Computational Statistics and Data Analysis 54 (2010) 1732–1743

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Prediction for Pareto distribution based on progressively Type-II censored samples

Mohammad Z. Raqab [a], A. Asgharzadeh [b,*], R. Valiollahi [c]

[a] Department of Mathematics, University of Jordan, Amman 11942, Jordan

[b] Department of Statistics, Faculty of Basic Science, University of Mazandaran, Post Code 47416-1407, Babolsar, Iran

[c] Department of Statistics, Faculty of Basic Science, Semnan University, Post Code 35195-363, Semnan, Iran

ARTICLE INFO

Article history:
Received 1 December 2008
Received in revised form 21 December 2009
Accepted 6 February 2010
Available online 13 February 2010

Keywords:
Progressive Type-II censoring
Pareto distribution
Best linear unbiased predictor
Maximum likelihood predictor
Prediction interval
Highest conditional density
Monte Carlo simulation

ABSTRACT

In this paper, we discuss different predictors of times to failure of units censored in multiple stages in a progressively censored sample from Pareto distribution. The best linear unbiased predictors, maximum likelihood predictors and approximate maximum likelihood predictors are considered. We also present two methods for obtaining prediction intervals for the times to failure of units. A numerical simulation study involving two data sets is presented to illustrate the methods of prediction.

1. Introduction

The two-parameter Pareto distribution (denoted by Pareto(μ, σ)) has the probability density function (pdf)

$$f(x; \mu, \sigma) = \frac{\alpha}{\sigma}\left(1 + \frac{x-\mu}{\sigma}\right)^{-(\alpha+1)} \quad x > \mu,\ \alpha > 0,\ \sigma > 0, \tag{1.1}$$

and the cumulative distribution function (cdf)

$$F(x; \mu, \sigma) = 1 - \left(1 + \frac{x-\mu}{\sigma}\right)^{-\alpha}, \quad x > \mu,\ \alpha > 0,\ \sigma > 0. \tag{1.2}$$

Here μ is the location parameter and σ is the scale parameter.

The Pareto distribution was originated by Pareto as a model for the distribution of income but is now used as a model in such widely diverse areas as insurance, business, economics, engineering, hydrology and reliability. The origin and other aspects of this distribution can be found in Johnson et al. (1994).

The prediction problems of the life time models are very important in medical and engineering sciences and have been studied, among others by Aitchison and Dunsmore (1975), Lawless (1982), Kaminsky and Rhodin (1985), Patel (1989), Raqab and Nagaraja (1995), Kaminsky and Nelson (1988), Awad and Raqab (2000) and Basak et al. (2006).

* Corresponding author.
E-mail addresses: mraqab@ju.edu.jo (M.Z. Raqab), a.asgharzadeh@umz.ac.ir (A. Asgharzadeh), rvaliollahi@semnan.ac.ir (R. Valiollahi).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.02.005

The progressive Type-II censoring, after starting the life-testing experiment with n units, arises as follows: n units are placed on a life-testing experiment and only m ($m < n$) units are completely observed until failure. The censoring occurs progressively in m stages. These m stages offer failure times of the m completely observed units. At the time of the first failure (the first stage), R_1 of the $n-1$ surviving units are randomly withdrawn (censored intentionally) from the experiment, at the time of the second failure (the second stage), R_2 of the $n-2-R_1$ surviving units are withdrawn, and so on. Finally, at the time of the mth failure (the mth stage), all the remaining $R_m = n-m-R_1-\cdots-R_{m-1}$ surviving units are withdrawn. We will refer to this as progressive Type-II right censoring scheme $(R_1, R_2, \ldots, R_m)$. It is clear that this scheme includes the conventional Type-II right censoring scheme (when $R_1 = R_2 = \cdots = R_{m-1} = 0$ and $R_m = n-m$) and complete sampling scheme (when $n = m$ and $R_1 = R_2 = \cdots = R_m = 0$). The ordered lifetime data which arise from such a progressive Type-II right censoring scheme are called progressively Type-II right censored order statistics.

When data are obtained by progressive censoring, inference problems for various models have been studied by many authors including Viveros and Balakrishnan (1994), Balakrishnan and Kannan (2000), Balakrishnan and Asgharzadeh (2005) and Asgharzadeh (2006). For further details on progressively censoring and relevant references, the reader may refer to the book by Balakrishnan and Aggarwala (2000).

Let $X_{1:m:n}, X_{2:m:n}, \ldots, X_{m:m:n}$, denote the above-mentioned progressively Type-II right censored observed sample from the Pareto(μ, σ) distribution. For notation simplicity, we will write X_i for $X_{i:m:n}$.

The aim of this paper is to discuss the prediction of the life lengths $Y_{s:R_i}$ ($s = 1, 2, \ldots, R_i$; $i = 1, 2, \ldots, m$) of all censored units in all m stages of censoring based on observed data $\mathbf{X} = (X_1, \ldots, X_m)$. Here $Y_{s:R_i}$ ($s = 1, 2, \ldots, R_i$; $i = 1, 2, \ldots, m$) denotes the sth order statistic from a sample of size R_i removed at stage $i = 1, 2, \ldots, m$. The best linear unbiased predictor (BLUP) is considered in Section 2. In Section 3, we provide the maximum likelihood predictor (MLP) and the approximate maximum likelihood predictor (AMLP) for $Y_{s:R_i}$ ($s = 1, 2, \ldots, R_i$; $i = 1, 2, \ldots, m$). In Section 4, we present two methods for obtaining prediction intervals (PIs) for the times to failure of units. Finally, in Section 5, we present two numerical examples as well as a Monte Carlo simulation study to illustrate all the prediction methods discussed in this paper.

2. Best linear unbiased predictor

Suppose $\mathbf{X} = (X_1, \ldots, X_m)$ denotes the m progressively Type-II right censored sample from a location-scale parameter family with cdf $F(x, \theta)$ and pdf $f(x, \theta)$ with $\theta = (\mu, \sigma)$, where μ is the location parameter and σ is the scale parameter such that

$$f(x; \mu, \sigma) = \frac{1}{\sigma} h\left(\frac{x-\mu}{\sigma}\right),$$

where $h(.)$ is the corresponding standardized distribution. Our interest is to predict $Y_{s:R_i}$ ($s = 1, 2, \ldots, R_i$; $i = 1, 2, \ldots, m$) in all m stages of censoring.

A statistic T is called a best linear unbiased predictor (BLUP) if it has the form $c_1X_1 + c_2X_2 + \cdots + c_mX_m$ for real c_i's such that the prediction error $T - Y_{s:R_i}$ has a mean zero and its variance is less than or equal to that of any other linear unbiased predictor of $Y_{s:R_i}$. The BLUP of $Y_{s:R_i}$, is given by (cf. Basak et al. (2006))

$$\widehat{Y}_{\text{BLUP}} = \mu^* + U_1\sigma^* + \mathbf{w}^{\mathbf{T}}\mathbf{\Sigma}^{-1}\left(\mathbf{X} - \mu^*\mathbf{1} - \sigma^*\mathbf{U}\right), \tag{2.1}$$

where μ^* and σ^* are the best linear unbiased estimates (BLUEs) of μ and σ, $\mathbf{1}$ is an $m \times 1$ vector of 1's, and $\mathbf{U}$ and $\mathbf{\Sigma}$ are, respectively, the vector of means and dispersion matrix of the m progressively Type-II right censored order statistics from the corresponding standardized distribution. The value U_1 is the expected value of the sth order statistic out of R_i units from the corresponding standardized conditional distribution and $\mathbf{w}^T = (w_1, w_2, \ldots, w_m) = (\sigma_{1s}, \sigma_{2s}, \ldots, \sigma_{ms})$ with $\sigma_{js} = \text{Cov}(X_j, Y_{s:R_i})$. For obtaining the BLUP of $Y_{s:R_i}$, we compute $\mu^*,\ \sigma^*,\ \mathbf{U},\ \mathbf{\Sigma}^{-1},\ U_1$ and $\mathbf{w}^T$.

Let $X_1, X_2, \ldots, X_m$ denote a progressively Type-II censored sample from the Pareto(μ, σ) distribution, with $(R_1, R_2, \ldots, R_m)$ being the progressive censoring scheme. We know that if $X \sim$ Pareto(μ, σ), then $Z = 1 + (X-\mu)/\sigma$ has the standard Pareto(α) distribution with pdf

$$g(z) = \alpha z^{-\alpha-1}, \quad z > 1 \tag{2.2}$$

and cdf

$$G(z) = 1 - z^{-\alpha}, \quad z > 1. \tag{2.3}$$

Now, let $Z_1, Z_2, \ldots, Z_m$ be a progressively Type-II censored sample from the standard Pareto(α) distribution with pdf (2.2). In other hand, $Z_j = 1 + (X_j - \mu)/\sigma$ where $X_j, j = 1, \ldots, m$ are progressively Type-II censored sample from the location-scale Pareto(μ, σ) distribution. If μ'_j and σ'_{jl} denote the $E(Z_j)$ and $\text{Cov}(Z_j, Z_l)$, respectively. Then (see Balakrishnan and Aggarwala, 2000, p. 26)

$$\mu'_j = \prod_{k=1}^{j} b_k$$

$$\sigma'_{jl} = \left(\prod_{k=1}^{j} d_k - \prod_{k=1}^{j} b_k\right)\prod_{k=1}^{l} b_k \quad 1 \le j \le l \le m, \tag{2.4}$$