Computational Statistics and Data Analysis 97 (2016) 169–183

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A bivariate Birnbaum–Saunders regression model

Filidor Vilca [a,*], Renata G. Romeiro [a], N. Balakrishnan [b]

[a] Departamento de Estatística, Universidade Estadual de Campinas, São Paulo, Brazil

[b] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:
Received 12 September 2014
Received in revised form 19 June 2015
Accepted 1 December 2015
Available online 11 December 2015

Keywords:
Sinh-normal distribution
Birnbaum–Saunders distribution
Log-linear model
Moment estimators
Consistent estimators
Maximum-likelihood estimators
Fisher information matrix
Asymptotic normality

ABSTRACT

In this work, we propose a bivariate Birnbaum–Saunders regression model through the use of bivariate Sinh-normal distribution. The proposed regression model has its marginal as the Birnbaum–Saunders regression model of Rieck and Nedelman (1991), which has been discussed extensively by various authors with natural applications in survival and reliability studies. This bivariate regression model can be used to analyze correlated log-lifetimes of two units, in which the dependence structure between observations arises from the bivariate normal distribution.

The main aim of this paper is to propose a bivariate Birnbaum–Saunders regression model and discuss some of its properties. Specifically, we have developed the moment estimation, the maximum likelihood estimation and the observed Fisher information matrix. Hypothesis testing is also performed by the use of the asymptotic normality of the maximum-likelihood estimators. Finally, the results of simulation studies as well as an application to a real data set are presented to illustrate the model and all the inferential methods developed here.

1. Introduction

It is quite common in survival and reliability analyses to observe two or more different lifetimes on the same subject; see, for example, Johnson and Wichern (1999), Klein and Moeschberger (1997), and Kundu et al. (2010). This indeed prompted the generalization of many univariate statistical models to multivariate models (Kotz et al., 2000; Balakrishnan and Lai, 2009). Usually, the first step in this process is to study the bivariate case, as can be seen in the works of Basu (1971), Oakes (1989), Kundu and Gupta (2009), and Navarro and Sarabia (2013). Kundu et al. (2010) developed the bivariate Birnbaum–Saunders (BS) distribution and pointed out some important characteristics and properties of that family of distributions; it provides an elegant extension of the well-known univariate BS distribution (Birnbaum and Saunders, 1969a,b). A further extension to the multivariate case has also been developed by Kundu et al. (2010, 2013) and Vilca et al. (2014a) who established several interesting and attractive properties of the bivariate and multivariate BS distributions by using the known properties of symmetric distributions.

In situations when some parameters of the model depend on covariates, one way of modeling lifetimes is to consider the logarithm of the positive random variable; for example, if T is a positive random variable following a BS distribution, $Y = \log(T)$ follows a Sinh-normal (SN) distribution (Rieck, Unpublished). This resulting distribution has received considerable attention in lifetime regression models, especially when the lifetime data involve modeling the scale of parameter of the BS distribution when it depends on covariates.

* Correspondence to: Departamento de Estatística, Universidade Estadual de Campinas, Caixa Postal 6065, SP, Brazil.
E-mail addresses: fily@ime.unicamp.br, filyps@yahoo.com.br (F. Vilca).

http://dx.doi.org/10.1016/j.csda.2015.12.003

1. *The BS and SN distributions*

The interest in the BS distribution arises from its attractive properties and its close relationship with the normal distribution. From the viewpoint of applications, it is a more attractive alternative to well-known Weibull, log–logistic, log-normal, gamma and inverse Gaussian models, since its derivation takes into account the basic characteristics of a fatigue process. The random variable T that follows a BS distribution is related to the normal distribution through the stochastic representation

$$T = \frac{\beta}{4}\left[\alpha Z + \sqrt{(\alpha Z)^2 + 4}\right]^2, \tag{1}$$

where $Z \sim \mathrm{N}(0, 1)$, and $\alpha > 0$ and $\beta > 0$ are shape and scale parameters, respectively. This distribution is usually denoted by $T \sim \mathrm{BS}(\alpha, \beta)$; for more details, see Birnbaum and Saunders (1969a,b). To introduce covariates into the BS distribution, the distribution of the transformed response variable $Y = \log(T)$ is usually considered; see Rieck and Nedelman (1991). The corresponding distribution is a special case of the SN distribution having its probability density function (pdf) as

$$f_Y(y) = \frac{1}{\sigma}\,\phi(\xi_2)\,\xi_1, \quad y \in \mathbb{R}, \tag{2}$$

where $\xi_1 = \xi_1(y; \alpha, \mu, \sigma) = \frac{2}{\alpha}\cosh\left(\frac{y-\mu}{\sigma}\right)$, $\xi_2 = \xi_2(y; \alpha, \mu, \sigma) = \frac{2}{\alpha}\sinh\left(\frac{y-\mu}{\sigma}\right)$, and $\phi(\cdot)$ denotes the pdf of the standard normal distribution. In this case, the notation $Y \sim SN(\alpha, \mu, \sigma)$ will be used.

2. *The multivariate BS distribution*

Kundu et al. (2013) proposed the multivariate BS distribution, where the multivariate normal distribution is used as a base kernel for the transformation. In fact, a random vector p-variate $\mathbf{T} = (T_1, \ldots, T_p)^\top$ is said to have a multivariate BS distribution if its cumulative distribution function (cdf) is of the form

$$F_{\mathbf{T}}(\mathbf{t}) = \Phi_p\big(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \boldsymbol{\Sigma}\big), \tag{3}$$

where $a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (a_{t_1}(\alpha_1, \beta_1), \ldots, a_{t_p}(\alpha_p, \beta_p))^\top$, with $a_{t_j}(\alpha_j, \beta_j) = (1/\alpha_j)\big(\sqrt{t_j/\beta_j} - \sqrt{\beta_j/t_j}\big)$, $j = 1, \ldots, p$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ in $\mathbb{R}^p_+$, which is the positive part of $\mathbb{R}^p$, and $\Phi_p\big(.; \mathbf{0}, \boldsymbol{\Sigma}\big)$ is the cdf of $\mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, with $\boldsymbol{\Sigma}$ being a $p \times p$ positive-definite correlation matrix, For the distribution in (3), we use the notation $\mathbf{T} \sim \mathrm{BS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$. For $p = 2$, we have the bivariate BS distribution that is denoted by $\mathbf{T} \sim \mathrm{BS}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho)$, where ρ is the correlation coefficient in the bivariate normal distribution.

From the bivariate BS distributions established by Kundu et al. (2010), it is natural to derive the bivariate log-BS distribution, or simply the bivariate SN distribution. This distribution can be used along the same lines as the univariate BS distributions in the context of regression models, following the idea of Rieck and Nedelman (1991). Recently, Lemonte (2013) developed a multivariate BS regression model in which the elements of the response vector are all independent. However, in bivariate or multivariate settings, we find many situations wherein it is necessary to consider a dependence structure between observations, especially when it involves the multivariate normal distribution. Thus, the main aim of this paper is to propose a bivariate BS regression model by considering a dependence structure between the observations and discuss some of its properties such as the moment estimation, the maximum likelihood estimation and the observed Fisher information matrix. We also discuss hypothesis testing based on the asymptotic normality of the maximum-likelihood estimators. All these results form a supplement to the results of Kundu et al. (2010).

The rest of this paper is organized as follows. In Section 2, we introduce the multivariate SN distribution and discuss some of its properties such as its marginal distributions, conditional distributions and some transformations. Section 3 describes the bivariate BS regression model. Section 4 describes some estimation methods such as the modified moment and the maximum likelihood methods. Section 5 presents the results of a simulation study along with an application to a real data set. Finally, some concluding remarks are made in Section 6.

2. The multivariate Sinh-normal distribution

Following the idea of bivariate Sinh-normal (SN) distribution introduced by Kundu (2015) and the skewed Sinh-normal distribution of Leiva et al. (2010), we now define the multivariate SN distribution through the following stochastic representation for random variables $Y_1, \ldots, Y_p$:

$$Y_j = \mu_j + \sigma_j \operatorname{arcsinh}(\alpha_j Z_j/2), \quad j = 1, \ldots, p, \tag{4}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_p)^\top \sim N_p(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi}$ being a $p \times p$ positive-definite correlation matrix. The multivariate random vector $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$ is said to have a multivariate SN distribution with shape, location, and scale parameters given by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^\top \in \mathbb{R}^p_+$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_p)^\top \in \mathbb{R}^p_+$, respectively, and it will be denoted by $\mathbf{Y} \sim \mathrm{SN}_p(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\Psi})$. Furthermore, as in the univariate case, each Z_j in (4) can be expressed as

$$Z_j = \frac{2}{\alpha_j}\sinh\left(\frac{Y_j - \mu_j}{\sigma_j}\right), \quad j = 1, \ldots, p. \tag{5}$$