Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 5927 – 5939
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A new family of life distributions for dependent data: Estimation

José A. Díaz-García[a,*], José Ramón Domínguez-Molina[b]

[a]*Department of Statistics and Computation, Universidad Autónoma Agraria Antonio Narro, 25315 Buenavista, Saltillo, Coahuila, México*
[b]*Centro de Investigación en Matemáticas, A.C., Gerencia de Consultoría Estadística, 36240, Jalisco S/N, Mineral de Valenciana, Guanajuato, Guanajuato, México*

Received 21 February 2005; received in revised form 3 November 2006; accepted 8 November 2006
Available online 1 December 2006

Abstract

One of the biggest problems in reliability analysis is determining an appropriate distribution of life data. Therefore, this paper develops the estimation aspect of a family of life distributions obtained from spherical distributions. Additionally, a new family of life distributions is proposed for dependent life data, together with an optimization algorithm based on the simulated annealing method. This algorithm is very efficient for optimization purposes and does not require any manipulation of the log-likelihood functions for the distributions proposed in this study.

MSC: 62E15; 62N05

Keywords: Generalized Birnbaum–Saunders distribution; Dependent life data; Life distributions; Reliability analysis; Elliptical distributions; Simulated annealing; *SIC*; Global optimization

1. Introduction

A very important area in the analysis of parametric survival and reliability is the study of probability distributions in order to model the faults in a product and/or the lifetime of a product or entity.

The Birnbaum–Saunders distribution was derived from a model showing that faults are caused by the development and growth of a dominant crack (see Birnbaum and Saunders, 1969a). Subsequently Díaz-García and Leiva-Sánchez (2005) suggested obtaining the Birnbaum–Saunders distribution from one of elliptical contours, rather than from the normal distribution, thus creating a whole family of lifetime distributions, in which there are multimodal distributions, those without moments, those with heavier or lighter tails, etc. (see Section 2). Most probabilistic models intended to describe lifetime data are chosen for one or more of the following reasons (Tobias, 2004):

- There exists a physical or statistical argument that, theoretically, corresponds to the fault mechanism.
- A particular model has been used previously and successfully for an identical or very similar fault mechanism.
- The model is convenient, as it provides an empirically adequate fit to the lifetime data.

* Corresponding author. Tel.: +52 844 4110293; fax: +52 844 4110228.
E-mail addresses: jadiaz@uaaan.mx (J.A. Díaz-García), jrdguez@cimat.mx (J.R. Domínguez-Molina).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.11.010

Whatever the method that is used to choose it, the model must be logical and must pass visual tests of fitness as well as statistical criteria. A common problem when few data are possessed is that many statistical models are so flexible that they seem to fit them very well, which is why arguments must exist to justify the use of a given distribution. For example, it has been reported that the extreme value-type argument justifies the use of the Weibull distribution, while the multiplicative degradation argument justifies the lognormal distribution and the fatigue argument justifies the use of the Birnbaum–Saunders distribution (Tobias, 2004). An additional problem in choosing a distribution is that, in general, the lifetime data in the sample are assumed to be independent. This assumption is not always justified, for example in an aquarium containing tropical fish, the lifetime of one fish is not independent of the lifetimes of the others in the aquarium. It is well known that when there is less competition for space and food, the lifetime of a fish is extended. Thus, when one fish dies, the others in the aquarium will undoubtedly live longer (under stable conditions). In this case, it is no longer possible to define the likelihood function as the product of the marginal densities; it is now necessary to directly propose the joint density function of the sample. Thus, the stochastic dependence between the variables $t_1, \ldots, t_n$ is taken into account in the model by the proposal of a joint distribution for the sample (likelihood function) in which there is a stochastic dependence among its elements t_i.

The population being studied is a univariate one. From this population, we extract a sample of size n, $\mathbf{t}' = (t_1, \ldots, t_n)$; which has an n-variate distribution (likelihood function) given by

$$L(\boldsymbol{\theta}; t_1, \ldots, t_n) = \begin{cases} \prod_{i=1}^{n} f_{T_i}(t_i; \boldsymbol{\theta}) & \text{independent case,} \\ f_{\mathbf{T}}(\mathbf{t}; \boldsymbol{\theta}) & \text{dependent case.} \end{cases}$$

Such densities are two alternative ways of defining likelihoods to obtain estimators of the parameters $\boldsymbol{\theta}$, assuming independence or dependence in the sample, respectively.

A problem that has been addressed by some authors, see for example Birnbaum and Saunders (1969b), is the one of the reparameterization or algebraic manipulation of the log-likelihood in order to have solutions of the likelihood equations in closed form. This issue has been present in problems in areas such as nonlinear regression and estimation of variance components. With these methodologies, what is normally required is the algebraic or numerical calculation (depending on the case) of the first and second derivatives of the log-likelihood function. Additionally, we must analyze the log-likelihood function in greater detail when it is multimodal, see Rieck and Nedelman (1991). These and other problems can be avoided or overcome by the application of alternative methods of optimization. Heuristic methods have recently played an important role in optimizing all kinds of functions arising in many areas of knowledge, especially that of statistical methodology, see Winker and Gilli (2004). Among the methods of heuristic optimization, simulated annealing (SA) stands out for its simplicity and high efficiency (Azencott, 1992).

The present paper includes the maximum likelihood estimators for the generalized Birnbaum–Saunders family of distributions, see Díaz-García and Leiva-Sánchez (2005). It also proposes a new family of distributions for the case in which the lifetime data in the sample are not independent. The maximum likelihood estimators of its parameters are also found for this family. In both cases, to maximize the log-likelihood, we propose a procedure based on heuristic optimization and its combination with the quasi-Newton method. In carrying out this optimization, special attention is paid when discrete or continuous parameters exist. The results are applied to two data sets, the first is available in the literature and the second was simulated.

2. Preliminary considerations

We now present some basic, preliminary results for the development of the present paper.

Let us define the random variable

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

where $Z \sim N(0, 1)$, $\alpha > 0$ and $\beta > 0$. The random variable S is said to have a Birnbaum–Saunders distribution, with the notation $S \sim \mathcal{BS}(\alpha, \beta)$. Furthermore, its density function is given by

$$f_S(s) = \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2\alpha^2}\left(\frac{s}{\beta} + \frac{\beta}{s} - 2\right)\right]\left(\frac{s^{-3/2}(s+\beta)}{2\alpha\beta^{1/2}}\right), \quad S > 0,$$