. The books by Maller and Zhou (1996) and Ibrahim, Chen and Sinha (2001), as well as the review article by Tsodikov, Ibrahim and Yakovlev (2003) and the recent article by Cooner et al. (2007), could be mentioned as key references. Our chief contribution is a probabilistic justification through the negative binomial distribution for the transformation introduced by Yin and Ibrahim (2005).

2 Model and main results

As in Yakovlev and Tsodikov (1996) and Yin and Ibrahim (2005), we formulate the model within a biological context. The promotion time (time to event) for the kth tumor cell is denoted by t_k, $k = 1, \ldots, N$, where N denotes the unobservable number of competing causes that can produce a detectable cancer. We assume that the variables t_k are i.i.d. with cumulative distribution function $F(t)$ and $S(t) = 1 - F(t)$. The observable time to relapse of cancer is defined as $T = \min\{t_0, t_1, \ldots, t_N\}$, where $\mathrm{P}(t_0 = \infty) = 1$. Further, we assume that N is independent of $t_1, t_2, \ldots$. Exponential, piecewise exponential, and Weibull distributions, for instance, can be used to represent $t_1, t_2, \ldots$. Under this setup, the survival function for the population is given by

$$S_{\mathrm{pop}}(t) = \mathrm{P}(N = 0) + \mathrm{P}(t_1 > t, \ldots, t_N > t | N \geq 1)\mathrm{P}(N \geq 1).$$

Key words and phrases. Survival analysis, cure rate models, long-term survival models, negative binomial distribution.

Received June 2008; accepted September 2008.

Tsodikov, Ibrahim and Yakovlev (2003), among others, proved that $S_{\mathrm{pop}}(t) = g(S(t))$, where $g(\cdot)$ is the probability generating function of the number of competing causes (N). The choice of a particular distribution for N entails some consequences. In this note the number of competing causes is modeled by the negative binomial distribution, leading to a formulation that encompasses some specific models found in the literature.

We suppose that the number of competing causes follows a negative binomial distribution with parameters α and θ (Piegorsch (1990); Saha and Paul (2005)), with probability function

$$p_m = \mathrm{P}(N = m) = \frac{\Gamma(\alpha^{-1} + m)}{\Gamma(\alpha^{-1})m!}\left(\frac{\alpha\theta}{1+\alpha\theta}\right)^m (1+\alpha\theta)^{-1/\alpha}, \qquad (2.1)$$

$m = 0, 1, 2, \ldots,$ for $\theta > 0$ and $\alpha > -1/\theta$, so that

$$\mathrm{E}(N) = \theta \quad \text{and} \quad \mathrm{var}(N) = \theta(1+\alpha\theta). \qquad (2.2)$$

The probability generating function is given by

$$g(s) = \sum_{m=0}^{\infty} p_m s^m = \{1 + \alpha\theta(1-s)\}^{-1/\alpha}, \qquad 0 \le s \le 1. \qquad (2.3)$$

Taking into account (2.3) we arrive at the improper survival function

$$S_{\mathrm{pop}}(t) = g(S(t)) = \{1 + \alpha\theta F(t)\}^{-1/\alpha}. \qquad (2.4)$$

The surviving fraction, defined as $p_0 = S_{\mathrm{pop}}(\infty) = \lim_{t\to\infty} S_{\mathrm{pop}}(t)$, comes out to be $p_0 = (1+\alpha\theta)^{-1/\alpha}$. Figure 1 displays the improper survival function corresponding to different values for α. In these plots the promotion time is exponentially distributed with scale parameter equal to 1 and $\theta = 0.8$.

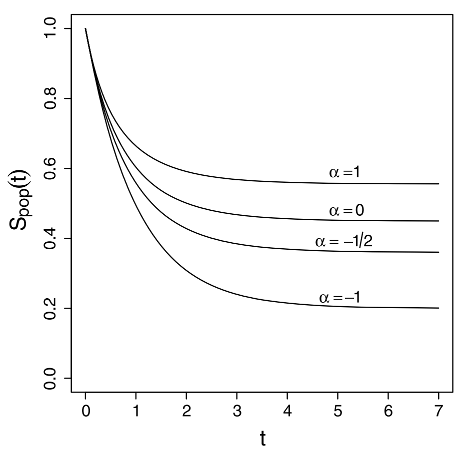

Figure 1 *Improper survival functions with* $F(t) = 1 - e^{-t}$ *and* $\theta = 0.8$.

Yin and Ibrahim (2005) proposed a class of cure rate models motivated by a transformation of the unknown population survival function. Next we prove that their proposal can be interpreted by means of the negative binomial distribution. The argument runs as follows. From (2.4) we have that

$$\frac{S_{\mathrm{pop}}(t)^{-\alpha} - 1}{\alpha} = \theta F(t), \tag{2.5}$$

an expression analogous to the Box–Cox transformation and that reduces to expression (4) in Yin and Ibrahim (2005) by taking $a = -\alpha$ in their paper. As $\alpha \to 0$, we obtain the Poisson distribution in (2.1) and $S_{\mathrm{pop}}(t)$ in (2.5) becomes $S_{\mathrm{pop}}(t) = \exp\{-\theta F(t)\}$, giving rise to the promotion time cure model (Yakovlev and Tsodikov (1996)).

Piegorsch (1990) pointed out that when $\alpha = -1/\kappa$, for κ a positive integer such that $\kappa > \theta$, the negative binomial distribution with parameters θ and $-1/\kappa$ gives the same probabilities as a binomial distribution with parameters κ and θ/κ. So, taking $\kappa = 1$ (Bernoulli distribution) we get $\alpha = -1$, $S_{\mathrm{pop}}(t)$ in (2.5) becomes $S_{\mathrm{pop}}(t) = 1 - \theta F(t)$, corresponding to the mixture cure model (Berkson and Gage (1952)). Therefore, α can be called a dispersion parameter (Saha and Paul (2005)). From (2.2) it follows that the variance of the number of competing causes under the negative binomial model is flexible. If $-1/\theta < \alpha < 0$, there is under-dispersion from the Poisson model. We illustrated this point with the mixture cure model. On the other side, if $\alpha > 0$ the counts are over-dispersed.

3 Conclusion

In a few words, under the negative binomial distribution for the number of competing causes, we present a probabilistic formulation of the cure rate model related to the transformation proposed by Yin and Ibrahim (2005). In our interpretation there is a strong connection between the parameter of the transformation (α) and the dispersion in the counts of competing causes.

Adopting the negative binomial distribution for the number of competing causes, we envision more general cure rate models with the dispersion parameter $\alpha \geq -1$ in (2.4). For instance, $\alpha = 1$ conduces to the geometric distribution. Promising work is in progress addressing issues such as model identifiability (Li, Taylor and Sy (2001)) and the existence of proper posterior distributions in a Bayesian context (Chen, Ibrahim and Sinha (1999), Cooner et al. (2007)).

Acknowledgment

The authors would like to thank an anonymous referee for the comments.

References

Berkson, J. and Gage, R. P. (1952). Survival cure for cancer patients following treatment. *Journal of the American Statistical Association* **47** 501–515.

Chen, M.-H., Ibrahim, J. G. and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association* **94** 909–919. MR1723307

Cooner, F., Banerjee, S., Carlin, B. P. and Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association* **102** 560–572. MR2370853

Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York. MR1876598

Li, C. S., Taylor, J. M. G. and Sy, J. P. (2001). Identifiability of cure models. *Statistics and Probability Letters* **54** 389–395. MR1861384

Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York. MR1453117

Piegorsch, W. W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics* **46** 863–867. MR1085821

Saha, K. and Paul, S. (2005). Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics* **61** 179–185. MR2135858

Tsodikov, A. D., Ibrahim, J. G. and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association* **98** 1063–1078. MR2055496

Yakovlev, A. Y. and Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.

Yin, G. and Ibrahim, J. G. (2005). Cure rate models: A unified approach. *The Canadian Journal of Statistics* **33** 559–570. MR2232380

M. de Castro
V. G. Cancho
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
Caixa Postal 668
13560-970, São Carlos-SP
Brasil
E-mail: mcastro@icmc.usp.br
garibay@icmc.usp.br

J. Rodrigues
Departamento de Estatística
Universidade Federal de São Carlos
Caixa Postal 676
13565-905, São Carlos-SP
Brasil
E-mail: vjosemar@ufscar.br