International Mathematical Forum, 3, 2008, no. 10, 473 - 478

Preliminary Test Estimation in the Pareto Distribution Using Minimax Regret Significance Levels

Ayman Baklizi

Department of Mathematics and Physics
Qatar University, Doha, Qatar
a.baklizi@qu.edu.qa

Abstract

We consider preliminary test estimator based on the maximum likelihood estimator of the parameter of the pareto distribution. The optimal significance levels for the preliminary test are obtained using the minimax regret criterion. The corresponding critical values of the preliminary test are calculated.

Mathematics Subject Classification: 62F10

Keywords: Maximum likelihood estimator, Minimax regret criterion, Preliminary test estimator, Optimum significance levels. Pareto distribution.

1. Introduction

The Pareto distribution has been used in a variety of fields, especially in economics. It is often used to model the distribution of income (Steindl, 1965; Mandelbrot, 1967, Ord, 1975). Various estimation methods for the parameter of the Pareto distribution have been discussed in the literature. The least squares and the moment estimators and their properties have been discussed by Quandt (1966). The maximum likelihood estimator is dicussed by Malik (1970). Estimators based on order statistics are discussed by (Ogawa, 1962; Koutrouvelis, 1981) among others. Minimax, Bayesian and other issues of estimation in this model are reviewed in Johnson et. al. (1994)

In some applications, the researcher possesses some knowledge about the parameter a in the form of a prior estimate a_0. This prior information may be incorporated to improve the estimation process using a preliminary test estimator (Ohtani and Toyoda, 1978; Toyoda and Wallace, 1975; Sawa and Hiromatsu, 1973). In this paper we present a preliminary test estimator for the parameter of the Pareto distribution. Similar work in this problem was done by Chiou (1978, 1988) and Baklizi (2005) among others. The procedure for obtaining the optimum values of the significance levels using the minimax regret criterion of Brook (1976) is developed in section 2. The results are given in the final section.

2. Preliminary test estimation

Consider a random sample $x_1,\ldots,x_n$ from the Pareto distribution with pdf

$$f(x) = ak^a x^{-(a+1)} \qquad a,k>0, \quad x \geq k.$$

The maximum likelihood estimator of a when k is unknown is given by (Johnson et. al., 1994) $\hat{a} = n\left[\sum_{i=1}^{n} \ln\left(\frac{x_i}{\min(x_i)}\right)\right]^{-1}$, It can be shown that. $2na/\hat{a} \sim \chi^2_{2(n-1)}$ (Johnson et. al., 1994). Assume that a_0 is a prior guess of a, Consider testing $H_0 : a = a_0$ against $H_1 : a \neq a_0$, the likelihood ratio test rejects H_0 when $2na_0/\hat{a} > c_1$ or $2na_0/\hat{a} < c_2$, a preliminary test estimator $\tilde{a}$ of a may be obtained as follows

$$\tilde{a} = \begin{cases} a_0, & c_1 < \dfrac{2n\,a_0}{\hat{a}} < c_2 \\ \hat{a}, & \text{Otherwise.} \end{cases}$$

where c_1 and c_2 are such that $p_{a_0}(W < c_1) = p_{a_0}(W > c_2) = \dfrac{\alpha}{2}$, where $W \sim \chi^2_{2(n-1)}$.

Our aim is to find the optimum values of α according to the minimax regret criterion. The mean of $\tilde{a}$ is given by

$$E(\tilde{a}) = a_0 E\left[I\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right)\right] + E\left[\hat{a}\left\{1 - I\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right)\right\}\right]$$

Notice that

$$E\left[I\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right)\right] = p\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right) = \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} g(w)\,dw,$$

where $g(w)$ is the pdf of a chi-squared random variable with $2(n-1)$ degrees of freedom. Also;

$$E\left[\hat{a}\left\{1 - I\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right)\right\}\right] = E(\hat{a}) - E\left(\hat{a}\, I\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right)\right)$$

Now, $E(\hat{a}) = 2naE\left(\frac{\hat{a}}{2na}\right) = 2naE\left(\frac{1}{W}\right) = \frac{na}{na-2}$, $W \sim \chi^2_{2(n-1)}$, and;

$$E\left(\hat{a}\, I\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right)\right) = 2naE\left(\frac{\hat{a}}{2na} I\left(\frac{c_1 a}{a_0} < \frac{2na}{\hat{a}} < \frac{c_2 a}{a_0}\right)\right) = 2na \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} \frac{1}{w} g(w)dw$$

Thus; $E(\tilde{a}) = a_0 \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} g(w)dw + \frac{na}{na-2} - 2na \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} \frac{1}{w} g(w)dw$

Similarly, the second moment of $\tilde{a}$ is given by

$$E(\tilde{a}^2) = a_0^2 \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} g(w)dw + \frac{n^2 a^2}{(n-2)^2(n-3)} + \left(\frac{na}{n-2}\right)^2 - (2na)^2 \int_{c_1\frac{a}{a_0}}^{c_2\frac{a}{a_0}} \left(1/w^2\right) g(w)dw$$

The mean squared error of $\tilde{a}$ is given by

$MSE(\tilde{a}) = E(\tilde{a}^2) - (E(\tilde{a}))^2 + (E(\tilde{a}) - a)^2 = E(\tilde{a}^2) - 2aE(\tilde{a}) + a^2$

Thus;

$$MSE(\tilde{a}) = a_0^2 \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} g(w)dw + \frac{n^2 a^2}{(n-2)^2(n-3)} + \left(\frac{na}{n-2}\right)^2 - (2na)^2 \int_{c_1\frac{a}{a_0}}^{c_2\frac{a}{a_0}} \left(1/w^2\right) g(w)dw - 2a\left(a_0 \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} g(w)dw + \frac{na}{na-2} - 2na \int_{\frac{c_1 a}{a_0}}^{\frac{c_2 a}{a_0}} \frac{1}{w} g(w)dw\right) + a^2$$

Now $\frac{MSE(\tilde{a})}{a^2}$ can be considered as a risk function (Chiou, 1988), let $\psi = \frac{\sigma_0}{\sigma}$ we get

$$RIS(\psi, \alpha) = \psi^2 \int_{\frac{c_1}{\psi}}^{\frac{c_2}{\psi}} g(w)dw + \frac{n^2}{(n-2)^2(n-3)} + \left(\frac{n}{n-2}\right)^2 - (2n)^2 \int_{\frac{c_1}{\psi}}^{\frac{c_2}{\psi}} \left(1/w^2\right) g(w)dw - 2\left(\psi \int_{\frac{c_1}{\psi}}^{\frac{c_2}{\psi}} g(w)dw + \frac{na}{na-2} - 2n \int_{\frac{c_1}{\psi}}^{\frac{c_2}{\psi}} \frac{1}{w} g(w)dw\right) + 1$$

Notice that the risk function depends on α through c_1 and c_2 which are determined such that $p_{a_0}(W < c_1) = p_{a_0}(W > c_2) = \frac{\alpha}{2}$, where $W \sim \chi^2_{2(n-1)}$.

If $\psi \to 0$ or ∞, then $RIS(\psi,\alpha)$ converges to $RIS(\psi,1)$ which is the risk of the maximum likelihood estimator $\hat{a}$. The general shapes of $RIS(\psi,\alpha)$ can be found in (Chiou, 1988; figure 1). An optimal value of α is $\alpha = 1$ if $\psi \le \psi_1$ or $\psi \ge \psi_2$ and $\alpha = 0$ otherwise, where ψ_1 and ψ_2 are intersections of $RIS(\psi,0) = (\psi - 1)^2$ with $RIS(\psi,1) = \frac{n^2}{(n-2)^2(n-3)} + \frac{4}{(n-2)^2}$ .The intersections are $\psi_1 = 1 - \left(\frac{n^2}{(n-2)^2(n-3)} + \frac{4}{(n-2)^2} \right)^{1/2}$ and $\psi_2 = 1 + \left(\frac{n^2}{(n-2)^2(n-3)} + \frac{4}{(n-2)^2} \right)^{1/2}$. Since ψ is unknown we seek an optimal value $\alpha = \alpha^*$ which gives a reasonable risk for all values of ψ. Going along the lines of Sawa and Hiromatsu (1973), the regret function is

$REG(\psi,\alpha) = RIS(\psi,\alpha) - \inf_{\alpha} RIS(\psi,\alpha)$, where

$$\inf_{\alpha} RIS(\psi,\alpha) = \begin{cases} RIS(\psi,1), & \psi \le \psi_1 \text{ or } \psi \ge \psi_2 \\ RIS(\psi,0), & \text{otherwise.} \end{cases}$$

For $\psi \le \psi_2$ $REG(\psi,\alpha)$ takes a maximum value at ψ_L. For $\psi > \psi_2$, $REG(\psi,\alpha)$ takes a maximum value at ψ_U, see (Chiou, 1988; figure 1). Thus the minimax regret criterion determines α^* such that $REG(\psi_L,\alpha^*) = REG(\psi_U,\alpha^*)$. An estimator for a that uses the minimax regret significance levels now can be defined as

$$\tilde{a} = \begin{cases} a_0, & c_1 < \frac{2na_0}{\hat{a}} < c_2 \\ \hat{a}, & \text{otherwise.} \end{cases}$$

where c_1 and c_2 are such that $p_{a_0}(W < c_1) = p_{a_0}(W > c_2) = \frac{\alpha^*}{2}$, where $W \sim \chi^2_{2(n-1)}$.

3. Results

We found numerically the optimum significance levels α^* and the corresponding critical values for n=4,5,…,20. The results are given in table 1.

Table 1: Optimum significance levels and the corresponding critical values.

N	4	5	6	7	8	9	10	11	12
α^*	0190	0819	1615	2426	3197	3910	4562	5159	5703
c_1	0.374	1.270	2.425	3.727	5.124	6.585	8.095	9.642	11.219
c_2	22.585	22.482	23.827	25.624	27.618	29.713	31.865	34.049	36.254
N	13	14	15	16	17	18	19	20	
α^*	6202	6660	7082	7471	7830	8164	8475	8764	
c_1	12.820	14.442	16.082	17.736	19.404	21.084	22.774	24.473	
c_2	38.471	40.695	42.923	45.153	47.385	49.616	51.846	54.075	

References

[1] Baklizi, A. Preliminary test estimation in the two parameter exponential distribution with time censored data, Applied Mathematics and Computation, 163(2005), 639-643.

[2] Brook, R.J. On the use of a regret function to set significance points in prior tests of estimation. Journal of the American Statistical Association, 71(1976), 126 – 131.

[3] Chiou, P. Shrinkage estimation of scale parameter of the extreme value distribution, IEEE Transactions on Reliability, 37(1988), 370 – 374.

[4] Chiou, P. and Han, C.P. Shrinkage estimation of threshold parameter of the exponential distribution, IEEE Transactions on Reliability, 38(1989), 449 – 453.

[5] Johnson, N.L, Kotz, S. and Balakrishnan, Continuous univariate distributions: vol 1 . Wiley, NewYork, 1994.

[6] Koutrouvelis, I.A. Large-sample quantile estimation in Pareto laws, Communications in Statistics, Theory and Methods, 10(1981), 189-201.

[7] Malik, H.J. Estimation of the parameter of the Pareto distribution, Metrika, 15(1970), 126-132.

[8] Mandelbort, B. The Pareto – Levy law and the distribution of income, International economic review, 1(1960), 79 – 106.

[9] Ohtani, K. and Toyoda, T. Minimax regret critical values for a preliminary test in pooling variances, Journal of the Japan Statistical Society, 8(1978), 15-20.

[10] Ord, J.K. Statistical models for personal income distributions, In Statistical distributions in scientific world, 2, G.P. Patil, S. Kotz, and J.K. Ord. (editors), 1975, 151-158, Dordrecht, Reidel.

[11] Quandt, R.E. Old and new methods of estimation and the Pareto distribution, Metrika, 10(1966), 55-82.

[12] Sawa, H. and Hiromatsu, T. Minimax regret significance points for a preliminary test in regression analysis, Econometrica, 41(1973), 1093 – 1101.

[13] Steindl, J. Random processes and the growth of firms, Hafner, New York, 1965.

[14] Toyoda, T. and Wallace, D. Estimation of variance after a preliminary test of homogeneity and optimum levels of significance for the pre-test. Journal of Econometrics, 3(1975), 395 – 404.

Received: September 18, 2007