Reliability Engineering and System Safety 129 (2014) 36–45

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

A new class of distribution functions for lifetime data

Filippo Domma*, Francesca Condino

Department of Economics, Statistics and Finance, University of Calabria, Via Pietro Bucci, Cubo 0C, 87036 Arcavacata di Rende (CS), Italy

ARTICLE INFO

Article history:
Received 28 May 2013
Received in revised form
18 March 2014
Accepted 27 April 2014
Available online 5 May 2014

Keywords:
Hazard rate
Differential equation
Flexibility
Log-concavity

ABSTRACT

This paper deals with the issue of building a parametric model from the empirical and/or qualitative information about the hazard rate. We propose a new class of models for survival data analysis. This class is characterized by a distribution function which includes, in its expression, a function that defines the sign of the first derivative of a monotonic transformation of the hazard rate. We show that certain parametric models used in survival analysis belong to the proposed class. Finally, by using the proposed method, we build two new distributions which allow us to achieve a highly flexible hazard rate. The first one is based on an m-degree polynomial and allows us to get BT, IFR and UBT-BT hazard rates, while the second, based on trigonometric functions, enables us to obtain monotonically increasing or decreasing hazard rates or hazard rates with a non-monotonic behavior. The usefulness of the new method is illustrated through two applications to real data.

1. Introduction

In different fields of applications the main focus of researchers is on the so-called time-to-event, i.e. the time taken for an event to occur. Certainly, in simple cases the event is the death of human beings, animals, cells, plants, etc. and the time of death is studied in demography, biology, botany, etc. Moreover, it is also very important in economic and social studies, for example, to have information about the required time to acceptance of a job offer for an unemployed person; in industrial applications, information about the lifetime of a unit or some component in a unit is necessary to assess the reliability of a system. The statistical studies dealing with these problems are classified, according to the field of application, as survival analysis, lifetime data analysis, reliability analysis, duration analysis and so on. In this paper, we propose a new class of models for the analysis of these kinds of data.

Let T be a non-negative random variable (rv), usually representing the lifetime of an individual or unit, with distribution function (df), survival function (sf) and probability density function (pdf), respectively, denoted by $F(t;\boldsymbol{\theta})$, $S(t;\boldsymbol{\theta}) = 1 - F(t;\boldsymbol{\theta})$ and $f(t;\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ with $p \geq 1$ and $p \in \mathbb{N}$. Moreover, let $f(t;\boldsymbol{\theta})$ be continuous and twice differentiable on $(0, \infty)$. One important function in survival analysis is the hazard rate (or failure rate) defined as

$$h(t;\boldsymbol{\theta}) = \lim_{\Delta t \to 0} \frac{P[t \leq T < t + \Delta t | T > t]}{\Delta t} = \frac{f(t;\boldsymbol{\theta})}{S(t;\boldsymbol{\theta})}. \tag{1}$$

That is, $h(t;\boldsymbol{\theta}) \cdot \Delta t$ can be thought of as the conditional probability that an event occurs in the interval $[t, t+\Delta t)$ given that the event has not occurred before time t.

In the literature, it is usual to denote the strictly increasing (decreasing) failure rate as IFR (DFR) and the hazard rate with a minimum or a maximum as *Bathtub* (BT) and *Upside-down Bathtub* (UBT), respectively. It is worth noting that in the continuous case the function in (1) is not a density function, since $h(t;\boldsymbol{\theta}) \geq 0$ and $\int_0^\infty h(t;\boldsymbol{\theta})\,dt = \infty$ (see [1, p. 7; 2, p. 9]). For a proper model, the condition $\int_0^\infty h(t;\boldsymbol{\theta})\,dt = \infty$ implies that $\lim_{t\to\infty} F(t;\boldsymbol{\theta}) = 1$. We highlight the fact that, in the literature on duration data, situations where the event of interest never takes place are provided, so that $\lim_{t\to\infty} F(t;\boldsymbol{\theta}) = k$, with $k < 1$. As a consequence, $\int_0^\infty h(t;\boldsymbol{\theta})\,dt$ does not diverge. Such models are called *defective* or *subprobability* models [3, p. 9; 4].

In survival analysis, the choice of a parametric model is often based on theoretical arguments dealing with the failure mechanism or aging properties [5] and/or on empirical analysis such as, for example, the total time on test (TTT) plot [6] or even on the empirical hazard rate to appraise the shape of the hazard rate. The linkage between probability density function and hazard rate expressed by

$$f(t;\boldsymbol{\theta}) = h(t;\boldsymbol{\theta}) \exp\left\{ - \int_0^t h(u;\boldsymbol{\theta})\,du \right\} \tag{2}$$

provides a tool for constructing a new pdf that corresponds to a hazard profile observed in the data. Thus, it is possible to develop a model by means of the hazard rate and determine later the distribution function linked to the hazard rate.

* Corresponding author.
E-mail addresses: filippo.domma@unical.it (F. Domma), francesca.condino@unical.it (F. Condino).

http://dx.doi.org/10.1016/j.ress.2014.04.026
0951-8320/

Over the past 30 years, a considerable part of the literature related to the building and specification of models for survival data has been directed towards the search for hazard rate with a bathtub shape. Among others, we recall the article by Glaser [7] who proposes a procedure to analyze the behavior of a particularly complex hazard rate, based on the reciprocal of $h(t;\boldsymbol{\theta})$. For a detailed review of the literature on this topic, see the work of Rajarshi S. and Rajarshi M.B. [8].

In the last two decades different models with a bathtub-shaped failure rate have been proposed. Some of these are based on the sum of two Weibull [9], or on the sum of two Burr XII random variables [10], or also on the sum of two inverse Weibull [11]. Others are based on particular generalizations of the Weibull random variables [12,13], or on transformations of the *total time on test* [14]; in this context, we remember the work of Chen [15] and Wang et al. [16]. Models with highly flexible hazard rate (IFR, DFR, UBT, BT) are found in the works of Mudholkar et al. [17,18] and Saha and Hilton [19]. More recently, a new generalization of the Weibull distribution has been proposed in order to get bathtub shaped failure rates (for a review see [20]). In this context Cordeiro et al. [21] introduce the beta extended Weibull family by using the beta generated distribution method. Such a method has been used by different authors also to achieve a more flexible hazard rate (see, for example, [22–24]). Finally, it is worth noting that Slymen and Lachembruch [25], and Louzado-Neto [26,27], Louzado-Neto et al. [28] and Bebbington et al. [29] discuss the issue of multi-modality of the hazard function.

In this paper, we propose a new class of distribution functions for modeling survival data, based on a transformation of the hazard rate. This class is very flexible because it contains a function, $\varphi(\cdot)$, that determines the sign of the derivate of the hazard rate (Section 2). In Section 3, some parametric models are reported as special cases of the proposed class. Starting from the proposed method, in Section 4 we construct and study two new parametric models, which provide a highly flexible hazard rate.

2. A new class of distribution functions

In the literature there are different systems of statistical distributions based on the specification of a differential equation. Certainly, the most widely known is the Pearson system in which every member has a density function $f(t;\boldsymbol{\theta})$ which satisfies the differential equation $f'(t;\boldsymbol{\theta})/f(t;\boldsymbol{\theta})=(t-m)/(at^2+bt+c)$, where m, a, b and c are constants determining a particular type of solution. Unlike the Pearson system, the differential equation specified by Burr [40] describes the distribution function and not the density function; in particular, the *df*'s of all Burr distributions satisfy the following differential equation: $F'(t;\boldsymbol{\theta})=F(t;\boldsymbol{\theta})[1-F(t;\boldsymbol{\theta})]g(t;\boldsymbol{\theta})$, where $g(\cdot)$ is a suitable nonnegative function. Several authors specify the differential equation so as to reproduce the characteristics of regularity observed in a given field of inquiry (see, for example, [30–33]); the functional form of distribution function (or density function) is the solution of the corresponding differential equation. In this section, we propose a new class of distribution function which turns out to be a solution of the following differential equation:

$$y''(t)=\alpha y'(t)\varphi(t) \tag{3}$$

where α is a positive constant and $\varphi(\cdot)$ is a real-valued function of t. Rearranging (3) we get

$$\frac{y''(t)}{y'(t)}=\alpha\varphi(t). \tag{4}$$

Integrating twice *w.r.t.* t, the set of solutions of the differential equation defined in (3) is given by

$$y(t)=c_2+e^{c_1}\int e^{\alpha\int\varphi(t)\,dt}\,dt$$

where c_1 and c_2 are two constants of integration. Putting $y(t)=-\ln[S(t;\boldsymbol{\theta})]$, after simple algebra, we obtain the following particular solution:

$$F(t;\boldsymbol{\theta})=1-k_2\exp\left\{-k_1\int e^{\alpha\int\varphi(t;\boldsymbol{\gamma})\,dt}\,dt\right\} \tag{5}$$

with $k_1=e^{c_1}$ and $k_2=e^{-c_2}$, and $\boldsymbol{\theta}=(\alpha,k_1,k_2,\boldsymbol{\gamma})$. The probability density function corresponding to (5) is given by

$$f(t;\boldsymbol{\theta})=k_1\cdot k_2\exp\left\{\alpha\int\varphi(t;\boldsymbol{\gamma})\,dt\right\}\times\exp\left\{-k_1\int\exp\left\{\alpha\int\varphi(t;\boldsymbol{\gamma})\,dt\right\}dt\right\}. \tag{6}$$

It is useful to note that $y(t)=-\ln[S(t;\boldsymbol{\theta})]$ is the integrated hazard rate, $y'(t)=h(t;\boldsymbol{\theta})$ is the hazard rate and $y''(t)=h'(t;\boldsymbol{\theta})$ is the first derivate *w.r.t.* t of the hazard rate. This allows us to clarify the role of function $\varphi(t;\boldsymbol{\gamma})$; in fact, the sign of $\varphi(\cdot;\cdot)$ in (3), when $y(t)=-\ln[S(t;\boldsymbol{\theta})]$, determines the sign of the first derivative of the hazard rate, i.e. $y''(t)$ and therefore, defines the behavior of the hazard rate. In particular, if we choose a positive (negative) function, $\varphi(t;\boldsymbol{\gamma})$, for all t, then we obtain an increasing (decreasing) hazard rate. While, if we choose a function such that for any $t\in(0,t^*)$, $\varphi(t;\boldsymbol{\gamma})>0$, $\varphi(t^*;\boldsymbol{\gamma})=0$ and for $t>t^*$, $\varphi(t;\boldsymbol{\gamma})<0$ then from (3) the hazard rate will be *UBT*. Similarly, to construct a *BT* hazard rate, we must choose a function such that for any $t\in(0,t^*)$, $\varphi(t;\boldsymbol{\gamma})<0$, $\varphi(t^*;\boldsymbol{\gamma})=0$ and for $t>t^*$, $\varphi(t;\boldsymbol{\gamma})>0$. Ultimately, we can conclude that the function $\varphi(t;\boldsymbol{\gamma})$, from Eq. (5) provides a mathematically equivalent way of specifying the distribution function of a continuous nonnegative random variable.

Remark. In a different context, when $y(t)$ is an increasing function, the left side of equation (4) defines the Arrow–Pratt coefficient of absolute risk aversion, so that function $y(t)=-\ln[S(t;\boldsymbol{\theta})]$ can be thought of as a utility function. Consequently, the sign of $\varphi(t;\boldsymbol{\gamma})$ determines the shape of the utility function. So, for example, if $\varphi(t;\boldsymbol{\gamma})$ is negative (positive) for all t, then the utility function is concave (convex), if $\varphi(t;\boldsymbol{\gamma})>(<)0$ for $t\in(0,t^*)$, $\varphi(t^*;\boldsymbol{\gamma})=0$ and for $t>t^*$, $\varphi(t;\boldsymbol{\gamma})<(>)0$ then the utility function is S-shaped (reversed S-shaped); for the non-canonical shape of utility function see, for example, [34–36]. From the aforesaid, Eq. (4) provides a link between the behavior of the utility function and the behavior of the hazard rate.

Moreover, we highlight the fact that in demographic studies the left side of Eq. (4) is called *age-specific rate of mortality change with age* [37].

2.1. Properties of the new class of distributions

In this section, we show that some characteristics and structural properties of the new class of distribution functions depend on the behavior of function $\varphi(t;\boldsymbol{\gamma})$. The results of this section show that function $\varphi(t;\boldsymbol{\gamma})$ not only determines the behavior of the hazard rate and allows us to build new distribution functions but also affects certain properties of the new distribution functions.

First of all, we remark that function $\varphi(t;\boldsymbol{\gamma})$ is a transformation of the hazard rate. Indeed, given that $y(t)=-\ln[S(t;\boldsymbol{\theta})]$, from relations (1) and (3), we have

$$\varphi(t;\boldsymbol{\gamma})=\frac{1}{\alpha}\frac{y''(t)}{y'(t)}=\frac{1}{\alpha}\frac{\partial\ln[h(t;\boldsymbol{\theta})]}{\partial t}.$$

Consequently, we can write the exponent of (5) as $-k_1\int h(t;\boldsymbol{\theta})\,dt$ and recalling the following relations:

$$\int_0^t h(u;\boldsymbol{\theta})\,du=\int_0^t\frac{-\partial S(u;\boldsymbol{\theta})}{\partial u}\cdot\frac{1}{S(u;\boldsymbol{\theta})}\,du=-\ln[S(u;\boldsymbol{\theta})]_0^t=-\ln[S(t;\boldsymbol{\theta})],$$