Estimation of Parameters of the Pareto Distribution Using a Minimization Technique

Rohan J. Dalpatadu
Department of Mathematical Sciences
University of Nevada
Las Vegas, United States.

Ashok K. Singh
William F. Harrah College of Hotel Administration
University of Nevada
Las Vegas, United States.

Abstract

Estimation of Parameters of the two-parameter Pareto distribution is considered in this paper. It is known that if a random variable has a Pareto distribution with parameters α *and* λ, *then* $Y = \ln\left(1 + X/\lambda\right)$ *is exponentially distributed with mean* $1/\lambda$. *This property is used to develop a method of estimation based on minimization of a distance function such as the Kolmogorov-Smirnov distance using the golden section search procedure. Simulation examples are provided.*

Keywords: Pareto Distribution, parameters, mean, Kolmogorov-Smirnov distance, golden-section search procedure

1. Introduction

The two-parameter Pareto distribution is a commonly used model in reliability and risk theory. Minimum variance unbiased estimates of the parameters of the Pareto distribution are not known. In this paper, we propose and investigate a method of estimation of the parameters of the Pareto distribution. This method is based on minimization of the Kolmogorov-Smirnov distance between the empirical cumulative distribution function (cdf) and the cdf of the Pareto distribution, using the golden section search procedure (See (1)).

2. The Model

The two-parameter Pareto distribution has the probability density function (pdf)

$$f(x;\alpha,\lambda) = \frac{\alpha\lambda^{\alpha}}{\left(\lambda + x\right)^{\alpha+1}}, \quad x > 0,\ \alpha > 0,\ \lambda > 0. \tag{2.1}$$

and the cdf

$$F\left(x;\alpha,\lambda\right) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\alpha} \tag{2.2}$$

It is known that if X has a Pareto distribution $PARETO\left(\alpha,\lambda\right)$, then the random variable

$$Y = \ln\left(1 + \frac{X}{\lambda}\right) \tag{2.3}$$

has an exponential distribution with mean $\frac{1}{\alpha}$ (See (2)).

3. Proposed Estimation Methods

Let $x_1, x_2, ..., x_n$ be an independent random sample, sorted in increasing order, from the population with pdf given by (2.1). The following procedure is used to estimate the parameters α *and* λ:

Step 1: Input an initial search interval (L, H) for the parameter λ.

Step 2: Set

$$I_0 = H - L,$$
$$k = \frac{\sqrt{5}-1}{2},$$
$$\lambda_1 = H - kI_0,$$
$$\lambda_2 = L + kI_0.$$

Step 3: Compute

$$y_{1i} = \ln\left(1 + \frac{x_i}{\lambda_1}\right),$$
$$y_{2i} = \ln\left(1 + \frac{x_i}{\lambda_2}\right),$$

for $i = 1, 2, ..., n$.

Step 4: Compute

$$\hat{\alpha}(\lambda_l) = \frac{n}{\sum_{i=1}^{n} y_{li}}, \quad l = 1, 2.$$

Step 5: Compute

$$D(\lambda_l) = \sup_x \left|\hat{F}(x) - F\left(x : \hat{\alpha}(\lambda_l), \lambda_l\right)\right|, \quad l = 1, 2,$$

where $\hat{F}(x)$ is the sample cdf.

Step 6: If $D(\lambda_l) < D(\lambda_2)$, then $H = \lambda_2$; else $L = \lambda_1$.

Steps (1) – (6) are the steps for the univariate minimization method of Golden Section Search (See (1)). These steps are repeated until $I_0 <$ desired tolerance.

4. Simulation Examples

The following simulation experiment is used to generate data from a Pareto distribution:

Step 1: Generate an ordered sample $u_1 > u_2 > ... > u_n$ from a uniform distribution over the interval $(0,1)$.

Step 2: Transform u_i, $i = 1, 2, ..., n$ to generate observations v_i, $i = 1, 2, ..., n$ from an exponential distribution with mean $\frac{1}{\alpha}$, using: $v_i = -\frac{1}{\alpha}\ln u_i$, $i = 1, 2, ..., n.$

Step 3: Generate $x_1, x_2, ..., x_n$ from the Pareto distribution (2.1), as follows:

$$x_i = \lambda\left(e^{v_i} - 1\right), \quad i = 1, 2, ..., n.$$

The generated data from the Pareto distribution is ordered as $x_1 < x_2 < ... < x_n$. With this ordering, the sample cdf is given by:

$$\hat{F}(x_i) = \frac{i}{n+1}.$$

The following examples are generated with the same input values:

$n = 25,\ \alpha = 1.5,\ \lambda = 8.$

Example1: The simulated sample of 25 observations is given below:

0.06, 0.14, 0.24, 0.32, 1.41, 1.61, 1.83, 2.74, 2.84, 3.06, 3.54, 3.94, 4.16, 4.38, 4.67, 6.11, 7.52, 10.77, 11.20, 11.42, 13.74, 16.61, 21.82, 36.39, 52.22.

The proposed method gave us the estimates: $\hat{\alpha} = 1.98$, $\hat{\lambda} = 10.02$, with $D_{\min} = 0.09$.

Example 2: The simulated sample of 25 observations is given below:

0.42, 1.02, 1.59, 1.64, 1.91, 1.97, 2.40, 2.83, 2.98, 4.51, 4.58, 7.54, 8.22, 8.79, 10.31, 10.77, 13.24, 14.19, 24.06, 24.83, 41.26, 52.88, 57.61, 61.73, 191.77.

The proposed method gave us the estimates: $\hat{\alpha} = 1.28$, $\hat{\lambda} = 10.45$, with $D_{\min} = 0.09$.

Conclusions

The proposed method gives us a fairly simple procedure to estimate the parameters of the Pareto distribution. The obtained estimates are fairly accurate if we take into consideration the sample size. The accuracy may be improved by using a larger sample.

References

RavindranA., Phillips D., and Solberg J.(1987), Operations Research: Principles and Practice, pp. 509-511.

Panjer, H. H. and Willmott G. E. (1992), Insurance Risk Models, Society of Actuaries, Schaumberg, IL.