Computational Statistics and Data Analysis 55 (2011) 3311–3319

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A unified view on lifetime distributions arising from selection mechanisms

Josemar Rodrigues [a,*], N. Balakrishnan [b,c], Gauss M. Cordeiro [d], Mário de Castro [e]

[a] Universidade Federal de São Carlos, Departamento de Estatística, Caixa Postal 676, 13565-905, São Carlos-SP, Brazil

[b] McMaster University, Department of Mathematics and Statistics, Ontario, Canada L8S 4K1

[c] King Saud University, Faculty of Science, Riyadh, Saudi Arabia

[d] Universidade Federal de Pernambuco, Recife-PE, Brazil

[e] Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos-SP, Brazil

ARTICLE INFO

Article history:
Received 28 December 2010
Received in revised form 13 June 2011
Accepted 15 June 2011
Available online 30 June 2011

Keywords:
Competing causes
Conditional personal cure rate
Conditional personal non-cure rate
Probability generating function
Selection mechanism
Survival analysis

ABSTRACT

In this paper, we formulate a flexible density function from the selection mechanism viewpoint (see, for example, Bayarri and DeGroot (1992) and Arellano-Valle et al. (2006)) which possesses nice biological and physical interpretations. The new density function contains as special cases many models that have been proposed recently in the literature. In constructing this model, we assume that the number of competing causes of the event of interest has a general discrete distribution characterized by its probability generating function. This function has an important role in the selection procedure as well as in computing the conditional personal cure rate. Finally, we illustrate how various models can be deduced as special cases of the proposed model.

1. Introduction

Recently, there has been a great interest among statisticians and applied researchers in constructing flexible families of distributions to facilitate better modeling of data. Consequently, a significant progress has been made in developing the generalizations of some well-known lifetime distributions and their successful application to problems in areas such as engineering, environmetrics, economics and biomedical sciences. The purpose of this work is to formulate a unified procedure with a biological and physical interpretation that includes as special cases many of these lifetime distributions. For formulating this procedure, we choose the selection approach discussed by Bayarri and DeGroot (1992) and Arellano-Valle et al. (2006). This selection approach is useful for obtaining flexible distributions from the original model based on the occurrence of some related selection random variables. Moreover, we introduce a new notion, called the conditional personal non-cure rate, for which we give an interpretation in terms of selection or weight function. Another related measure is the conditional personal cure rate which is of interest when, for example, successfully treated cancer patients may die from a cause other than the diagnosed cancer.

The rest of this article is organized as follows. In Section 2, the unified model is developed from the selection mechanism viewpoint and the idea of the conditional personal probability is introduced. In Section 3, many of the recently introduced lifetime distributions are obtained as special cases from the proposed unified model, and some new interpretations from

* Corresponding author.
E-mail address: vjosemar@ufscar.br (J. Rodrigues).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.06.018

a biological viewpoint are given to them. Section 4 deals with some mathematical properties of the unified model. Two applications of some distributions, discussed in Section 3, are given in Section 5. Finally, Section 6 offers some concluding remarks.

2. A unified selection distribution

Selection mechanisms arise when a random sample from the entire population might be too difficult or too expensive to secure and so flexible models must be developed to incorporate this constraint on the observations. We formulate the selection distributions here within a biological context, where the population is restricted to patients not cured from an event of interest such as disease or tumor. In biological context, we mean that the damaged cells are competing to produce detectable tumors. The time for the jth damaged cell (clonogens) to transform into a detectable tumor (promotion time) is denoted by X_j, $j = 1, \ldots, N$, where N denotes the unobservable number of damaged cells that can produce the event of interest. In the sequel, we suppose that N has its probability mass function (pmf) given by

$$p_n = P(N = n), \quad n = 0, 1, \ldots. \tag{1}$$

Let $A_N(s) = \sum_{n=0}^{\infty} p_n s^n$ be the corresponding probability generating function (pgf) for $0 < s < 1$, and p_0 the cure rate. We assume that, conditional on N, that the X_j's are i.i.d. having density function $g(x)$ and survival function $S(x) = 1 - G(x)$. Usually, exponential, piecewise exponential (Chen and Ibrahim, 2001) and Weibull distributions are used to represent $g(x)$.

Given $N = n$ and the lifetime $T = t$, let Z_j, $j = 1, \ldots, n$, be independent random variables, independently of N, following a Bernoulli distribution with success probability $G(t)$ indicating the presence of the jth competing cause (or clonogens) at time t. The discrete variable N_t, representing the total number of competing causes among the N initial competing causes that are present at time t, is then given by

$$N_t = \begin{cases} Z_1 + Z_2 + \cdots + Z_N, & \text{if } N > 0, \\ 0, & \text{if } N = 0. \end{cases} \tag{2}$$

It follows from the fundamental formula for conditional probabilities that

$$P(N_t = j) = \sum_{n=j}^{\infty} p_n \overbrace{P(N_t = j | N = n)}^{\text{Binomial}(n, G(t))},$$

and its corresponding pgf (Feller, 1968) is

$$A_{N_t}(s) = A_N[1 - (1 - s)G(t)]. \tag{3}$$

The long-term survival function (Rodrigues et al., 2008) can be obtained from (3) as

$$S_{\text{Pop}}(t) = P(T \geq t) = P(N_t = 0) = A_{N_t}(0) = A_N[S(t)], \tag{4}$$

where $A_N(.)$ is the pgf of the discrete random variable N.

Motivated by the work of Arellano-Valle et al. (2006), we start with a definition of a selection distribution and its association with the pgf $A_{N_t}(s)$ and density function $g(x)$ of the promotion time random variable X. First, we assume that the population is divided into two sub-populations of cured and non-cured patients defined by the following binary random variable for any time t:

$$U_t = \begin{cases} 1, & \text{if } N_t \geq 1, \\ 0, & \text{if } N_t = 0, \end{cases} \tag{5}$$

where $P(U_t = 1) = 1 - P(N_t = 0) = 1 - p_0$.

Definition 2.1 (*Selection Distribution*)**.** Let T be a non-negative lifetime random variable and X the promotion time with probability density function (pdf) $g(x)$. We define the selection distribution of T as the conditional distribution of X, given $U_t = 1$.

This definition simply states that the selection probability distribution of T is the probability distribution of X, truncated by non-cured patients. We show that this viewpoint is quite useful to obtain new classes of flexible lifetime distributions and also to unify many models proposed recently in the literature.

Indeed, if X in Definition 2.1 has pdf $g(x)$, then T has a pdf $f_T(t)$ given by

$$f_T(t) = \frac{g(t)\, P(U_t = 1 \mid X \leq t)}{P(U_t = 1)} = \frac{g(t)\, P(U_t = 1 \mid X \leq t)}{1 - p_0}. \tag{6}$$

In fact, (6) can be expressed as a weighted distribution (Bayarri and DeGroot, 1992)

$$f_T(t) = \frac{w(t)\, g(t)}{E[w(X)]}, \tag{7}$$