DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 4 | Issue 2 Article 23

11-1-2005

Inference on $(Y < X)$ in a Pareto Distribution

M. Masoom Ali
Ball State University, mali@bsu.edu

Jungsoo Woo
Yeungnam University

Recommended Citation

Ali, M. Masoom and Woo, Jungsoo (2005) "Inference on $(Y < X)$ in a Pareto Distribution," *Journal of Modern Applied Statistical Methods*: Vol. 4 : Iss. 2 , Article 23.
DOI: 10.22237/jmasm/1130804520
Available at: http://digitalcommons.wayne.edu/jmasm/vol4/iss2/23

Journal of Modern Applied Statistical Methods
November, 2005, Vol. 4, No. 2, 583-586

1538 – 9472/05/$95.00

BRIEF REPORTS
Inference on *P(Y < X)* in a Pareto Distribution

M. Masoom Ali
Department of Mathematical Sciences
Ball State University

Jungsoo Woo
Department of Statistics
Yeungnam University

Inference on the reliability $R = P(Y < X)$ in a Pareto distribution with a known scale parameter is considered. Point estimates and confidence intervals of R are obtained a test of hypothesis is also considered.

Key words: MLE, MSE

Introduction

A Pareto distribution is given by

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta(1+x/\beta)^{\alpha+1}}, x > 0, \alpha, \beta > 0.$$

Pareto law has been universal and inevitable, regardless of taxation and social and political conditions. More recently, attempts have been made to explain many empirical phenomena using the Pareto distribution (see Moothathu, 1984; Arnold & Press, 1983). Ali, et al, (2005a and 2005b) considered the problem for some other distributions. The probability that a Weibull random variable *Y* is less than another independent Weibull random variable *X* was considered by McCool (1991). Baklizi (2003) considered the confidence interval of $P(X < Y)$ in the exponential case with common location.

M.Masoom Ali is George & Frances Ball Distinguished Professor of Statistics. His research interests are in order statistics, Bayesian statistics, statistical inference, and distribution problems. Email him at mali@bsu.edu. Jungsoo Woo is Professor of Statistics. His research interests are in Bayesian statistics, statistical inference and distribution problems.

The problem of estimating and of drawing inferences about the probability that a random variable *Y* is less than another independent random variable *X* arise in reliability studies.

When *Y* represents the random variable of a stress that a device will be subjected to in service and *X* represents the strength that varies from item to item in the population of devices, then the reliability *R*, i.e., the probability that a randomly selected device functions successfully, is equal to $P(Y < X)$. The same problem also arises in the context of statistical tolerance where Y represents, say, the diameter of a shaft and *X* the diameter of a bearing that is to be mounted on the shaft. The probability that the bearing fits without interference is the $P(Y < X)$. In biometry, Y represents a patient's remaining years of life if treated with drug A and X represents the patient's remaining years when treated with drug B. If the choice of drug is left to the patient, person's deliberations will center on whether $P(Y < X)$ is less than or greater than 1/2.

In this article, the problem of estimating $P(Y < X)$ in a Pareto distribution with a known scale parameter, including point and interval estimation is considered and also a test of hypothesis.

Inference on $P(Y < X)$

Let X and Y be independent random variables from Pareto distributions with parameters (α_x, β) and (α_y, β) respectively.

Then from formula 3.381(4) in Gradshteyn and Ryzhik (1965), the following fact is obtained.

Fact 1:

$$R \equiv P(Y < X) = 1 - \frac{\alpha_x}{\alpha_x + \alpha_y} = \frac{\rho}{1+\rho}$$

is a monotone function of ρ, where $\rho \equiv \frac{\alpha_y}{\alpha_x}$.

Proof:

$$R = P(Y < X) = 1 - \iint_{0<y<x<\infty} f_X(x;\alpha_x,\beta) f_Y(y;\alpha_y,\beta)dxdy$$

where f_X is the Pareto density with parameters (α_x,β) and f_Y is the Pareto distribution with parameters (α_y,β). By formula 3.381(4) in Gradshteyn and Ryzhik (1965), one can integrate and obtain the following.

$$R = P(Y < X) = 1 - \alpha_x \cdot B(1,\alpha_x + \alpha_y),$$

where, $B(a,b)$ is a beta function. Using $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b), a > 0, b > 0,$ the above result is obtained.

Because R is a monotone function of ρ, inference on ρ is equivalent to inference on R. Attention is confined to the parameter ρ (see McCool, 1991). Assume $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$ are independent random samples from $f_X(x;\alpha_x,\beta_0)$ and $f_Y(y;\alpha_y,\beta_0)$, respectively, where β_0 is known. From Johnson *et al* (1995), MLE's of α_x and α_y are

$$\hat{\alpha}_x = \frac{m}{\sum_{i=1}^{m} \ln(1 + X_i/\beta_0)}$$

and

$$\hat{\alpha}_y = \frac{n}{\sum_{i=1}^{n} \ln(1 + Y_i/\beta_0)}.$$

The following results in Fact 2 are well-known.

Fact 2: (a) Assume $X_1, X_2, ..., X_m$ be a random sample from a Pareto distribution with parameters (α_x, β_0). Then $\sum_{i=1}^{m} \ln(1 + X_i/\beta_0)$ follows a gamma distribution with a shape parameter m and a scale parameter $1/\alpha_x$. (b) If a random variable X follows a gamma distribution with shape α and scale β then

$$E(1/X^k) = \frac{\Gamma(\alpha - k)}{\Gamma(\alpha)} \cdot \frac{1}{\beta^k} \text{ if } \alpha > k.$$

From the definition of $\hat{\rho} \equiv \frac{\hat{\sigma}_y}{\hat{\sigma}_x}$, the MLE of ρ

$$\text{is } \hat{\rho} = \frac{\hat{\alpha}_y}{\hat{\alpha}_x} = \frac{n}{m} \cdot \frac{\sum_{i=1}^{m} \ln(1 + X_i/\beta_0)}{\sum_{i=1}^{n} \ln(1 + Y_i/\beta_0)}.$$

From Fact 2(a) and (b), one can obtain the following fact.

Fact 3:

$$E(\hat{\rho}) = \frac{n}{n-1}\rho$$

and

$$Var(\hat{\rho}) = \frac{n^2(m+n-1)}{m(n-1)^2(n-2)}\rho^2, n > 2.$$

From Johnson *et al* (1995),

$$\tilde{\alpha}_x = \frac{m-1}{\sum_{i=1}^{m} \ln(1 + X_i/\beta_0)},$$

and

$$\tilde{\alpha}_y = \frac{n-1}{\sum_{i=1}^{n} \ln(1+Y_i/\beta_0)}$$

are UMVUE of α_x and α_y, respectively.

Define $\tilde{\rho} = \frac{\tilde{\sigma}_y}{\tilde{\sigma}_x} = \frac{n-1}{m-1} \cdot \frac{\sum_{i=1}^{m} \ln(1+X_i/\beta_0)}{\sum_{i=1}^{n} \ln(1+Y_i/\beta_0)}$.

Then one can obtain the following expectation and variance.

$$E(\tilde{\rho}) = \frac{m}{m-1}$$

and

$$Var(\tilde{\rho}) = \frac{m}{(m-1)^2}\rho^2.$$

Therefore, it is obtained:

Fact 4: $MSE(\hat{\rho}) < MSE(\tilde{\rho})$.

To consider a confidence interval for ρ, the following random variables are defined. Let

$$Z \equiv \sum_{i=1}^{m} \ln(1+X_i/\beta_0),$$
$$W \equiv \sum_{i=1}^{n} \ln(1+Y_i/\beta_0)$$

and $U \equiv Z/W$.

By formula 3.381(4) in Gradshteyn and Ryzhik (1965) and the quotient pdf of two independent random variables, the pdf of U is obtained as follows.

$$f_U(u) = \frac{u^{m-1}}{B(m,n)\rho^m}(1+\frac{u}{\rho})^{-m-n}, u > 0,$$

where $B(m,n)$ is the Beta function. From the density of $U = Z/W$, one can easily find the distribution of $B \equiv \frac{U}{\rho+U}$.

Fact 5: Let $B \equiv \frac{U}{\rho+U}$. Then, B follows a beta distribution with parameters *m* and *n*. Based on the pivot quantity B, a confidence interval of ρ is considered. From the beta distribution function, for a given $0 < \alpha < 1$, there exists $0 < b_\alpha < 1$ such that

$$\alpha = \int_0^{b_\alpha} \frac{1}{B(m,n)} x^{m-1}(1-x)^{n-1} dx.$$

Here, for a given $0 < \alpha < 1$, b_α can be easily evaluated by inverse function of the beta distribution using statistical software. Hence, a $(1-\alpha)100\%$ confidence interval of ρ can be obtained as

$$\left(\frac{m}{n} \cdot \frac{1-b_{1-\alpha/2}}{b_{1-\alpha/2}} \cdot \hat{\rho}, \frac{m}{n} \cdot \frac{1-b_{\alpha/2}}{b_{\alpha/2}} \cdot \hat{\rho} \right)$$

and from the result of Fact 3, its expected length is

$$E(L) = \frac{m}{n-1}\left(\frac{1}{b_{\alpha/2}} - \frac{1}{b_{1-\alpha/2}} \right)\rho.$$

Next, the null hypothesis is tested

$H_0: \alpha_x = \alpha_y$ against $H_1: \alpha_x \neq \alpha_y$. Let
$\Theta = \{(\alpha_x, \alpha_y) \mid \alpha_x > 0, \alpha_y > 0\}$, and
$\theta = (\alpha_x, \alpha_y)$.

Then the joint probability density function of $X_1, X_2, ..., X_m, Y_1, Y_2, ..., Y_n$ is

$$L(\theta) = f_\theta(x,y) = \frac{\sigma_x^m \sigma_y^n}{\beta_0^{m+n}} \prod_{i=1}^{m} (1 + x_i / \beta_0)^{-\alpha_x - 1}$$

$$\prod_{i=1}^{n} (1 + y_i / \beta_0)^{-\alpha_y - 1}.$$

Differentiating with respect to α_x and α_y, the MLE's are obtained as follows.

$$\hat{\alpha}_x = \frac{m}{\sum_{i=1}^{m} \ln(1 + x_i / \beta_0)}$$

and

$$\hat{\alpha}_y = \frac{n}{\sum_{i=1}^{n} \ln(1 + y_i / \beta_0)}.$$

If $\alpha_x = \alpha_y = \alpha$, then the MLE of α is

$$\hat{\alpha} = \frac{m+n}{\sum_{i=1}^{m} \ln(1 + x_i / \beta_0) + \sum_{i=1}^{n} \ln(1 + y_i / \beta_0)}.$$

From the definition of likelihood ratio test, the likelihood ratio test function is given by

$$\Lambda(x,y) = \left(\frac{m+n}{m}\right)^m \left(\frac{m+n}{n}\right)^n \times$$
$$\frac{1}{(1 + 1/U)^m} \frac{1}{(1+U)^n},$$

where

$$U = \frac{\sum_{i=1}^{m} \ln(1 + X_i / \beta_0)}{\sum_{i=1}^{n} \ln(1 + Y_i / \beta_0)}.$$

Therefore, $\Lambda(x,y) < c$ is equivalent to $U < c_1$ or $U > c_2$. Under $H_0 : \alpha_x = \alpha_y$, i.e., $\rho = 1$, from Fact 5, the statistic

$$B_0 = \frac{U}{\rho + U} = \frac{U}{1+U}$$

follows a beta distribution with m and n. Because B_0 is a monotone increasing function of U, so $U < c_1$ or $U > c_2$ is equivalent to $B_0 < b_1$ or $B_0 > b_2$. b_1 and b_2 can be obtained by inverse function of a beta distribution and using a statistical software.

References

Ali, M. Masoom & Woo, J. (2005a). Inference on reliability P(Y<X) in a p-dimensional rayleigh distribution. *Mathematical and Computer Modelling - An International Journal*, *42*, 367-373.

Ali, M. Masoom & Woo, J. (2005b). Inference on reliability P(Y<X) in the Levy distribution. *Mathematical and Computer Modelling – An International Journal*, *41*, 965-971.

Abramowitz, M. & Stegun, I. A. (1970). *Handbook of Mathematical Functions*. Dover Publications: New York.

Arnold, B. C. & Press, S. (1983). Bayesian inference for Pareto population. *Journal of Econometrics*, *21*, 287-306.

Baklizi, A. (2003). Confidence intervals for P(X<Y) in the exponential case with common location parameter. *Journal of Modern Applied Statistical Methods*, *2*(2), 341-349.

Gradshteyn, I. S. & Ryzhik, I. M. (1965). *Table of integrals, series and products*. Academic Press: New York.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous distributions*. Houghton Mifflin Co.: Boston.

McCool, J. I. (1991). Inference on P(X < Y) in the Weibull case. *Communication in Statistics, Simulations*. *20*(1), 129-148.

Moothathu, T. S. K. (1984). Characterizations of power function and Pareto distribution in terms of their Lorenz curves. *Research Reviews*, *3*, 118-123.

Obenhettinger, F. (1974). *Tables of mellin transforms*. Springer-Verlag: New York.