The Korean Communications
in Statistics Vol. 8, No. 1, 2001
pp. 249-256

Estimations of Lorenz Curve and Gini Index in a Pareto Distribution

Jung Soo Woo[1] and Gi Ern Yoon[2]

Abstract.

We shall derive the MLE and UMVUE of Lorenz Curve and Gini Index in a Pareto distribution with the pdf(1.1) and their variances. And compare mean square errors(MSE) of the MLE and UMVUE of the Lorenz Curve and Gini Index in a Pareto distribution with pdf(1.1).

Keywords : Lorenz curve, Gini index, MLE, UMVUE.

1. Introduction

Let $Y_1, Y_2, \ldots, Y_n$ be a simple random sample from a Pareto distribution with the pdf:

$$f(y;\beta)=\frac{\beta}{\beta-1}\cdot y^{\frac{2\beta-1}{1-\beta}}, \quad 1<y, \quad 1<\beta, \tag{1.1}$$

which has mean β and variance $\beta(\beta-1)^2/(2-\beta)$, if $1<\beta<2$

A Pareto distribution with the pdf(1.1) is a special case of the Pareto distribution with the scale parameter $\theta=1$ and the shape parameter $x=\beta/(\beta-1)$(see Johnson(1970)), and the density function is decreasing if $\beta>1$. Let $\beta=\frac{\theta}{2\theta-1}$ in a power-function random variable with the pdf $f(x;\theta)=\frac{\theta}{1-\theta}\cdot x^{\frac{2\theta-1}{1-\theta}}$, $0<x<1$, $1/2<\theta<1$ (see Ali. etal(1999)). Then an inverse of the power-function random variable has a Pareto distribution with the pdf (1.1). Hung & Bier(1998) considered the properties of the conjugate prior for the non-homogeneous

1) Professor Department of Statistics, Yeungnam University, Kyungsan, 712-749, Korea
E-mail : jswoo@yu.ac.kr
2) Instructor, Department of statistics, Yeungnam University, Kyungsan, 712-749, Korea

poisson procss with a power-function law intensity. Ali, Woo & Yoon(1999) studied the UMVUE for mean and right-tail probability in a power-function distribution , that is, an inverse of a Pareto random variable with the pdf (1.1).

The Lorenz Curve proved to be a powerful tool for the analysis of a variety of scientific problems: e.g., to measure the income of inequality within a population of income receivers, as a criterion to perform a partial ordering of social welfare states, to assess the progressiveness of a tax system, to extend the concept of the Lorenz Curve to functions of income or other variables, to study the stochastic properties of the Lorenz Curve, as well as upper and lower bounds for the Gini ratio.

The cumulative proportion $F(y)$ of units earning less than a given leval of income is measured on the horizontal axis, and the corresponding cumulative proportion $L(y)$ of total income earned by these units is measured on the vertical axis. The values of $L(y)$ and $F(y)$ for various levels of income are then plotted in a series of points and joined by a curve, known as the Lorenz Curve of the distribution. It will start at the origin, corresponding to the lowest income, where both $F(y)$ and $L(y)$ are zero, and go to the point corresponding to the maximum income, where both $F(y)$ and $L(y)$ are unity.

Its formal representation is

$$L(y) = \int_0^y x dF(x)/E(Y), \tag{1.2}$$

where Y is a nonnegative income variable for which the mathematical expectation $E(Y)$ exists, $p = F(y) = 1 - y^{\frac{\beta}{1-\beta}}$ is the cumulative distribution function (cdf) of the population of income receivers. Since the cumulative distribution function of all specified models of income distribution are strictly increasing and continuously differentiable functions, $y = F^{-1}(p) = (1-p)^{\frac{1-\beta}{\beta}}$, $0 \le p \le 1$ is well defined. Replacing it in (1.2),

$$L(p) = \int_0^p F^{-1}(x) dx / E(Y). \tag{1.3}$$

Some properties of the Lorenz Curve for the class of continuously differentiable income distribution models with finite mathematical expectation are summarized by Kotz and Johnson (1981).

The proportion of units will be less than the proportion of income, and so the Lorenz Curve will be a convex curve lying below the diagonal line. The Lorenz Curve is a cdf with $L(0) = 0$ and with $L(1) = \lim_{p \to 1-} L(p) = \lim_{y \to \infty} F(y) = 1$.

Moothathu(1985 & 1990) derived the UMVUE and a strongly consistent asymptotically normal unbiased estimator (SCANUE) of the Lorenz Curve, the Gini Index, and the Theil Entropy Index of the Pareto distribution. Enrique, Ali and Jose(1998) studied a method for estimating Lorenz Curves.

In the paper we shall derive the MLE and UMVUE of the Lorenz Curve and Gini Index in a Pareto distribution with the pdf(1.1) and their variances. And compare mean square errors(MSE) of the MLE and UMVUE of the Lorenz Curve and Gini Index in a Pareto distribution with the pdf(1.1).

2. Estimators of Lorenz Curve $L(p, \beta)$

The Lorenz Curve $L(p, \beta)$ defined in (1.3) of the Pareto distribution with the pdf(1.1) is given by

$$L(p\ ,\ \beta) = 1-(1-p)^{\frac{1}{\beta}},\ \ (0 \le p \le 1). \tag{2.1}$$

Fact 1 (Moothathu(1985)). Let $Y_1, Y_2, \ldots, Y_n$ be a simple random sample from a Pareto distribution with the pdf(1.1) then $S(Y_1, \ldots, Y_n) = \sum_{i=1}^{n} \ln Y_i$ is a complete sufficient statistic for β and $S(Y_1, \ldots, Y_n) = \sum_{i=1}^{n} \ln Y_i$ has a gamma distribution with a shape parameter n and a scale parameter $(\beta - 1)/\beta$.

Since the MLE of β is $\hat{\beta} = (1 - \sum_{i=1}^{n} \ln Y_i\ /\ n)^{-1}$, $\sum_{i=1}^{n} \ln Y_i < n$, the MLE of the Lorenz Curve(2.1) as follows;

$$\hat{L}(p, \beta) = 1-(1-p)^{\frac{1}{\hat{\beta}}} = 1-(1-p)\cdot e^{-\frac{\ln(1-p)}{n}\cdot S},\ \ \ S > 0$$

From the moment generating function, the expectation of the MLE can be obtained as

$$1-(1-p)\left[1+\frac{\beta-1}{\beta n}\ln(1-p)\right]^{-n} \tag{2.2}$$

And the variance of the MLE can be obtained as

$$(1-p)^2\left\{\left[1+\frac{2(\beta-1)}{\beta n}\ln(1-p)\right]^{-n} - \left[1+\frac{\beta-1}{\beta n}\ln(1-p)\right]^{-2n}\right\} \tag{2.3}$$

We can show the following consistency as the result of Moothathu(1985):

Fact 2 (Moothathu(1985)). The MLE of the Lorenz Curve(2.1) in a Pareto distribution with the pdf(1.1) are consistent.

Since the regular conditions are satisfied(see Rohatgi(1976)), the Frechet-Cramer-Rao lower bound(FCRLB) for an unbiased estimator of the Lorenz Curve(2.1) is

$$\frac{1}{n\beta^2}(1-\beta)^2(1-p)^{-\frac{2}{\beta}}[\ln(1-p)]^2.$$

From Fact 1, we can obtain the UMVUE of the Lorenz Curve(2.1) in a Pareto distribution with the pdf(1.1) and its variance.

So we can express the Lorenz Curve(2.1) in a Pareto distribution with the pdf(1.1) as follows;

$$L(p\ ,\ \beta)=1-(1-p)\sum_{k=0}^{\infty}\frac{[-\ln(1-p)]^k}{k!}(1-\frac{1}{\beta})^k.$$

Fact 3. The UMVUE of the Lorenz Curve(2.1) in a Pareto distribution with the pdf(1.1) is

$$\widehat{Lu}(p\ ,\beta)=1-(1-p)\sum_{k=0}^{\infty}\frac{[-\ln(1-p)]^k}{k!}\frac{1}{(n)_k}S^k\ ,$$

and its variance is

$$(1-p)^2\sum_{i,j=0}^{\infty}\frac{[-\ln(1-p)]^{i+j}}{i!j!}\frac{(n)_{i+j}}{(n)_i\ (n)_j}(1-\frac{1}{\beta})^{i+j}-(1-p)^{\frac{2}{\beta}}. \qquad (2.4)$$

Proof. Since $S(Y_1,\dots,Y_n)=\sum_{i=1}^{n}\ln Y_i$ is complete sufficient , from the moment generating function, $\frac{1}{(n)_k}S^k$ is unbiased estimator of $\left(1-\frac{1}{\beta}\right)^k(k=1,2,\dots)$,

where $(n)_k=n(n+1)\cdots(n+k-1)$and $(n)_0=1$, from Lehmann-ScheffeTheorem, we can obtain the UMVUE and the variance of the Lorenz Curve(2.1) in a Pareto distribution with the pdf(1.1).

From the results (2.1) through (2.4) of variances for $\hat{L}(p,\beta)$ $\widehat{Lu}(p,\beta)$, Table 1 shows mean square errors of the MLE and UMVUE for the Lorenz Curve(2.1) in a Pareto distribution with the pdf (1.1) when β=1.5, p=0.01(0.05, 0.5, 0.95, 0.99) and n=10(5)30. From Table 1, the MLE, the UMVUE are almost same when β=1.5, p=0.01(0.05) and n=10(5)30, but the UMVUE of the Lorenz Curve tends to be more efficient than the MLE when β=1.5, p=0.5(0.95, 0.99) and n=10(5)30.

3. Estimators of Gini Index $g(\beta)$

Here we shall consider estimation of Gini Index $g(\beta)$ in a Pareto distribution with the pdf(1.1). From the Lorenz Curve(2.1), the Gini Index $g(\beta)$ of the Pareto distribution with the pdf(1.1) is given by

$$g(\beta)=1-2\int_0^1 L(p,\beta)\,dp$$
$$=1-\frac{2}{\beta-2}. \tag{3.1}$$

We can express the Gini Index(3.1) in a Pareto distribution with the pdf(1.1) as follows;

$$g(\beta) = \sum_{k=1}^{\infty}\frac{1}{2^k}\left(1-\frac{1}{\beta}\right)^k$$

From the facts that the MLE of the Gini Index(3.1) is the MLE of β,
the MLE of Gini Index(3.1) as follows;

$$\widehat{g(\beta)} = \sum_{k=1}^{\infty}\frac{1}{(2n)^k}S^k.$$

Since $\frac{1}{(n)_k}S^k$ is unbiased estimator of $\left(1-\frac{1}{\beta}\right)^k (k=1,2,\ldots)$, where $(n)_0=1$ and $(n)_k=n(n+1)\cdots(n+k-1)$,
the expectation of the MLE can be obtained as

$$\sum_{k=1}^{\infty}\frac{1}{(2n)^k}(n)_k\left(1-\frac{1}{\beta}\right)^k \tag{3.2}$$

And the variance and the bias of the MLE can be obtained as

$$\sum_{i,j=1}^{\infty}\frac{1}{2^{i+j}n^{i+j}}\left((n)_{i+j}-(n)_i(n)_j\right)\left(1-\frac{1}{\beta}\right)^{i+j},$$

$$BIAS(\widehat{g(\beta)}) = \sum_{k=1}^{\infty} \frac{1}{2^k}\left(\frac{(n)_k}{n_k} - 1\right) \cdot \left(1 - \frac{1}{\beta}\right)^k. \quad (3.3)$$

We can show the following consistency:

Fact 4. The MLE of the Gini Index in a Pareto distribution with the pdf(1.1) are consistent.

Since the regular conditions are satisfied(see Rohatgi(1976)), the Frechet-Cramer-Rao lower bound(FCRLB) for an unbiased estimator of the Gini Index is

$$\frac{4}{n} \cdot \frac{\beta^2(\beta-1)^2}{(1+\beta)^4}$$

Since $S(Y_1, \ldots Y_n) = \sum_{i=1}^{n} \ln Y_i$ is a complete sufficient statistics, from Lehmann-Scheffe Theorem, we can obtain the UMVUE of the Gini Index(3.1) in a Pareto distribution with the pdf(1.1) and its variance.

Fact 5. The UMVUE of the Gini Index(3.1) in a Pareto distribution with the pdf(1.1) is

$$\widehat{g_U(\beta)} = \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{1}{(n)_k} S^k$$

and its variance is

$$\sum_{i,j=1}^{\infty} \frac{1}{2^{i+j}}\left(\frac{n_{i+j}}{n_i n_j} - 1\right)\left(1 - \frac{1}{\beta}\right)^{i+j} \quad (3.4)$$

Proof. Since $S(Y_1, \ldots, Y_n) = \sum_{i=1}^{n} \ln Y_i$ is complete sufficient , from the moment generating function, $\frac{1}{(n)_k} S^k$ is unbiased estimator of $\left(1 - \frac{1}{\beta}\right)^k (k = 1, 2, \ldots)$, where $(n)_0 = 1$and $(n)_k = n(n+1)\cdots(n+k-1)$, from Lehmann-ScheffeTheorem, we can obtain the UMVUE and the variance of the Gini Index(3.1) in a Pareto distribution with the pdf(1.1).

From the results (3.1) through (3.4) of variances for $\widehat{g(\beta)}$, $\widehat{g_U(\beta)}$, Table 2 shows mean square errors of the MLE and UMVUE for the Gini Index(3.1) in a Pareto distribution with the pdf (1.1) when β=1.5 and n=10(5)30. From Table 2, the UMVUE of the Gini Index tends to be more efficient than the MLE when β=1.5 and n=10(5)30.

Table 1. The MSE of the MLE and UMVUE of the Lorenz Curve(2.1) in a Pareto distribution with the pdf(1.1) when β=1.5

p	n	MLE	UMVUE	FCRLB
0.01	10	0.0000011	0.000011	0.0000011
	15	0.0000007	0.0000007	0.0000007
	20	0.0000006	0.0000006	0.0000006
	25	0.0000004	0.0000004	0.0000004
	30	0.0000004	0.0000004	0.0000004
0.05	10	0.0000274	0.0000273	0.0000273
	15	0.0000182	0.0000182	0.0000182
	20	0.0000137	0.0000137	0.0000137
	25	0.0000109	0.0000109	0.0000109
	30	0.0000091	0.0000091	0.0000091
0.5	10	0.0022418	0.0021237	0.0021185
	15	0.0014663	0.0014147	0.0014124
	20	0.0010894	0.0010606	0.0010593
	25	0.0008666	0.0008483	0.0008474
	30	0.0007195	0.0007068	0.0007062
0.95	10	0.0027577	0.0019224	0.0018368
	15	0.0015929	0.0012634	0.0012245
	20	0.0011156	0.0009405	0.0009184
	25	0.0008573	0.0007490	0.0007347
	30	0.0006958	0.0006222	0.0006123
0.99	10	0.0011553	0.0005658	0.0005077
	15	0.0005696	0.0003646	0.0003384
	20	0.0003718	0.0002686	0.0002538
	25	0.0002745	0.0002125	0.0002031
	30	0.0002171	0.0001758	0.0001692

Table 2. The MSE of the MLE and UMVUE of the Gini Index (3.1) in a Pareto distribution with the pdf(1.1).

β	n	MLE	UMVUE	FCRLB
1.5	10	0.21092	0.00580	0.00576
	15	0.00415	0.00386	0.00384
	20	0.00305	0.00290	0.00288
	25	0.00241	0.00231	0.00230
	30	0.00199	0.00193	0.00190

References

[1] Castillo, E, Hadi, A. S. & Sarabia, J. M.(1998), A Method for Estimating Lorenz Curves, Communication Statistics-Theory Meth., 27(8), 2037-2063.

[2] Hung, Y.S. & Bier, V.M.(1998), A Natural Conjugate Prior for the non-homogeneous Poisson Process with a Power Law Intensity Function, Communication Statistics-Simulation, 27(2), 525-551.

[3] Johnson, N.L. & Kotz, S.(1970), Continuous Univariate Distributions-2, Houghton Mifflin Company, Boston.

[4] Moothathu, T. S. K.(1985), Sampling distributions of Lorenz curve and Gini index of the Pareto distribution. *Sankhya*, Series B, Vol.47, 247-258.

[5] Moothathu, T. S. K.(1990), The Best Estimator of Lorenz Curve, Gini Index and Theil Entropy Index of Pareto Distribution, *Sankhya*, Series B, Vol.52, 115-127.

[6] Rohatgi, V.K.(1976), An Introduction to Probability Theory and Mathematical Statististics, John Wiley & Sons, New York.

[7] Woo, J.S. Ali, M.M. & Yoon, G.E.(1999), The UMVUE for Mean and Right-tail Probability in a Power-Function Distribution, to appear it in ESTADISTICA(India)

[8] Woo, J.S. & Yoon, G.E.(1999), The Mean and Right-Tail Probability Estimations in a Pareto Distribution, in preparation at Yeungnam University, South Korea.