Computational Statistics and Data Analysis 56 (2012) 698–718

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

A log-linear regression model for the β-Birnbaum–Saunders distribution with censored data

Edwin M.M. Ortega [a], Gauss M. Cordeiro [b], Artur J. Lemonte [c,*]

[a] Departamento de Ciências Exatas, Universidade de São Paulo, Piracicaba/SP, Brazil
[b] Departamento de Estatística, Universidade Federal de Pernambuco, Recife/PE, Brazil
[c] Departamento de Estatística, Universidade de São Paulo, São Paulo/SP, Brazil

ARTICLE INFO

Article history:
Received 18 April 2011
Received in revised form 25 July 2011
Accepted 16 September 2011
Available online 22 September 2011

Keywords:
Birnbaum–Saunders distribution
Censored data
Cure fraction
Generating function
Lifetime data
Maximum likelihood estimation
Regression model

ABSTRACT

The β-Birnbaum–Saunders (Cordeiro and Lemonte, 2011) and Birnbaum–Saunders (Birnbaum and Saunders, 1969a) distributions have been used quite effectively to model failure times for materials subject to fatigue and lifetime data. We define the log-β-Birnbaum–Saunders distribution by the logarithm of the β-Birnbaum–Saunders distribution. Explicit expressions for its generating function and moments are derived. We propose a new log-β-Birnbaum–Saunders regression model that can be applied to censored data and be used more effectively in survival analysis. We obtain the maximum likelihood estimates of the model parameters for censored data and investigate influence diagnostics. The new location-scale regression model is modified for the possibility that long-term survivors may be presented in the data. Its usefulness is illustrated by means of two real data sets.

1. Introduction

The fatigue is a structural damage that occurs when a material is exposed to stress and tension fluctuations. Statistical models allow to study the random variation of the failure time associated to materials exposed to fatigue as a result of different cyclical patterns and strengths. The most popular model for describing the lifetime process under fatigue is the Birnbaum–Saunders (BS) distribution (Birnbaum and Saunders, 1969a,b). The crack growth caused by vibrations in commercial aircrafts motivated these authors to develop this new family of two-parameter distributions for modeling the failure time due to fatigue under cyclic loading. Relaxing some assumptions made by Birnbaum and Saunders (1969a), Desmond (1985) presented a more general derivation of the BS distribution under a biological framework. The relationship between the BS and inverse Gaussian distributions was explored by Desmond (1986) who demonstrated that the BS distribution is an equal-weight mixture of an inverse Gaussian distribution and its complementary reciprocal. The two-parameter BS model is also known as the fatigue life distribution. It is an attractive alternative distribution to the Weibull, gamma and log-normal models, since its derivation considers the basic characteristics of the fatigue process. Furthermore, it has the appealing feature of providing satisfactory tail fitting due to the physical justification that originated it, whereas the Weibull, gamma and log-normal models typically provide a satisfactory fit in the middle portion of the data, but oftentimes fail to deliver a good fit at the tails, where only a few observations are generally available.

In many medical problems, for example, the lifetimes are affected by explanatory variables such as the cholesterol level, blood pressure, weight and many others. Parametric models to estimate univariate survival functions for censored

* Corresponding author.
E-mail address: arturlemonte@gmail.com (A.J. Lemonte).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.09.018

data regression problems are widely used. Different forms of regression models have been proposed in survival analysis. Among them, the location-scale regression model (Lawless, 2003) is distinguished since it is frequently used in clinical trials. Recently, the location-scale regression model has been applied in several research areas such as engineering, hydrology and survival analysis. Lawless (2003) also discussed the generalized log-gamma regression model for censored data. Xie and Wei (2007) developed the censored generalized Poisson regression models, Barros et al. (2008) proposed a new class of lifetime regression models when the errors have the generalized BS distribution, Carrasco et al. (2008) introduced a modified Weibull regression model, Silva et al. (2008) studied a location-scale regression model using the Burr XII distribution and Silva et al. (2009) worked with a location-scale regression model suitable for fitting censored survival times with bathtub-shaped hazard rates. Ortega et al. (2009a,b) proposed a modified generalized log-gamma regression model to allow the possibility that long-term survivors may be presented in the data, Hashimoto et al. (2010) developed the log-exponentiated Weibull regression model for interval-censored data and Silva et al. (2010) discussed a regression model considering the Weibull extended distribution.

For the first time, we define a location-scale regression model for censored observations, based on the β-Birnbaum–Saunders (βBS for short) introduced by Cordeiro and Lemonte (2011), referred to as the log-βBS (LβBS) regression model. The proposed regression model is much more flexible than the log-BS regression model proposed by Rieck and Nedelman (1991). Further, some useful properties of the proposed model to study asymptotic inference are investigated. For some recent references about the log-BS linear regression model the reader is refereed to Lemonte et al. (2010), Lemonte and Ferrari (2011a,b,c), Lemonte (2011) and references therein. A log-BS nonlinear regression model was proposed by Lemonte and Cordeiro (2009); see also Lemonte and Cordeiro (2010) and Lemonte and Patriota (2011).

Another issue tackled is when in a sample of censored survival times, the presence of an immune proportion of individuals who are not subject to death, failure or relapse may be indicated by a relatively high number of individuals with large censored survival times. In this note, the log-βBS model is modified for the possible presence of long-term survivors in the data. The models attempt to estimate the effects of covariates on the acceleration/deceleration of the timing of a given event and the surviving fraction, that is, the proportion of the population for which the event never occurs. The logistic function is used to define the regression model for the surviving fraction.

The article is organized as follows. In Section 2, we define the LβBS distribution. In Section 3, we provide expansions for its moment generating function (mgf) and moments. In Section 4, we propose a LβBS regression model and estimate the model parameters by maximum likelihood. We derive the observed information matrix. Local influence is discussed in Section 5. In Section 6, we propose a LβBS mixture model for survival data with long-term survivors. In Section 7, we show the flexibility, practical relevance and applicability of our regression model by means of two real data sets. Section 8 ends with some concluding remarks.

2. The LβBS distribution

The BS distribution is a very popular model that has been extensively used over the past decades for modeling failure times of fatiguing materials and lifetime data in reliability, engineering and biological studies. Birnbaum and Saunders (1969a,b) define a random variable T having a BS distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, $T \sim \mathrm{BS}(\alpha, \beta)$ say, by $T = \beta[\alpha Z/2 + \{(\alpha Z/2)^2 + 1\}^{1/2}]^2$, where Z is a standard normal random variable. Its cumulative distribution function (cdf) is defined by $G(t) = \Phi(v)$, for $t > 0$, where $v = \alpha^{-1}\rho(t/\beta)$, $\rho(z) = z^{1/2} - z^{-1/2}$ and $\Phi(\cdot)$ is the standard normal distribution function. Since $G(\beta) = \Phi(0) = 1/2$, the parameter β is the median of the distribution. For any $k > 0$, $kT \sim \mathrm{BS}(\alpha, k\beta)$. The probability density function (pdf) of T is then $g(t) = \kappa(\alpha, \beta)t^{-3/2}(t+\beta)\ \exp\{-\tau(t/\beta)/(2\alpha^2)\}$, for $t > 0$, where $\kappa(\alpha, \beta) = \exp(\alpha^{-2})/(2\alpha\sqrt{2\pi\beta})$ and $\tau(z) = z + z^{-1}$. The fractional moments of T are $E(T^p) = \beta^p I(p, \alpha)$, where

$$I(p, \alpha) = \frac{K_{p+1/2}(\alpha^{-2}) + K_{p-1/2}(\alpha^{-2})}{2K_{1/2}(\alpha^{-2})} \tag{1}$$

and the function $K_\nu(z)$ denotes the modified Bessel function of the third kind with ν representing its order and z the argument (see Watson, 1995). Kundu et al. (2008) studied the shape of its hazard function. Results on improved statistical inference for this model are discussed by Wu and Wong (2004) and Lemonte et al. (2007, 2008). Díaz-García and Leiva (2005) proposed a new family of generalized BS distributions based on contoured elliptical distributions, whereas Guiraud et al. (2009) introduced a non-central version of the BS distribution.

The βBS distribution (Cordeiro and Lemonte, 2011), with four parameters $\alpha > 0$, $\beta > 0$, $a > 0$ and $b > 0$, extends the BS distribution and provides more flexibility to fit various types of lifetime data. Its cdf is given by $F(t) = I_{\Phi(v)}(a, b)$, where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function, $\Gamma(\cdot)$ is the gamma function, $I_y(a, b) = B_y(a, b)/B(a, b)$ is the incomplete beta function ratio and $B_y(a, b) = \int_0^y \omega^{a-1}\,(1-\omega)^{b-1}d\omega$ is the incomplete beta function. The density function of T has the form (for $t > 0$)

$$f_T(t) = \frac{\kappa(\alpha, \beta)}{B(a, b)} t^{-3/2}(t+\beta)\ \exp\left\{-\tau(t/\beta)/(2\alpha^2)\right\} \Phi(v)^{a-1}\{1-\Phi(v)\}^{b-1}. \tag{2}$$

The βBS distribution contains, as special sub-models, the exponentiated BS (EBS), Lehmann type-II BS (LeBS) and BS distributions when $b = 1$, $a = 1$ and $a = b = 1$, respectively. If T is a random variable with density function (2), we